Brain Talk : discourse with and in the brain : papers from the first Birgit Rausing Language Program Conference in Linguistics, Lund, June 2008

Alter, Kai; Horne, Merle; Lindgren, Magnus; Roll, Mikael; von Koss Torkildsen, Janne

2009

Papers from the first Birgit Rausing Language Program Conference in Linguistics, Lund, June 2008

Discourse with and in the brain

Edited by

Kai Alter
Merle Horne
Magnus Lindgren
Mikael Roll
Janne von Koss Torkildsen

Brain Talk

Discourse with and in the brain

Editors:

Kai Alter
Merle Horne
Magnus Lindgren
Mikael Roll
Janne von Koss Torkildsen

Series Editor:

Sven Strömqvist

LUNDS UNIVERSITET

This book appears in the Birgit Rausing Language Program Conference Series in Linguistics.

Editors: Kai Alter, Merle Horne, Magnus Lindgren, Mikael Roll, Janne von Koss Torkildsen
Editorial assistance: Marcus Uneson

Layout: Marcus Uneson
This book is set in Adobe Garamond, with headings in Frutiger.

Cover: Marcus Uneson
Brain images used for the cover:
Sarah Bihler, letters ‘B’ and ‘L’
Iris-Corinna Schwarz, letter ‘N’
Yury Shtyrov, all others

Printed and bound by Media-Tryck, Lund, Sweden, 2009

ISBN 978-91-633-5561-5

Contents

Pragmatics

Prosody

Processing and modelling

Clinical aspects

Appendix

Series Editor's Preface

It is with great pleasure we now present to the readership the first volume of a new series publishing conference proceedings from the Birgit Rausing Language Program in Linguistics. The series is dedicated to new and vital trends in linguistics and neighbouring sciences, focussing on language and its role in communication and cognition. The first volume testifies to the rapidly growing interface between linguistics and cognitive neuroscience.

We would like to express our gratitude to Dr. Birgit Rausing for her deep engagement in the field of language studies and for her generous support of Lund University, making this conference series possible. A special thanks also goes to Kai Alter, Merle Horne, Magnus Lindgren, Mikael Roll, and Janne von Koss Torkildsen for editing this first volume, and to Marcus Uneson for editorial assistance and layout.

Lund, October 2009

Sven Strömqvist
Series Editor

Editors' Preface

Kai Alter
Institute of Neuroscience, Newcastle University

Merle Horne
Dept. of Linguistics, Lund University

Magnus Lindgren
Dept. of Psychology, Lund University

Mikael Roll
Dept. of Linguistics, Lund University

Janne von Koss Torkildsen
Dept. of Biologial and Medical Psychology, University of Bergen

Volume editors and members of the
Brain Talk organizing committee

During two sunny spring days, June 2–3 2008, the Center for Languages and Literature, Lund University was the venue for the first Birgit Rausing Language Program conference in linguistics, Brain Talk. The conference, opened by Dr. Birgit Rausing herself, was an interdisciplinary event, involving researchers in linguistics, neuropsychology, and speech therapy.[1] Brain Talk had as its theme different aspects of language processing in the brain. The contributions focussed on the importance of context for the neurocognitive processing of language and speech in various communicative situations.

The present volume, the first in the Birgit Rausing Language Program conference series, contains papers from the Brain Talk conference. They have been grouped into six general areas of research in language processing in the brain: Lexicon, Syntax, Pragmatics, Prosody, Processing and modelling, and Clinical aspects. Needless to say, there is some overlapping between the categories (e.g. some of the papers in the prosody section also deal with syntax, lexicon, pragmatics, and clinical aspects) but we felt that it would be helpful to the reader to make a division of the contributions according to the

[1] A list of the speakers and poster presenters appears in Appendix A.

language parameter that constituted the main focus of investigation. Below we present a summary of the contributions within the various sections.

Before doing so, however, the organizing committee would like to extend their deepest gratitude to Birgit Rausing for providing the generous funding that made Brain Talk as well as future conferences in the Birgit Rausing Language Program Conference Series in Linguistics possible. We would also like to thank Marcus Uneson for painstaking editorial assistance and layout work for this first volume in the conference series.

Lexicon

In their study, "Auditory repetition of words and pseudowords: An fMRI study" Sarah Bihler, Dorothee Saur, Stefanie Abel, Dorothee Kümmerer, Walter Huber, Jürgen Dittmann, and Cornelius Weiller compare the neural correlates of repetition of real words and pseudowords using fMRI. They find that pseudowords engage brain areas related to phonological processing to a larger extent, whereas real words activate areas associated with both implicit and explicit semantic processing.

Francesca Citron, Brendan Weekes and Evelyn Ferstl present in "Evaluation of lexical and semantic features for English emotion words" evidence from a large lexical rating study which indicates that the two emotional dimensions "valence" and "arousal" are independent of each other, as arousal is not reducible to the intensity of valence. Moreover, the authors find different levels of arousal for positive and negative words, suggesting independence between positive and negative affects. The ratings collected by Citron and colleagues will be useful for researchers designing language comprehension experiments in the future.

Drawing on evidence from ERPs and reaction times, Armina Janyan, Ivo Popivanov, and Elena Andonova argue in "Concreteness effect and word cognate status: ERPs in single word translation" that cognates – words in a second language that are phonetically similar to the corresponding first-language words – are interpreted by direct activation of conceptual representations. Access to the conceptual representations of noncognate words, however, seems to be more mediated by words from the first language.

In her contribution, "Electrophysiological correlates of word learning in toddlers", Janne von Koss Torkildsen presents an ERP-based data set that provides insights into the processes of early word learning. She shows striking effects of productive vocabulary on the learning process in toddlers.

Syntax

Zoltán Bánréti proposes in "Restricted discrimination between local and global economy in agrammatic aphasia" that the limited use of binding principles in agrammatic aphasia can be explained in terms of impaired discrimination between global and local economy. To support this view, he presents evidence from a sentence-picture matching test with two Hungarian speakers with Broca's aphasia.

In their chapter "Negative polarity items and complementizer agreement in Basque", Leticia Pablos and Douglas Saddy investigate the interplay between negative polarity item (NPI) and complementizer agreement licensing in Basque in a self-paced reading experiment. They find effects of the semantic environment on the licensing of complementizer agreement, which are blocked by interfering NPIs.

Using a typological perspective in their study "When semantic P600s turn into N400s", Matthias Schlesewsky and Ina Bornkessel-Schlesewsky present recent evidence for the extended Argument-Dependency Model by investigating languages with different syntactic parameters. In a series of EEG studies, they show that the P600 reflects argument linking rather than purely semantic processes.

Pragmatics

In a set of Eye-Tracking and visual and auditory ERP-experiments, Kerry Ledoux investigates discourse phenomena affecting the processing of coreference. Results reported in her study "On-line studies of repeated name and pronominal coreference" show that factors such as discourse prominence, lexical repetition, and implicit causality interact when anaphoric reference is established.

In their contribution "Can context affect gender processing?", Hamutal Kreiner, Sibylle Mohr, Klaus Kessler, and Simon Garrod report results from two ERP-experiments showing distinct neural signatures for reference resolution of stereotypical gender role nouns (e.g. *minister*) compared to definitional gender role nouns (e.g. *king*). Their findings suggest that discourse constraints influence the processing of the two noun types in different ways.

Prosody

In her contribution, "Electrophysiological studies of prosody", Mireille Besson provides an overview on EEG studies and prosodic processing in different languages. She especially focuses on the interfaces between syntax, semantic, and informational structure investigating the impact of each information level on prosodic realisation and prosodic processing.

In their chapter "The processing of emotional utterances: Roles of prosodic and lexical information", Carly Metcalfe, Manon Grube, Damien Gabriel, Susanne Dietrich, Hermann Ackermann, Vivian Cook, Andy Hanson and Kai Alter investigate the contributions of prosodic and lexical features on the processing of interjections in English. Perceptual ratings from a number of participants indicate that prosodic cues may play a more important role than lexical information in emotional interpretation. The study identifies optimal stimuli for future experiments on interjection processing.

Inger Moen's chapter "Prosody and the brain", reviews a number of clinical studies, with particular emphasis on Norwegian data, which throw light on the question of hemispheric specialization as regards the processing of prosody.

Reviewing studies on speech production and processing, Mikael Roll argues in "The role of a left-edge tone in speech processing" that not only right-edge prosody is relevant in the syntactic processing of Swedish. A left-edge boundary tone is proposed as an important cue to syntactic parsing.

In "Fishing for information: The interpretation of focus in dialogs", Ulrike Toepel, Ann Pannekamp, and Elke van der Meer investigate the influence of pragmatic context on the processing of novelty and correction focus. Positive ERP effects show that the participants interpret constituents as being focused when they are associated with pragmatic focus due to the context, regardless of their prosodic realization. In addition, N400 effects show integration difficulty when the intonation pattern is unexpected due to the pragmatic context.

Processing and modelling

Giorgos Argyropoulos proposes in "Neocerebellar emulation of language processing" a neurolinguistic model, where the posterolateral cerebellum is involved in aspects of language processing concerning speed, prediction, and noise-resistance. He also shows the role the cerebellum might play in grammaticalization in connection with its proposed automatizing function.

In his contribution, "On the emergence of early linguistic functions: A biological and interactional perspective", Francisco Lacerda describes a theoretical framework for early language development, Ecological Theory of Language Acquisition, in which he integrates the importance of perception and interaction for the emergence of the early lexicon.

Iris-Corinna Schwartz, Francisco Lacerda, Heléne Walberg and Ellen Marklund explored in their study "Cerebral activation patterns of speech perception in 4-months-olds and adults" the effects of infant-directed speech on ERP lateralization in 4-month old infants and adults. Infants tended to discriminate between Swedish, manipulated Swedish and Portuguese.

In "Rapid, automatic and parallel language processing in the brain", Yury Shtyrov and Friedemann Pulvermüller provide an overview of language processing studies by means of MEG investigating the Mismatch Negativity (MMN). They demonstrate that size and topography of the MMN reflect the activation of memory traces for language elements in the human brain even under unattentional conditions.

Bengt Sigurd and Bengt Nilsson present in their contribution "Modelling brain activity understanding center-embedded sentences" two computer programs modeling different parsing strategies of complex sentences. In relation to Eye-Tracking data, they further discuss how working memory and attention might constrain the syntactic processing.

In their chapter "Hemifield asymmetries in parafoveal word processing", Jaana Simola, Kenneth Holmqvist and Magnus Lindgren review different approaches to the right visual field advantage in processing text that is not in current visual focus. They also discuss results from a combined Eye-Tracking and EEG study supporting a perceptual learning account, suggesting that the asymmetry is a result of long-term reading from left to right.

Clinical aspects

Stefan Heim, Julia Tschierse, Katrin Amunts, Marcus Wilms, Simone Vossel, Klaus Willmes, Anna Grabowska, Elisabeth Bay, Marion Grande, and Walter Huber demonstrate in "Cognitive subtypes of dyslexia" that dyslexic children can be classified into different groups characterized by distinct patterns of cognitive deficits. These findings illuminate the heterogeneity of results in the research literature concerning cognitive impairments in dyslexia.

In his chapter "Speech perception and brain function: Experimental and clinical studies", Kenneth Hugdahl gives a thorough review of speech perception during dichotic listening. These findings are extended to include children at risk for dyslexia.

In "Linguistic analysis of spontaneous language in aphasia", Elisabeth Meffert, Katja Hussmann, and Marion Grande review the literature on the analysis of spontaneous language in aphasia research and clinical practice. They point out some general problems of analysis, and introduce a computer-assisted instrument which offers solutions to some of these problems.

Lexicon

Auditory repetition of words and pseudowords: An fMRI study[1]

Sarah Bihler, Dorothee Saur,
Dorothee Kümmerer, Cornelius Weiller
Department of Neurology, University Hospital Freiburg

Stefanie Abel, Walter Huber
Neurolinguistics at the Department of Neurology, University Hospital RWTH Aachen

Jürgen Dittmann
Institute of German Language, University of Freiburg

Abstract

Cognitive routes involved in repetition and especially the role of semantic processing are not yet fully understood. A further open question is the impact of a new phoneme sequence on the repetition process. We therefore intended to investigate cognitive states by measuring neural correlates with functional imaging. An auditory repetition task was used: healthy subjects heard single words and pseudowords in the MRI scanner and repeated them overtly. As a result, repetition of both words and pseudowords activated bilateral auditory and motor areas as well as the left BA 44 which is responsible for phonological processing. Pseudowords showed a clear left hemispheric activation network associated with phonological processing and articulatory planning. Hence, these processes are affected by the new phoneme sequence and a dominant use of the nonlexical pathway in pseudoword repetition is obvious. Words specifically revealed areas associated with semantic processing, indicating that the lexical-semantic route is dominant in real word repetition.

[1] Full author and affiliation details appear in Appendix.

1. Introduction

Phonetic and phonological de- and encoding processes are known to be relevant for overt word and pseudoword repetition (cf. Indefrey & Levelt, 2004). The involvement of a semantic processing stage, however, is still a matter of debate. In repetition, two processing routes are discussed which can be used together or alone: a lexical-semantic route which contains access to lexical entries, versus a nonlexical route of repetition that is based on acoustic-phonemic conversion and does not involve lexical processing (Dell et al., 2007; Hanley et al., 2004).

It is an unresolved question to which extent the routes are used during word and pseudoword repetition. In contrast to real words, pseudowords have no semantic content and no lexeme entry in the mental lexicon. Therefore, they must be predominantly processed via a nonlexical repetition route. Hence, the engagement of nonlexical processing is likely to be greater in pseudoword than in real word repetition. There is a wide range of patient studies, however, that indicate an additional participation of the lexical-semantic route in pseudoword repetition (e.g. Gathercole & Baddeley, 1990; Glosser et al., 1998; Hanley et al., 2002; Saito et al., 2003). Real words possess semantic and phonological representations in the mental lexicon (lexical concepts and lexemes). Therefore, a repetition via a lexical-semantic route is likely. However, it is still unsettled as to which route is dominantly used (cf. Hanley et al., 2004) and whether repetition of real words actually involves semantic processing.

Furthermore, pseudowords have an unfamiliar phoneme sequence which should have an impact on the neuronal resources necessary for cognitive processing. It is unclear to which extent the new phoneme sequence influences pseudoword repetition. We hypothesize that both phonological en- and decoding processes are affected by the unfamiliar phoneme sequence.

Cognitive processes as well as cognitive routes in repetition may be elucidated by analyses of neural correlates. We therefore examined neural regions that are specific for word and pseudoword repetition. In real word repetition, a strong activation of regions responsible for semantic processing would emphasize the dominance of a lexical-semantic route compared to pseudoword repetition. In addition, we hypothesize that the new phoneme sequence of a pseudoword has an impact on the cognitive processing of pseudowords. This is clarified by contrasting pseudoword repetition with word repetition. High demands on phonological processing areas in pseudoword repetition could be a possible outcome. Likewise, these neural activations should reveal the dominant cognitive route in pseudoword repetition.

2. Materials and methods

Thirty-four native German speakers (18 male, 16 female; median age: 27.5 years; age range: 18–69 years) without any neurological, psychiatric, or hearing deficit participated in the study. We included 19 right-handed and 15 left-handed participants.

During scanning in a Siemens 3 Tesla MR scanner 60 German single words (e.g. *blume* 'flower', *praline* 'praline') and 60 pseudowords (e.g. *kliehe*, *freimelei*), each set including 30 disyllabic and 30 trisyllabic stimuli, were presented via headphones, and the subjects were instructed to overtly repeat them immediately. Hence, four conditions were measured: disyllabic real words, trisyllabic real words, disyllabic pseudowords and trisyllabic pseudowords. The stimuli were balanced for length, syllable structure, and stress and frequency of the stressed syllable; the real words were also balanced for word frequency, imageability, and semantic field. An event-related design with two sessions was used with a trial duration of 8 s.

T1-weighted anatomical images were acquired using MP-RAGE with 160 sagittal slices, 1x1x1 mm^3 voxel size, TR/TE of 2200 ms/2.15 ms, flip angle of 12° and a matrix size of 256x256 $pixel^2$. Echoplanar images were acquired with 30 axial slices, 3x3x3.6 mm^3 voxel size, TR/TE of 1600 ms/30 ms, flip angle of 70° and a matrix size of 64x64 $pixel^2$.

Using SPM5 (`http://www.fil.ion.ucl.ac.uk/spm/`) the images were adjusted for slice-acquisition timing, co-registered, segmented, normalized, spatially smoothed using a Gaussian kernel with a FWHM of 9 mm. Stimulus-offset was modelled as response-onset in the GLM. One-sample *t*-tests were performed for the main effects (words > rest, pseudowords > rest), flexible-factorial ANOVAs for the specific effects (pseudowords > words, words > pseudowords) and a two-sample *t*-test for handedness (left-handed > right-handed).

3. Results

We observed a difference between the neural processing of pseudowords and real words during repetition. There was no effect of length: the number of syllables neither of real words nor of pseudowords had an effect on the neural processing. We revealed only a small effect of handedness in the right putamen ($p < .001$, uncorrected, $t = 3.37$) when contrasting left-handed with right-handed subjects in both word and pseudoword repetition. There were no significant activations for right-handed contrasted with left-handed subjects. Hence, disyllabic and trisyllabic stimuli and the data of right- and left-handed subjects were analyzed together.

Compared to the resting state, repetition of both words and pseudowords ($p < .005$, FWE-corrected) activated bilateral primary and secondary auditory areas, bilateral motor areas, SMA, the mid cingulate gyrus, cerebellum, basal ganglia, thalamus, and other subcortical regions (Figure 1a). Using a less conservative level of significance ($p < .05$, FWE-corrected) we additionally found activation in left inferior frontal gyrus (BA 44).

Contrasted with words, pseudowords activated clear left hemispheric regions ($p < .001$, uncorrected), including left pars opercularis of the inferior frontal gyrus (BA 44) and the ventral premotor cortex, the left pars triangularis of the inferior frontal gyrus (BA 45) and the anterior insula, as well as the left posterior superior temporal

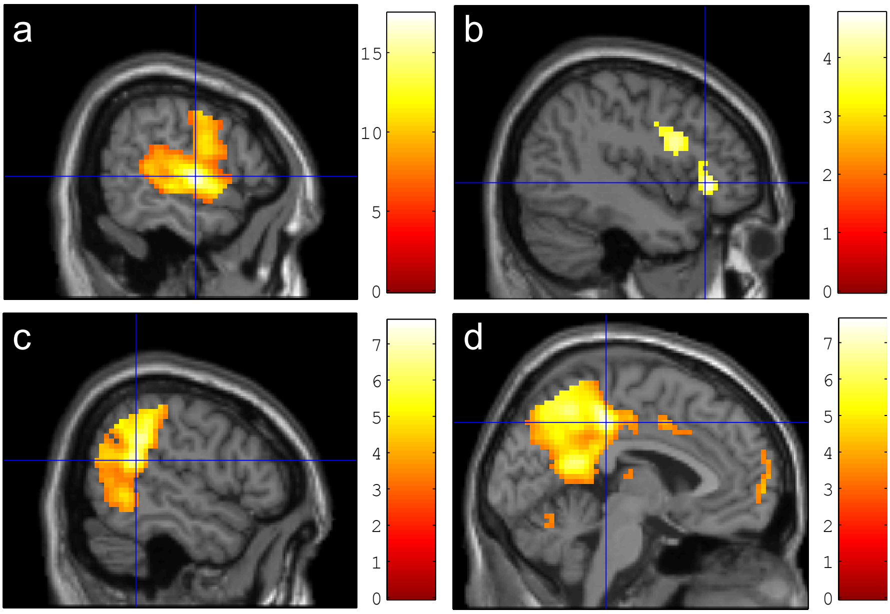


Figure 1. Functional activation maps overlaid on average MP-RAGE: (a) word repetition vs. rest ($p < .005$, FWE-corrected); (b) pseudoword repetition vs. word repetition ($p < .001$, uncorrected); (c–d) word repetition vs. pseudoword repetition ($p < .001$, uncorrected).

sulcus, the left anterior superior temporal gyrus and the SMA (Figure 1b).

Activation for words contrasted with pseudowords was stronger than the other way around. The parietotemporal cortex was bilaterally activated (Figure 1c) including supramarginal gyrus and angular gyrus, the bilateral mesial cortex (Figure 1d) including the precuneus, posterior and mid cingulate gyrus, the rostral and dorsolateral prefrontal cortex including superior and middle frontal gyrus as well as lateral and medial parts of the frontal pole, the left cerebellum as well as bilateral anterior temporal areas, posterior insula, thalamus, and brainstem. A good portion of the activation shown for words is due to deactivation for pseudowords. Activation for words can be found in the right posterior STG and supramarginal area, activations for both words and pseudowords in the bilateral posterior insula.

4. Discussion

Repetition of words and pseudowords activated neural areas which are necessary for auditory and motor processing. Thus, our investigation of cognitive processes by means of overt speech in the MR scanner was successful. The activation of the middle cingulate gyrus can be explained with the translation of intention to action. The region is probably responsible for the voluntary initiation of speech movements. Activation of the left inferior frontal gyrus (BA 44) which is a part of Broca's area and responsible for phonological processing (e.g. Gitelman et al., 2005; Heim, 2005) could also be shown. The activation of BA 44 was much more prominent at the individual level, which is in line with the fact that Broca's area is inter- and intraindividually very variable (cf. also Burton et al., 2001).

Contrasting pseudoword repetition with word repetition shows a clear left hemispheric activation of brain areas associated with phonological processing and articulatory planning. The anterior STG activation demonstrates phonological decoding processes during speech perception, the posterior STS activation is both a correlate of phonological de- and encoding, and the inferiofrontal activation is a correlate of enhanced phonological encoding processes. Thus, pseudowords unlike real words require more neuronal resources to segment the unfamiliar phoneme sequence during perception and for phonological encoding during word production (cf. Castro-Caldas et al., 1998), because there is no way of accessing a represented lexeme and a new phoneme sequence has to be produced. Furthermore, the new phoneme sequence has an impact on articulatory planning and motor control which is consistent with the findings of Klein et al. (2006) and is illustrated by left hemispheric activations of the SMA, inferior premotor regions and the anterior insula. One may conclude that phonetic encoding also incorporates planning of the phoneme sequence as a whole, not only planning of single phonemes and small phoneme combinations. We could not find neural correlates of lexical-semantic processing or semantic associations. This implies that pseudoword repetition requires a greater engagement of the nonlexical route than real word repetition, which is underpinned by our findings of enhanced phonological de- and encoding processes and enhanced processes of articulatory planning and motor control.

In contrasting real word with pseudoword repetition we revealed a differential increase of neural activity in bilateral areas that are associated with semantic processing: temporoparietal, dorsolateral and rostral prefrontal, as well as anterior temporal areas. These areas and the bilateral precuneus/posterior cingulate region showed deactivations for both conditions, though more deactivations for pseudoword repetition, and are located in the so-called "default-mode network" (Raichle et al., 2001). This network is activated during the resting state, and deactivated during performance of specific tasks. It is therefore thought to reflect brain functions which are ongoing and intrinsic as opposed to transient and evoked brain functions. According to Binder et al. (1999), this network can be seen as a "conceptual processing network" which is active in direct comparisons between semantic and perceptual processing tasks. We conclude that there is less (maybe no) semantic processing in pseudoword repetition than in real word repetition. Hence, the involvement of a lexical-semantic route in word repetition is indeed stronger than in pseudoword repetition. We found additional activations for real words in contrast to pseudowords in the posterior STG and posterior insula. The posterior STG is considered to be responsible for implicit semantic retrieval (Ohyama et al., 1996), which implies that a difference between explicit and implicit semantic processing in repetition has to be assumed. This is supported by our finding of activation in the posterior insula, which is active when speech tasks get more and more automated or real words are repeated under conditions that minimize semantic processing (Castro-Caldas et al., 1998). We conclude that both explicit and implicit semantic processes take place during real word repetition, which emphasizes our assumption of dominance of the lexical-semantic processing route in real word repetition.

To sum up, our data indicates that the production of a new phoneme sequence in pseudoword repetition requires great efforts on phonological processing, articulatory planning and motor control. In real word repetition, both explicit and implicit semantic processing have to be assumed. We could reveal the dominant processing routes for word and pseudoword repetition with fMRI: the engagement of a nonlexical route is stronger in pseudoword repetition than in repetition of real words, and the engagement of a lexical-semantic route is stronger in word repetition compared to pseudoword repetition. Furthermore, overt speech in the MR scanner is possible.

5. Acknowledgements

Hansjörg Mast provided technical assistance. Leo Schilbach advised helpfully.

References

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M. & Cox, R. W. (1999). Conceptual processing during the conscious resting state: A functional MRI study. *Journal of Cognitive Neuroscience, 11*(1), 80–93.

Burton, M. W., Noll, D. C. & Small, S. L. (2001). The anatomy of auditory word processing: Individual variability. *Brain and Language, 77*(1), 119–131.

Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S. & Ingvar, M. (1998). The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain. *Brain, 121*(6), 1053–1063.

Dell, G. S., Martin, N. & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56*(4), 490–520.

Gathercole, S. E. & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*(4), 439–454.

Gitelman, D. R., Nobre, A. C., Sonty, S., Parrish, T. B. & Mesulam, M. M. (2005). Language network specializations: An analysis with parallel task designs and functional magnetic resonance imaging. *Neuroimage, 26*(4), 975–985.

Glosser, G., Friedman, R. B., Kohn, S. E., Sands, L. & Grugan, P. (1998). Cognitive mechanisms for processing nonwords: Evidence from Alzheimer's disease. *Brain and Language, 63*(1), 32–49.

Hanley, J. R., Dell, G. S., Kay, J. & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology, 21*(2–4), 147–158.

Hanley, J. R., Kay, J. & Edwards, M. (2002). Imageability effects, phonological errors, and the relationship between auditory repetition and picture naming: Implications for models of auditory repetition. *Cognitive Neuropsychology, 19*(3), 193–206.
Heim, S. (2005). The structure and dynamics of normal language processing: Insights from neuroimaging. *Acta Neurobiologiae Experimentalis, 65*(1), 95–116.
Indefrey, P. & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144.
Klein, D., Watkins, K. E., Zatorre, R. J. & Milner, B. (2006). Word and nonword repetition in bilingual subjects: A PET study. *Human Brain Mapping, 27*(2), 153–161.
Ohyama, M., Senda, M., Kitamura, S., Ishii, K., Mishina, M. & Terashi, A. (1996). Role of the nondominant hemisphere and undamaged area during word repetition in poststroke aphasics. A PET Activation Study. *Stroke, 27*(5), 897–903.
Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A. & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676–682.
Saito, A., Yoshimura, T., Itakura, T. & Lambon Ralph, M. A. (2003). Demonstrating a wordlikeness effect on nonword repetition performance in a conduction aphasic patient. *Brain and Language, 85*(2), 222–230.

Appendix

The full author list including affiliations of this paper is as follows:

Sarah M. E. Bihler[1,2], Dorothee Saur[1], Stefanie Abel[1,3], Dorothee Kümmerer[1], Walter Huber[3], Jürgen Dittmann[2], Cornelius Weiller[1]

[1]Dept. of Neurology, University Hospital Freiburg, Germany
[2]Institute of German Language, University of Freiburg, Germany
[3]Neurolinguistics, Dept. of Neurology, University Hospital RWTH Aachen, Germany

Evaluation of lexical and semantic features for English emotion words

Francesca Citron, Brendan Weekes, Evelyn Ferstl
Department of Psychology, University of Sussex

Abstract

A general consensus has been achieved in emotion research on the two-dimensional structure of affect, constituted by *emotional valence* and *arousal* (Feldman Barrett & Russell, 1999; Russell, 2003). However, two issues are still in debate. First, are the two dimensions intrinsically associated, or are they independent from one another? And second, does emotional valence represent a bipolar continuum of affect (positive versus negative), or are positive and negative valences two independent axes? In order to address these issues, we let 82 English native speakers rate 300 words. Emotional variables were valence and arousal. Lexical/semantic variables were age of acquisition, familiarity, and imageability. Different patterns of correlations with the lexical variables were found for emotional valence and for arousal, supporting a two-dimensional structure of affect and suggesting independence between emotional valence and arousal, with arousal not being reducible to the intensity of valence. Furthermore, different levels of arousal for positive and negative words were found, suggesting independence between positive and negative affects, instead of being opposites of a bipolar continuum.

1. Introduction

In emotion research there seems to be a general consensus on the two-dimensional structure of affect: *valence* describes the extent to which an affect is pleasant or unpleasant (positive, negative), whereas *arousal* refers to the intensity of an affect, how exciting or calming it is (Feldman Barrett & Russell, 1999; Russell, 2003). Arousal has often been defined as the intensity of a valenced affective response, in fact emotionally valenced words are typically more arousing than neutral words (Bradley & Lang, 1999; Võ et al., 2006). Empirical research often considers arousal as intrinsically associated with emotional valence (Scott et al., 2009; Kanske & Kotz, 2007; Kissler et al., 2007), although theoretical research suggests independence of valence and arousal (Reisenzein, 1994). Furthermore, no consensus has yet been achieved on the relation between positive and negative affect, which can be independent or bipolar opposites (Feldman Barrett & Russell, 1999).

Support for a two-dimensional structure of affect comes from theoretical models of emotion (Russell, 1980, 2003; Reisenzein, 1994), behavioural and physiological research (Lewis et al., 2007; Anders et al., 2004; Kersinger & Corkin, 2004; Kersinger & Shacter, 2006), as well as linguistic evidence (Russell, 1991). Emotion language can play a role in creating conscious emotional experience through the application of language to affective feelings (Feldman Barrett & Russell, 1999).

In particular, emotional content of verbal stimuli influences cognitive processing, as revealed by ERPs, fMRI, and behavioural performance, during lexical decision tasks (Scott et al., 2009; Kanske & Kotz, 2007; Kuchinke et al., 2005), valence decision tasks (Võ et al., 2006), reading (Kissler et al., 2007; Herbert et al., in press), self-referential tasks (Lewis et al., 2006), oddball tasks (Delplanque et al., 2004) and memory tasks (Maratos et al., 2000).

Beyond emotional valence and arousal ("warm" features), other lexical/semantic features ("cold") have been shown to have an influence on single word processing. These include word frequency (Scott et al., 2009; Kuchinke et al., 2007), familiarity, age of acquisition (Juhasz, 2005), imageability, and concreteness (Kanske & Kotz, 2007).

Table 1 summarizes these features. Spoken and written word frequency refers to the frequency of occurrence of a word in a large corpus of words and can be determined by using corpora such as CELEX or BNC; familiarity is the subjective frequency of exposure to a word; and age of acquisition represents the age at which a word was learnt. These features are similar and typically show high correlations; in fact, frequent words are usually also familiar and are likely to have been acquired relatively early in life (e.g. *cup*) (Morrison et al., 1997; Stadthagen-Gonzales & Davies, 2006). Imageability represents the ease with which a word evokes a sensory mental image, whereas concreteness represents the ability to experience something in a sensory modality. High correlations between these two features were also shown (Paivio et al., 1968; Altarriba et al., 2004), as well as correlations between imageability and age of acquisition (Morrison et al., 1997; Bird et al., 2001), which suggest that highly imageable words are acquired earlier than less imageable items.

Table 1. Lexical, semantic, and affective features.

Feature	Definition	Example
Word length	In letters, syllables, or phonemes; can be determined by looking at the surface structure.	strawberry > tree
Frequency of use	The frequency of exposure to a word; depends on the relationship to a larger corpus of words; can be determined by using corpora (such as CELEX or BNC).	stuff > flattery
Emotional valence	The ability of a word to evoke positive or negative feelings, as well as neutral ones.	flower / prison / chair
Arousal	The intensity of a word, how exciting or agitating it is.	tornado > rain
Concreteness	The ability to experience something in a sensory modality.	pencil > faith
Imageability	The ease with which a word evokes a sensory mental image.	fingernail > permission
Familiarity	The subjective frequency of exposure to a word.	joke > riddle
Age of acquisition	The age at which a word was learnt.	apple > avocado

The aim of the present study was to assess the relationship between emotional valence and arousal and the independence or bipolarity of positive and negative valence. In a within-subject design, 300 words were rated for emotional valence, arousal, familiarity, age of acquisition and imageability. Correlations between all the features were calculated and interpreted.

The first hypothesis was that arousal is not reducible to the intensity of emotional valence. This hypothesis predicts differential patterns of correlations of emotional valence and of arousal with the other "cold" features. Moreover, a non-perfect correlation between arousal and absolute valence (non-identity) is expected.

The second hypothesis concerns the independence of positive and negative valence. If they are not simply bipolar opposites of an affect continuum, they should differ not only with respect to their valence ratings, but also with respect to other features such as arousal. Therefore, different levels of arousal are predicted for words with positive valence compared to words with negative valence.

2. Methods

2.1. Participants

Eighty-two psychology students from the University of Sussex (71 women, 11 men; mean age: 20.5 years; age range: 18–42 years; *SD* = 3.98) took part in the experiment. They were all native speakers of English. They received course credits for their participation.

2.2. Materials

For the experiment, 300 words were collected. Words from a translation of the BAWL (Võ et al., 2006) and from the Compass DeRose guide to emotion words (DeRose, 2005) were supplemented by additional words so that approximately one third of the items were of neutral, positive, and negative valence. Of the 300 words, 86 were emotion words (e.g. *cheerful*). Word length in letters, syllables, phonemes, and frequency of use for all the words were obtained using the web-based CELEX database (Max Planck Institute for Psycholinguistics, 2001). Concreteness was determined by classifying the words as either concrete or abstract referring to Paivio's definition of concreteness (Paivio et al., 1968).

2.3. Rating procedure

Capitalized feature names will be used when referring to empirical measures. Online questionnaires were created using the software Macromedia Dreamweaver MX 2004. For each of the five features, instructions including several examples for words high and low on the particular feature scale were given. Then, the 300 words were presented one at a time, at the centre of the page, with the 7-point scale immediately below it. The extremes were labelled as follows: the scale for Emotional Valence ranged from –3 (very negative) to +3 (very positive), Arousal, Familiarity, and Imageability were scaled from 1 (not at all) to 7 (very high), and for Age of Acquisition the 7-point scale was labelled with the following age ranges: 0–2, 2–4, 4–6, 6–9, 9–12, 12–16, older than 16. At the right end of each scale an additional bullet allowed the participants to indicate if they did not know the word at all. When all 300 words were rated for one feature, instructions for the next feature rating appeared. The order of the five features and order of the words within each rating task were varied for each participant.

2.4. Data analysis

Means and standard deviations were calculated for each feature and each word. In addition, a categorical variable "Valence Category" was created. Words with a rating of +3 to +1 were labelled positive, words with a rating from +0.8 to –0.8 were labelled neutral, and words with a rating of –1 to –3 were labelled negative. We then created a dummy variable called "Absolute Valence" to obtain low and high emotionality ratings independent of valence (positive/negative).

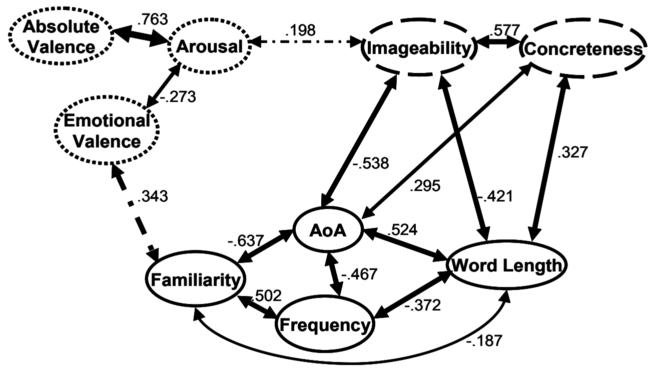


Figure 1. Spearman correlations (ϱ) between "warm" and "cold" features ($p < .001$). The arrow thickness refers to the strength of the correlations. Ovals with continuous lines contain lexical features, ovals with dashed lines contain semantic features, and ovals with dotted lines contain affective features. Correlations among cold features and among warm ones are continuous, whereas correlations between cold and warm features are dashed. Word Length is reported only in number of phonemes for ease of representation. AoA: Age of Acquisition.

Spearman correlations (ϱ) were calculated between lexical/semantic features: emotional valence, absolute valence, arousal, familiarity, age of acquisition, imageability, concreteness, frequency of use, and word length (number of letters, syllables, and phonemes). A conservative significance level of 99.9% was applied throughout.

3. Results

3.1. Reliability analysis

A reliability analysis was carried out comparing ratings of a subset of words from the current corpus with ratings from other corpora. The high correlations for both "warm" and "cold" features (Figure 1) confirm high reliability of our ratings with the previous ones.

The ANEW (Bradley & Lang, 1999) contains valence and arousal ratings for 113 out of the 300 words used in the current study. Spearman correlations between the ANEW ratings and the present items were highly significant (Valence: $\varrho = .864$, $p < .001$; Arousal: $\varrho = .703$, $p < .001$).

The MRC Psycholinguistic database (Wilson, 1988) contains familiarity and imageability ratings for 181 out of the 300 words used, and age of acquisition ratings for 72 words. Spearman correlations between the MRC ratings and the present items were highly significant (Fam: $\varrho = .788$, $p < .001$; Imag: $\varrho = .908$, $p < .001$; AoA: $\varrho = .92$, $p < .001$). Finally, the Bristol Norms (Stadthagen-Gonzales & Davies, 2006) contain familiarity, imageability, and age of acquisition ratings for 53 out of the 300 words used and Spearman correlations were even higher than the previous ones (Fam: ϱ = .891, $p < .001$; Imag: ϱ = .944, $p < .001$; AoA: ϱ = .958, $p < .001$), probably because the latter corpus contains more recent ratings.

3.2. Emotional Valence and Arousal

A highly positive correlation between Absolute Valence and Arousal was found (ϱ = .763, $p < .001$), with highly valenced words being highly arousing. A negative correlation between Emotional Valence and Arousal was also found (ϱ = –0.273, $p < .001$), suggesting that highly negative words were more arousing than highly positive words (Figure 2). A Mann-Whitney test revealed that positive and negative words significantly differed in Arousal (U = 1995.5, $p < .001$), whereas they did not differ in Absolute Valence.

Emotional Valence was positively correlated with Familiarity (ϱ = .343, $p < .001$), and negatively correlated with Age of Acquisition (ϱ = –0.169, $p < .01$).

A positive correlation between Imageability and Arousal (ϱ = .198, $p < .001$) suggests that highly imageable words were more arousing than less imageable words. However, this trend seems to characterize only emotionally valenced words, not neutral items.

3.3. Correlations between "cold" features

A highly negative correlation between Familiarity and Age of Acquisition was found (ϱ = –0.637, $p < .001$), as well as a highly positive correlation between Familiarity and Frequency of use (ϱ = .502, $p < .001$) and a negative correlation between Frequency of use and Age of Acquisition (ϱ = –0.467, $p < .001$).

A significantly negative point-biserial correlation between Imageability and Concreteness was found (r_{pb} = –0.577, $p < .001$). Concreteness correlates also with AoA (r_{pb} = –0.295, $p < .001$) and Word length in phonemes (r_{pb} = .327, $p < .001$).

A highly negative correlation between Imageability and Age of Acquisition (AoA) was found (ϱ = –0.538, $p < .001$).

Word Length in letters, syllables, and phonemes showed highly positive correlations with one another (ϱ > .79, $p < .001$). In order to be concise only the correlations of Number of phonemes are shown (see Figure 1). This variable negatively correlates with Familiarity (ϱ = –0.187, $p < .001$), Imageability (ϱ = –0.421, $p < .001$) and Frequency of use (ϱ = –0.372, $p < .001$) and positively correlates with AoA (ϱ = 0.524, $p < .001$).

4. Discussion

The correlations found between familiarity, age of acquisition and frequency of use are in line with previous findings (Morrison et al., 1997; Stadthagen-Gonzales et al., 2006), suggesting that words which are familiar are also frequently used and are acquired early. Imageability and concreteness also correlated as measures of similar features (Paivio et

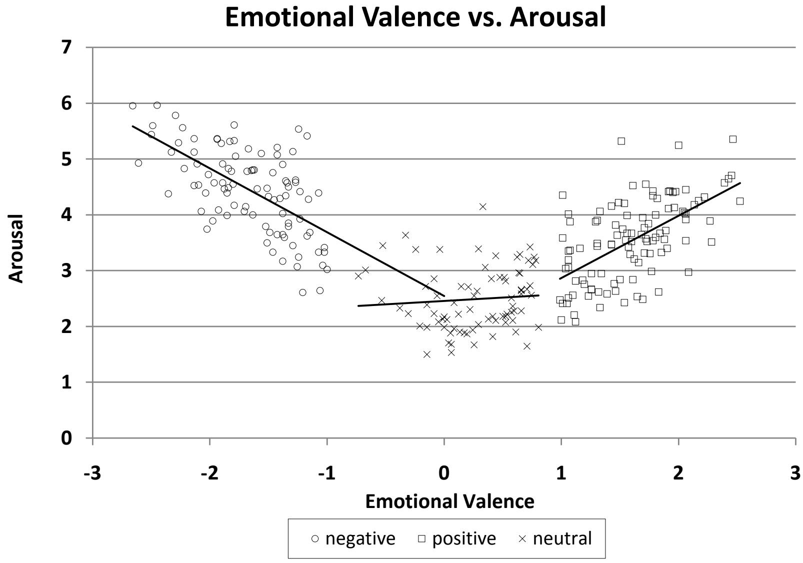


Figure 2. Emotional Valence ratings plotted with Arousal ratings. Valence ratings are categorized as positive, neutral, and negative.

al., 1968). Interestingly, imageability correlated with age of acquisition, suggesting that highly imageable words were acquired earlier than less imageable words (Morrison et al., 1997; Bird et al., 2001). Finally, word length in phonemes correlated with familiarity, imageability, frequency of use, and age of acquisition, suggesting that shorter words are more familiar, more imageable, more frequently used and earlier acquired than longer words (Morrison et al., 1997; Stadthagen-Gonzales et al., 2006).

Our first hypothesis of an independent relationship between emotional valence and arousal, i.e. a non identification of arousal as the intensity of emotional valence, was supported by our results. Firstly, there was a highly positive but not perfect correlation between absolute valence and arousal (non-identity). Secondly, positive and negative words differed in levels of arousal but not of absolute valence. Thirdly, we obtained different patterns of correlations of valence and of arousal with the other cold features. In fact, emotional valence correlated with familiarity and age of acquisition, whereas arousal correlated with imageability.

It is suggested here that positive words are first acquired and hence are also more familiar. Correlations between warm features and cold ones have not been investigated so far. However, a high correlation between imageability and age of acquisition was found by Bird et al. (2001), who suggested a relationship between the order of imageability across word classes (nouns > verbs > function words) together with the order of acquisition of the words (nouns are usually acquired earlier than verbs and function words). A "response bias" could also account for these results, with participants being reluctant to admit that they are very familiar with negative words. In a self-referential task, Lewis et al. (2007) presented participants with affective words and asked them to indicate whether each word could be used to describe themselves (i.e. "yes"/"no"). A tendency to respond "yes" for more positive words was found.

There are no prior data on the relationship between arousal and imageability. The present results suggest that words which easily evoke a mental image lead to higher arousal than abstract items, which in turn might facilitate the maintenance of a mental image. This relationship holds for emotionally valenced words only, not for neutral ones, reinforcing the view of arousal as a dimension of the structure of affect, together with valence (Feldman Barrett & Russell, 1999; Russell, 2003).

Our second hypothesis of an independence of positive and negative valence was also supported by the results. Positive and negative words had different levels of arousal, with negative words being more arousing than positive ones. Therefore positive and negative words with comparable levels of absolute valence are not just bipolar opposites, but rather they differ with regard to other features such as arousal, familiarity, and age of acquisition.

Further support for these conclusions comes from an interesting fMRI study by Lewis et al. (2007). They found a double dissociation in patterns of brain activation between emotional valence and arousal. This finding supported a characterization of valence in terms of independent axes but not in terms of a bipolar continuum.

The different levels of arousal found for positive and negative words could be accounted for by the selection of particular words. For example, in a rating study with the aim of validating the Velten Mood Induction Statements, Jennings et al. (2000) obtained an opposite pattern of results: higher arousal ratings for positive statements than negative ones. Although this study is not comparable with the current one because the stimuli rated are different, the contrasting results could be due to the selection of particular material. It is here suggested, though, that highly negative words are naturally more arousing than highly positive ones, as very intense positive words like ecstasy are likely to elicit positive feelings up to a certain threshold, which would then transform into negative ones.

In addition to these conclusions, the collected data will facilitate the design of psycholinguistic and neuroscientific studies of word and language comprehension. In particular, the arousal ratings will enable us to compare different arousal levels within each level of valence. Furthermore, the differentiated knowledge of "cold" features, such as age of acquisition, familiarity, and imageability will be useful for controlling cognitive influences over and above the effects of emotion processing.

5. Acknowledgements

We would like to thank Kathrin Klingebiel for her useful help in preparing the questionnaire material for the experiment and some graphic representations of the results, as well as Daniel Hyndman for his IT support.

References

Altarriba, J. & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract and concrete words. *American Journal of Psychology, 117,* 389–410.

Anders, S., Lotze, M., Erb, M., Grodd, W. & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fMRI study. *Human Brain Mapping, 23,* 200–209.

Bird, H., Franklin, S. & Howard D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33,* 73–79.

Bradley, M. M. & Lang, P. J. (1999). *Affective norms for English words (ANEW): Simuli, instruction manual and affective ratings.* Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Delplanque, S., Lavoie, M. E., Hot, P., Silvert, L. & Sequeira, H. (2004). Modulation of cognitive processing by emotional valence studied through event-related potentials in humans. *Neuroscience letters, 356,* 1–4.

DeRose, S. J. (2005). *The Compass DeRose guide to emotion words.* Retrieved November 2, 2007, `http://www.derose.net/steve/resources/emotion/`.

Feldman Barrett, L. & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science, 8,* 10–14.

Herbert, C., Junghofer, M. & Kissler, J. (2008). Event-related potentials to emotional adjectives during reading. *Psychophysiology, 45,* 487–498.

Jennings, P. D., McGinnis, D., Lovejoy, S. & Stirling, J. (2000). Valence and arousal ratings for Velten mood induction statements. *Motivation and emotion, 24,* 285–297.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin, 131,* 684–712.

Kanske, P. & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research, 1148,* 138–148.

Kersinger, E. A. & Corkin, S. Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Science* (St. Luis, MO, March 2004), 101:3310–101:3315.

Kersinger, E. A. & Shacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience, 6,* 110–126.

Kissler, J., Herbert, C., Peyk, P. & Junghofer, M. (2007). Buzzwords. Early cortical responses to emotional words during reading. *Psychological Science, 18*, 475–480.

Kuchincke, L., Võ, M. L.-H., Hofmann, M. & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology, 65,* 132–140.

Leech, G. N., Rayson, P. & Wilson, A. (2001). *Word frequencies in written and spoken English.* Harlow: Longman.

Lewis, P. A., Critchley, H. D., Rotshtein, P. & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex, 17,* 742–748.

Maratos, E. J., Allan, K. & Rugg, M. D. (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia, 38,* 1452–1465.

Max Planck Institute for Psycholinguistics (2001). *Web-based CELEX*. Retrieved November 23, 2007, `http://celex.mpi.nl/`.

Morrison, C. M., Chappell, T. D. & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50A,* 528–559.
Pavio, A., Yuille, J. C. & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 76, 1–25.
Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67,* 525–539.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172.
Russell, J. A. (1991). Culture and the categorisation of emotions. *Psychological Bulletin, 110,* 426–450.
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.
Scott, G. G., O'Donnell, P. J., Leuthold, H. & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology, 80*, 95–104.
Stadthagen-Gonzales, H. & Davis, C. (2006). The Bristol norms for age of acquisition, imageability and familiarity. *Behavior Research Methods, 38,* 598–605.
Võ, M. L.-H., Jacobs, A. M. & Conrad, M. (2006). Cross-validating the Berlin affective word list. *Behavior Research Methods, 38,* 606–609.
Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioural Research Methods, Instruments and Computers, 20,* 6–11.

Concreteness effect and word cognate status: ERPs in single word translation

Armina Janyan, Ivo Popivanov
Central and East European Center for Cognitive Science, New Bulgarian University

Elena Andonova
University of Bremen

Abstract

The study examines the extent of the involvement of semantics (concreteness) in cognate versus non-cognate processing in oral translation of visually presented single words. Behavioural results replicated the robust cognate facilitation effect and revealed no concreteness effect. However, analyses of ERPs did not show the robust cognate effect but instead indicated a topographically distinct concreteness effect in cognate processing and an absence of the effect in non-cognate processing. The differential processing of the two types of words is interpreted in terms of models of bilingual language processing and theories of semantic system organization.

1. Introduction

The present study addresses the issue of bilingual language representation and processing in translation by testing the degree of semantic/conceptual involvement in processing (the concept-mediation hypothesis, see Potter et al., 1984). We crossed a semantic variable (concreteness) and the cognate status of words in an oral translation task.

Cognates are words that have a similar form and meaning in both languages (L1 and L2) such as Bulgarian лампа [lampa] and English *lamp* whose phonological representations overlap. Non-cognates have different forms in L1 and L2 but similar meaning such as слон [slon] in Bulgarian and *elephant* in English. The general finding is that cognates are produced, recognized, and translated faster than non-cognates (e.g. Kroll & Stewart, 1994; Dijkstra et al., 1999; Costa et al., 2005). The faster production, recognition, and translation of cognates are usually attributed to a common set of form-based representations (e.g. orthographic, phonological, morphological) that are used to process them in both languages. Bilinguals may rely more on form-based features than on not form-based ones when translating cognates, that is, translation processing is restricted to the lexical levels of L1 and L2 (word association, or lexical mediation hypothesis, Potter et al., 1984; Kroll & De Groot, 1997). However, the distributed feature model (Kroll & De Groot, 1997) predicts faster cognate translation not necessarily because translation processing might be restricted to the lexical level but because cognates have a higher level of feature overlap on both the lexical and the conceptual levels. Non-cognates, on the other hand, may be processed differently depending on the level of L2 proficiency: the more proficient a person is, the more meaning representation(s) will be activated and involved in L2 word processing (see also the Revised Hierarchical Model, Kroll & Stewart, 1994). Here we aim to test the models' predictions and to examine the extent of involvement of semantics (concreteness) in cognate versus non-cognate processing.

Concrete words are typically found to have a processing advantage over abstract words in a variety of language and memory tasks (e.g. Schwanenflugel et al., 1988; Hamilton & Rajaram, 2001; Allen & Hulme, 2006). Neurophysiological studies typically find that concrete words are associated with more negative event-related potentials (ERPs) than abstract words (West & Holcomb, 2000; Zhang et al., 2006; Lee & Federmeier, in press). One of the most influential theories that tries to explain this processing difference is the dual coding theory (Paivio, 1991) which regards concrete words as encoded both verbally and non-verbally (in an image-based system) so access to concrete words is faster due to the double-system coding. On the other hand, a single-coding (verbal only) context-availability account (Schwanenflugel et al., 1988) assumes that access to concrete words is faster due to stronger/better associations to relevant contextual knowledge within a single representational system.

Previous ERP research on bilingual single-word processing has focused on a number of issues such as, for instance, the role of proficiency in the processing of semantic information (Elston-Güttler et al., 2005), code switching effects (Chauncey et al., 2008), or the processing of interlingual homographs (Kerkhofs et al., 2006). The present study tests the concept-mediation hypothesis by contrasting processing of concreteness and cognate status of single words in a translation task. Dual-coding theory (Paivio, 1991) posits that concrete words have verbal and non-verbal representations, that is, they are coded in at least two modalities while abstract words are only coded in one. This means that concrete and abstract words should have access to different cognitive and neural processing structures (Swaab et al., 2002; Zhang et al., 2006). However, the context-availability hypothesis (Schwanenflugel et al., 1988) is based on the verbal modality only, that is, concrete and abstract words should have access to the same neuronal structures. In other words, the emergence of a concreteness effect without an interaction with scalp distribution would be in favour of the context-availability hypothesis and of the lexical-mediation hypothesis in translation. The absence of a concreteness effect would also speak in favour of lexical mediation. However, a differential scalp distribution of the concreteness effect would mean differential processing (Swaab et al., 2002; Zhang et al., 2006), hence, would speak in favour of the dual-coding theory and of conceptual mediation. The same line of reasoning concerns cognate status. A differential scalp distribution would mean a functional difference between cognate and non-cognate processing.

Thus, we are interested to see at what point the concreteness effect emerges during L2 word processing in a translation task, how a word's cognate status affects the concreteness effect, and whether this effect is distributed over the scalp.

2. Method

2.1. Participants

Twenty-two native Bulgarian speakers (19 female, 3 male; age in years: *M* = 22, *SD* = 3.1) with English as a second language participated in the experiment. Eighteen of them were students of English philology, two were young English language teachers, and two were advanced bilinguals with a specialized English-language school background and sporadic translation work. All of them had normal or corrected to normal vision. All participants filled in an English (L2) language history questionnaire.

The three main self-report measures were collapsed subjective proficiency rating of speaking, reading, writing, and understanding native speakers on a 7-point scale (SubProf, with 7 the most proficient): *M* = 5.9, *SD* = 0.7; age of L2 acquisition (AoA): *M* = 8.9, *SD* = 2.3; and length in years of L2 study (Length): *M* = 12.5, *SD* = 3.9.

2.2. Stimuli and design

Stimuli for a 2 (Concreteness: Concrete vs. Abstract) x 2 (Cognate status: Cognate vs. NonCognate) design were selected from a rich database source (Clark & Paivio, 2004), 52 items per condition. Word frequency data were derived from Balota et al. (2007). The words were matched as much as possible on the most important word characteristics (Table 1).

Table 1. Means and standard deviations (in parentheses) of word stimuli characteristics for each condition. Columns Concr(eteness) and Imag(eability) refer to a 7-point subjective rating (7 as maximum); Freq(uency) refers to $\log_{10}$(HALFrequency per million + 1) derived from Balota et al. (2007); and Length refers to word length measured in number of letters.

Condition	Concr.	Imag.	Freq.	Length
Cognate, Abstract	2.6 (.5)	3.0 (.7)	1.8 (.7)	7.3 (1.6)
Cognate, Concrete	6.6 (.3)	6.2 (.5)	1.5 (.6)	7.0 (1.6)
Noncognate, Abstract	3.1 (.5)	2.3 (.6)	1.9 (.8)	7.2 (1.8)
Noncognate, Concrete	6.7 (.3)	6.3 (.4)	1.5 (.4)	6.7 (1.6)

2.3. Procedure

Four pseudorandomized lists were constructed so that no more than two consecutive trials were in the same condition. Participants were asked to translate aloud the word they saw on the screen from English into Bulgarian as quickly and accurately as possible, to avoid hesitation markers and remain silent if they did not know the translation of the stimulus word. In addition, participants were asked to try not to blink or move their eyes away from the stimulus appearance till the response. After each response an asterisk ("*") appeared for 1500 ms indicating the possibility to blink.

Each stimulus was presented only once to each participant. Each trial started with a fixation cross ("+") for a time randomly varying between 400 and 600 ms followed by a blank screen for 800 ms. Stimuli were presented in white uppercase letters (typeface Arial, size 25 pt) against a black background. The stimulus remained on the screen for 300 ms. The intertrial interval was set to be a random number between 1200 and 1500 ms; it did not include the 1500 ms blink interval. E-prime software (Schneider et al., 2002) controlled stimulus presentation and the recording of response times (RTs). A serial response box connected to a microphone recorded voice onset RT. RT was measured from the offset of the stimulus word with the response window set to 3 s. In addition, E-prime sent TTL signals to an EEG channel to mark stimulus appearance and microphone triggering.

Participants' responses were recorded by the experimenter. The practice session contained 15 items. The experiment took about 20–25 minutes.

2.4. EEG recording

The EEG was recorded by means of Ag/AgCl ring electrodes placed symmetrically over the left and right hemispheres. The set of electrodes covered frontal, temporal, parietal, and occipital areas. The electrodes were denoted respectively as F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1 and O2 (10/20 system). EOG was recorded monopolarly (from the point 2 cm dorsal to the left ektokanthion and from the right cantus 1 cm superior to the nasion) to detect both vertical and horizontal eye movements. Linked mastoids served as reference for both EEG and EOG recordings. The EEG and EOG signals were amplified using a Guger gBSamp biosignal amplifier (`http://www.gtec.at/content.htm`) with a sampling rate of 250 Hz and a frequency range of 0.5–100 Hz. The noise of 50 Hz was removed from the EEG-signal via application of a 50 Hz notch filter.

3. Results and discussion

The data of two left-handed participants were excluded from the analyses. In addition, the data of one participant with approximately 20% valid responses only were also excluded. For technical reasons, the data of one participant were lost. Thus, the reaction time and EEG analyses are based on the data of 18 right-handed bilinguals.

3.1. Reaction time

The following types of errors were identified and excluded from the data prior to analysis. Trials on which no response was registered (9.8%) and trials with technical errors (0.4%) were excluded from the analyses. In addition, trials with response names different from the target equivalent (4.3%) were discarded. Further, response times lying more than ±2 standard deviations from the mean per subject and condition were also removed (3.2% of overall responses).

Thus, a total of 82.3% of the originally collected RT data were included in further analyses. The reported results are based on both item and subject analyses. Table 2 presents item means and standard deviations for each condition in ms.

Table 2. Mean response times (in ms) and standard deviations (in parentheses) for four experimental conditions, item means.

	Cognate	Non-cognate
Concrete	821 (150)	983 (216)
Abstract	781 (160)	1047 (305)

A 2x2 ANOVA on item means obtained a significant main effect of cognate status ($F(1, 203) = 50.79$, $p < .001$) which showed that cognates were translated faster (801 ms) than non-cognates (1015 ms). There was no effect of concreteness ($p > .6$). A marginal cognate status by concreteness interaction ($F(1, 203) = 3.01$, $p = .08$) showed a large difference in cognate status ($p < .001$) and no difference in concreteness in spite of the observed tendency ($p > .1$) (cf. Table 2). A 2x2 repeated measures ANOVA on subject means replicated the cognate facilitation effect ($F(1, 17) = 43.02$, $p < .001$) and obtained no concreteness effect and interaction ($p > .2$).

To summarize, the results replicated the cognate effect (e.g. Kroll & Stewart, 1994) and showed no concreteness effect in either cognate condition. That is, the RT results supported the lexical mediation hypothesis for both word types.

3.2. Event-Related Potentials

The EEG data were screened for eye movement, EMG and other artifacts in a critical window ranging from 200 ms before to 800 ms after the word onset. In addition to behavioral and technical errors (14.5%), trials containing eye and muscle artifacts were rejected (15.3%). Thus, 70.2% of the data remained for the analysis.

For each subject, average waveforms were computed across all accepted trials per condition aligned to a 200 ms baseline. The grand average ERPs and electrode Cz separately are shown in Figure 1. Mean amplitude ERPs were calculated for the 300–500 ms latency window after the onset of the word. This window was determined by a visual inspection of grand average waveforms (Figure 1) and previous reports (e.g. West & Holcomb, 2000; Swaab et al., 2002; Zhang et al., 2006). The data were submitted to repeated measures analyses of variance with the following within-subject factors: cognate status (cognate versus non-cognate), concreteness (concrete versus abstract), topography (frontal versus central – i.e., central collapsed with temporal – versus parietal), and hemisphere (left versus right). The occipital sites were not included in the analysis. A Greenhouse-Geisser correction was applied for all the repeated-measures that had more than one degree of freedom in the numerator. Unless indicated differently, the Greenhouse-Geisser corrected p values will be reported.

The analysis of variance indicated that overall, concrete words evoked more negative deflection than abstract words ($F(1, 34) = 34.18$, $p < .001$). This result is consistent with previous research concerning the time of occurrence of the concreteness effect and its amplitude trend (West & Holcomb, 2000; Swaab et al., 2002; Zhang et al., 2006; Lee & Federmeier, 2008).

There was a marginally significant main effect of cognate status ($F(1, 34) = 3.60$, $p < .07$) and a significant main effect of topography ($F(2, 68) = 10.61$, $p < .001$). The main effect of hemisphere ($F(1, 34) = 8.57$, $p < .01$) showed that the right hemisphere produced overall more negative amplitude than the left one. More importantly, an interaction between cognate status and concreteness ($F(1, 34) = 7.32$, $p < .01$) showed that the concreteness effect emerged in the cognate conditions ($p < .05$) but not in the non-cognate ($p > .3$). The concreteness effect was maximal over centro-temporal sites ($p < .001$), present over frontal sites ($p < .01$) and absent over parietal ($p > .1$) as was evident from a significant interaction between concreteness and topography ($F(2, 68) = 5.62$, $p < .01$). Thus, a topographic difference of the concreteness effect was observed. Finally, a marginally significant three-way interaction of cognate status, concreteness and hemisphere ($F(1, 34) = 3.31$, $p < .08$) indicated that the concreteness effect in cognates was maximal over the sites of the right hemisphere ($p < .01$), present over the sites of the left hemisphere ($p < .05$), and confirmed the absence of a concreteness effect in non-cognates over sites of both hemispheres ($p > .2$). The interaction is shown in Figure 2.

The hemispheric asymmetries of the concreteness effect in cognates seen in Figure 2 are in perfect accordance with the dual coding theory (Paivio, 1991) which assumes that a larger concreteness effect should be observed in the right hemisphere because concrete words are processed both in the imagery system (mainly associated with the right hemisphere) and in the verbal system (mainly associated with the left hemisphere).

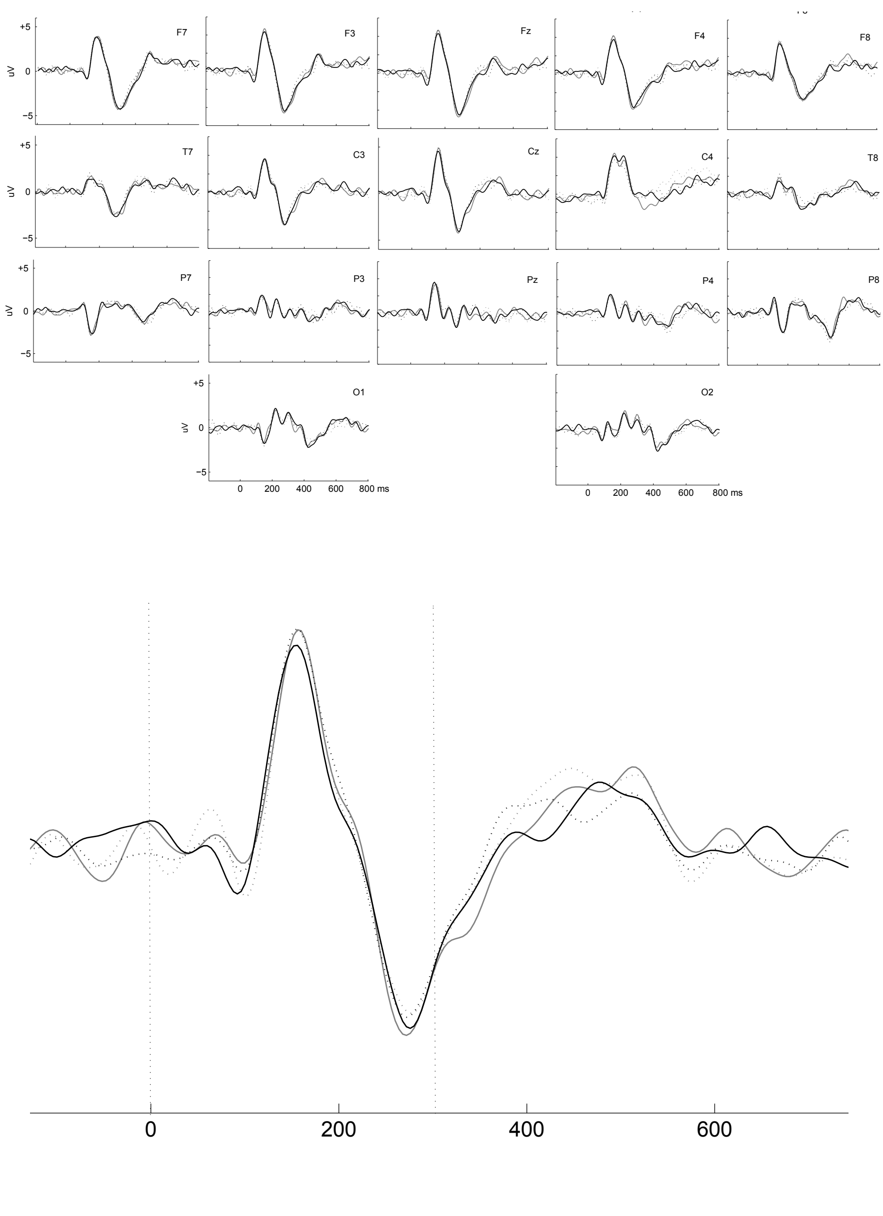


Figure 1. Grand average ERP waveforms (top) for each condition from word onset (0 ms) to 800 ms after the word onset including the 200 ms baseline. Vertical dashed lines on the Cz close-up (bottom) show stimulus presentation window (300 ms). Dashed lines represent abstract word condition, solid lines represent concrete; grey lines stand for cognates, black ones for non-cognates. Note the stable difference between abstract and concrete cognates (grey lines) and the unstable one between abstract and concrete non-cognates (black lines) in 300–500 ms post-stimulus window.

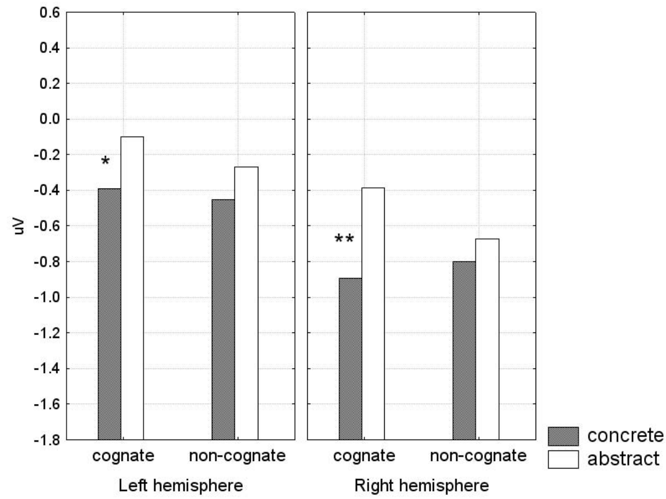


Figure 2. A concreteness by cognate by hemisphere interaction.

Abstract words, on the other hand, are processed only in the verbal system. Thus, the difference in patterns of activation appears to originate in the imagery system (right hemisphere).

Importantly, the results show no indication of a concreteness effect in non-cognate processing which unambiguously suggests non-accessibility of conceptual/semantic information during processing. To summarize, the results suggested differential processing of cognates and non-cognates, in particular, it seems that the processing of cognates was conceptually mediated, and of non-cognates lexically mediated.

4. Conclusion

The main purpose of the study was to test the concept-mediation hypothesis in translation. We manipulated the cognate status of words and their concreteness level. According to some models of the bilingual lexicon, cognates are the first candidates for the lexical mediation hypothesis in translation. Conceptual or lexical involvement in the translation of non-cognates, however, can be a matter of degree depending on the proficiency of an individual (Kroll & De Groot, 1997). Our participants were young university students learning English and not highly proficient bilinguals which may explain the results obtained concerning lexical mediation in non-cognate processing.

We combined the lexical and conceptual mediation hypotheses with the assumptions of dual coding theory and the context-availability theory assuming that different scalp distribution for the concreteness effect would have functional implications, namely, that concrete and abstract words access different neural structures. Hence, it would provide support for the dual-coding theory and the concept-mediation hypothesis. Absence of a concreteness effect or its uniform distribution over the scalp would speak in favour of the context availability model and the lexical mediation hypothesis.

The results of the behavioural data show a robust cognate facilitation effect and no concreteness effect. However, the ERP analyses added more information on processing. The results showed a differential scalp distribution of the concreteness effect in cognate processing and no concreteness effect in non-cognate processing. In other words, conceptual mediation was involved in cognate processing but was absent in non-cognate processing. To conclude, the study showed for the first time that cognate processing is conceptually mediated, and non-cognate is not, using a novel combination of hypotheses and methods. The results can be interpreted in terms of the distributed feature model (Kroll & De Groot, 1997) in the way that the higher featural overlap of cognates on a lexical level resulted in the activation of conceptual representations while the lower or absent featural overlap of non-cognates on the same lexical level restricted the processing of non-cognates to the lexical level. Importantly, the study provided support for the idea of verbal and image-based representational systems and showed that access to these systems can be restricted by cognate status of words processed and L2 proficiency level.

5. Acknowledgements

We thank our participants for their patience and Stefka Chincheva for her help in data collection. This study was supported by a research grant under the auspices of the 6th European Framework for Research and Technological Development for the project FP6-517590 "Development of human-computer monitoring and feedback tools and technologies for the purposes of studying cognition and translation".

References

Allen, R. & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55,* 64–88.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D., Simpson, G. B. & Treiman, R. (2007). The English lexicon project. *Behaviour Research Methods, 39*(3), 445–459. http://elexicon.wustl.edu/.

Chauncey, K., Grainger, J. & Holcomb, P. J. (2008). Code-switching effects in bilingual word recognition: A masked priming study with event-related potentials. *Brain and Language, 105,* 161–174.

Clark, J. M. & Paivio. A. (2004). Extensions of the Paivio, Yuille, and Madigan (1964) norms. *Behavior Research Methods, Instruments, and Computers, 36*(3), 371–383.

Costa, A., Santesteban, M. & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language, 94,* 94–103.

Dijkstra, T., Grainger, J. & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496–518.

Elston-Güttler, K. E., Paulmann, S. & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience, 17*(10), 1593–1610.
Hamilton, M. & Rajaram, S. (2001). The concreteness effect in implicit and explicit memory tests. *Journal of Memory and Language, 44,* 96–117.
Kerkhofs, R., Dijkstra, T., Chwilla, D. J. & de Bruijn E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research, 1068,* 170–183.
Kroll, J. F. & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp.169–200). Mahwah, NJ: Lawrence Erlbaum Publishers.
Kroll, J. P. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149–174.
Lee, C. & Federmeier, K. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language, 104,* 145–158.
Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287.
Potter, M. C., So, K.-F., Von Eckardt, B. & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23,* 23–38.
Schneider, W., Eschman, A. & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh: Psychology Software Tools Inc.
Schwanenflugel, P. J., Harnishfeger, K. K. & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27,* 499–520.
Swaab, T. Y., Baynes, K. & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: An ERP study. *Cognitive Brain Research, 15,* 99–103.
West, W. C. & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024–1037.
Zhang, Q., Guo, C., Ding, J. & Wang, Z. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language, 96,* 59–68.

Electrophysiological correlates of word learning in toddlers

Janne von Koss Torkildsen
Department of Biological and Medical Psychology, University of Bergen

Abstract

Towards the end of the second year most children experience a vocabulary spurt where the rate of productive word acquisition increases dramatically. However, it is unclear whether this striking development in the productive domain is accompanied by similar gains in receptive word processing abilities. In the present event-related potential (ERP) study, we compared the performance of 20-month-olds with high and low productive vocabularies in a word learning task where novel words were associated with objects. Results show that children in both vocabulary groups learned to recognize the novel word forms, but there were group differences in the efficiency of word form encoding. Moreover, only children with high vocabularies showed evidence of having formed associations between word forms and referents by displaying an N400-like incongruity effect when word-referent pairings were switched. Findings of the present study thus indicate that the rapid productive vocabulary development in the second year is related to receptive word recognition and the ability to fast map between words and referents.

1. Introduction

The speed with which young children learn the vocabulary of their native language has impressed both parents and researchers. Behavioral studies have shown that children from two years onwards are capable of fast mapping, i.e. forming an initial association between a word and its referent in just a single or a few exposures (Carey & Bartlett, 1978; Gershkoff-Stowe & Hahn, 2007; Heibeck & Markman, 1987). Traditionally, the ability to fast map has been thought to emerge around the vocabulary spurt, a period of dramatic growth in the rate of productive word acquisition which usually occurs in the second half of the second year (Benedict, 1979; Fenson et al., 1994; Goldfield & Reznick, 1990). The predominant view in language acquisition research is that the vocabulary spurt marks a qualitative shift in the way words are acquired (MacShane, 1979; Nazzi & Bertoncini, 2003): before the spurt word learning is assumed to be slow and laborious because the child uses associationist mechanisms that require a large number of co-presentations of a word and its referent. Then, during the latter half of the second year, children supposedly gain the insight that words refer to things, and this referential knowledge enables much more efficient word learning.

However, the claim that fast mapping emerges around the vocabulary spurt has recently been called into question by a number of experimental studies. These experiments have shown that infants are capable of receptive fast mapping already around age 13–15 months, about half a year before the productive word learning rate starts to accelerate (Schafer & Plunkett, 1998; Werker et al., 1998; Woodward et al., 1994). This has lead to the hypothesis that there is a receptive spurt around age one (Werker et al., 1998), and that the productive vocabulary spurt may not result from general cognitive changes affecting both comprehension and production, but rather from production specific changes such as increased articulatory control (Woodward et al., 1994) or increased motivation to talk (Ninio, 1995).

Some recent Eye-Tracking studies have provided evidence which questions the production specific account of the vocabulary spurt. Investigating children across the second year, Fernald et al. (2006) and Zangl et al. (2001) showed a significant relation between productive vocabulary size and the speed and accuracy of receptive word recognition. However, these experiments both employed stimulus words which were familiar to the children. A problem with using known words is that effects of general word processing skills and effects of experience with individual words cannot be easily separated. Even when the experimental words are rated as comprehended by parents of all participating children, there may be great differences in the *amount* of experience with specific words, and these differences may influence both speed and accuracy of word recognition. Thus, the effects of productive vocabulary size on word recognition in the above studies may have been due to increased experience with the experimental words in the high vocabulary children compared to the low vocabulary children.

The present study investigated receptive word learning in two groups of 20-month-old children: children who had attained the productive vocabulary spurt and children who had not yet reached this developmental milestone. In order to control for effects of previous linguistic experience, we studied learning of novel, rather than familiar,

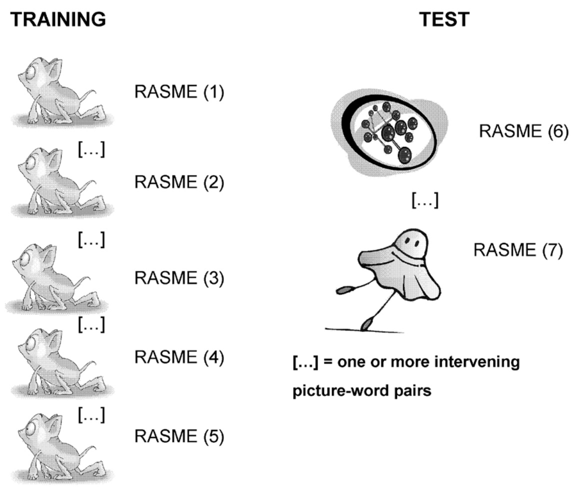


Figure 1. Outline of the experimental design. In the training phase each word was paired with its respective picture five times. In the test phase, the words were presented together with two other, but equally familiar, pictures.

words. The experimental technique was event-related potentials (ERPs) which provides information about neural activity in the brain on a millisecond level. We aimed to investigate both the dynamics of on-going learning and to assess the learning outcome after a fixed training session. A finding of nonexistent or minimal differences in receptive word learning between the two production groups would indicate that production specific advances play a significant role in the achievement of the vocabulary spurt. On the other hand, substantial group differences in receptive word learning would suggest a relation between the productive vocabulary spurt and general cognitive changes affecting both production and comprehension.

2. Methods

2.1. Stimuli and procedure

The idea behind the experimental design was to build up associations between specific novel words and pictures through repeated co-presentations, and then violate these associations by presenting the words together with different, but equally familiar pictures (Figure 1). Real words and phonotactically legal novel words, 30 of each, were used as auditory stimuli. Visual stimuli were 30 color drawings depicting referents for the real words, and 30 color drawings depicting fantasy objects and creatures which served as referents for the novel words. The real pairs were used as a reference condition and to emphasize the referential relation between words and pictures.

Stimuli were presented in 10 blocks which employed three novel and three real words each. The blocks were divided into a familiarization phase and a test phase. In the familiarization phase, each of the six words was associated with the same picture five times. The test phase of the block was an incongruity condition where words were associated with a “wrong” picture. There were two incongruity trials for each of the six words. In the incongruity trials each word was associated with two of the other pictures from the same block. There was no marker which indicated where the familiarization phase ended and the test phase of the block began.

The whole experiment lasted 24.5 minutes. Trials were 2500 ms long with an inter-trial interval of 1000 ms. The picture was displayed on the screen during the entire trial while word onset was at 1000 ms. Only event-related potentials time-locked to the auditory stimuli are reported here.

2.2. Participants

Forty-four typically developing children participated in the study. Data from an additional 34 children were excluded due to refusal to wear the electrocap ($N = 11$), technical problems during the recording ($N = 2$), or too few artefact free trials in one or more of the experimental conditions ($N = 21$). To assess vocabulary size, parents completed a Norwegian adaptation (Smith, unpublished) of the MacArthur-Bates Communicative Development Inventory (MCDI) (Fenson et al., 1993). The mean productive vocabulary on the MCDI was 117.7 words ($SD = 99.6$).

As the 50–75-word point is considered pivotal in language development and is supposed to coincide with the productive vocabulary spurt (e.g. Bates et al., 1988; Bates et al., 2002), participants were divided post hoc into a high production group and a low production group, with 75 words as a cut-off criterion. This division resulted in a high production group consisting of 25 subjects (15 girls) and a low production group consisting of 19 subjects (8 girls). Children in the high production group had a mean total productive vocabulary of 178.8 words ($SD = 91.6$), produced 23.0 ($SD = 4.8$) and comprehended 27.1 ($SD = 3.5$) of the 30 real words used in the experiment on average. Children in the low vocabulary group had a mean total productive vocabulary of 37.3 words ($SD = 22.0$), produced 4.6 ($SD = 3.5$) and comprehended 22.5 ($SD = 7.2$) of the 30 real words used in the experiment on average.

3. Word form familiarization

In this section only results for novel words will be presented. For a fuller description of statistical analyses and findings, including results for real words, see Torkildsen et al. (in press). The main finding for novel words was that the two production groups did not differ in ERPs to the initial presentation, but there were several group differences in repetition effects.

3.1. Early repetition effect

As can be seen in Figures 2 and 3, only high producers showed statistically significant effects of repetition in early time intervals (200–400 ms). This early repetition effect, which was largest at frontal sites, resembled the N200–500 effect which was found in a previous word familiarization study by Mills et al. (2005). This study employed a paradigm where words were either repeated without a referent or associated with a referent a fixed number of times. Results showed that pairing a novel word with a referent led to an increased negativity in the 200–500 ms time window. However, the opposite pattern, i.e. decreasing negativities with repetition, was seen for novel words which had merely been repeated, and for familiar words in both the paired and unpaired conditions. Thus, the authors concluded that this early negativity reflected the establishment of word meaning. This interpretation is compatible with findings from the present study as there is evidence from the test phase of the study (see Section 4) that high producers, but not low producers, managed to establish associations between novel words and pictures (Torkildsen et al., 2008). In other words, it appears that repetition rendered the novel words meaningful to the high producers, but not to the low producers.

3.2. Later repetition effect

Both high and low producers displayed statistically significant effects of repetition in later time intervals, indicating that participants in both groups achieved some recognition of the novel word forms within five repetitions (Torkildsen et al., in press). The topographical distribution of the effect, with a fronto-central dominance, was also similar for the high and the low producers. Within the experimental design of the current study, this scalp distribution, together with the peak latency around 500–600 ms, suggests that the observed effect was a negative central (Nc), one of the most studied ERP components in infants and children (Csibra et al., 2008; De Haan, 2007). As the Nc is thought to reflect attentional processes such as novelty detection, it is likely that this component is affected by repetition.

While the topographical distribution of the repetition effects was similar for the two groups of participants, the dynamic pattern of effects with the course of repetition was quite different for high and low producers (Figure 4). For high producers there was a non-linear pattern of effects (a significant quadratic contrast) with an increasing negativity from the first to the third presentation and then a decrease in negativity from the third to the fifth presentation. For low producers the negativity increased with repetition from the first throughout fifth presentation.

Since the stimulus words were novel to all participants, the differential repetition effect cannot be attributed to experience with individual words. It is more likely that the findings reflect group differences in the allocation of attentional resources to repeated presentations of stimuli. In this regard it is interesting to note a discrepancy in results from previous studies of the Nc in repetition paradigms: while two studies of the visual Nc in infants reported that this component decreased with repetition (Hill Karrer et al., 1998; Wiebe et al., 2006), another study reported that the visual Nc increased with repetition (Webb & Nelson, 2001). On the basis of these apparently conflicting observations De Haan (2007) suggested that the Nc may display non-linear amplitude changes in response to repetition. No hypothesis about the specific nature of the non-

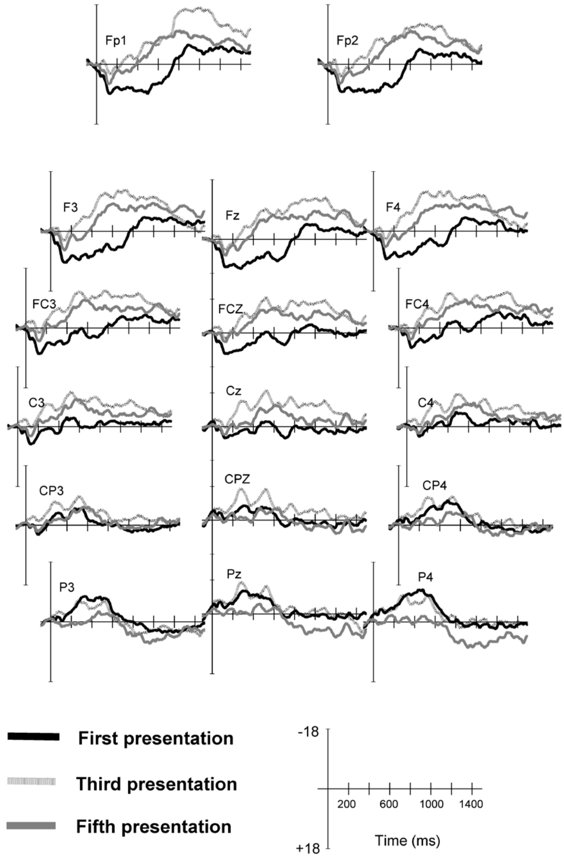


Figure 2. High production group. Novel words. Grand-average waveforms for the first, third, and fifth presentations. Negative is plotted up on this an all other grand-averages.

linear changes has been put forward, but it is relevant to observe that the two studies which found decreases in Nc amplitude with repetition collapsed ERPs to a quite large number of repetitions, while the study which found an increase in Nc amplitude with repetition compared the first and the second presentation of a stimulus. Thus, it is possible that the Nc tends to increase in amplitude until a stimulus has been encoded to such a degree that the child no longer needs to allocate its full attention to it, and then decreases gradually with further repetition. If this interpretation is correct, the results of the current study points to a difference in rate of learning between children in the two production groups. While children in the high production group may have a fairly stable encoding of the novel words after the third presentation, children in the low production group may still need to allocate substantial attentional resources to the encoding process through the fifth presentation. This interpretation is supported by the pair wise comparisons between the first presentation and repetitions of the novel words.

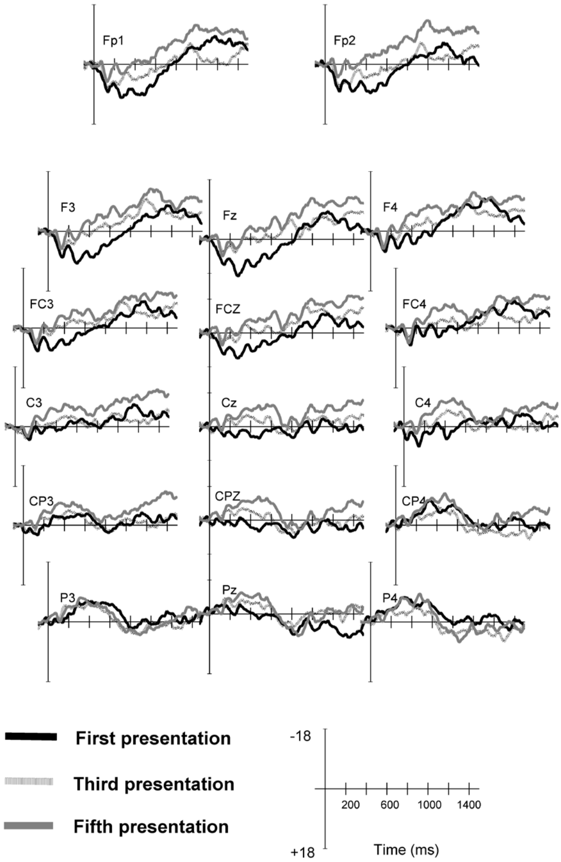


Figure 3. Low production group. Novel words. Grand-average waveforms for the first, third, and fifth presentations.

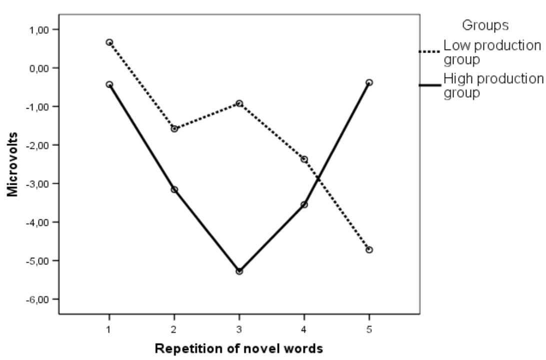
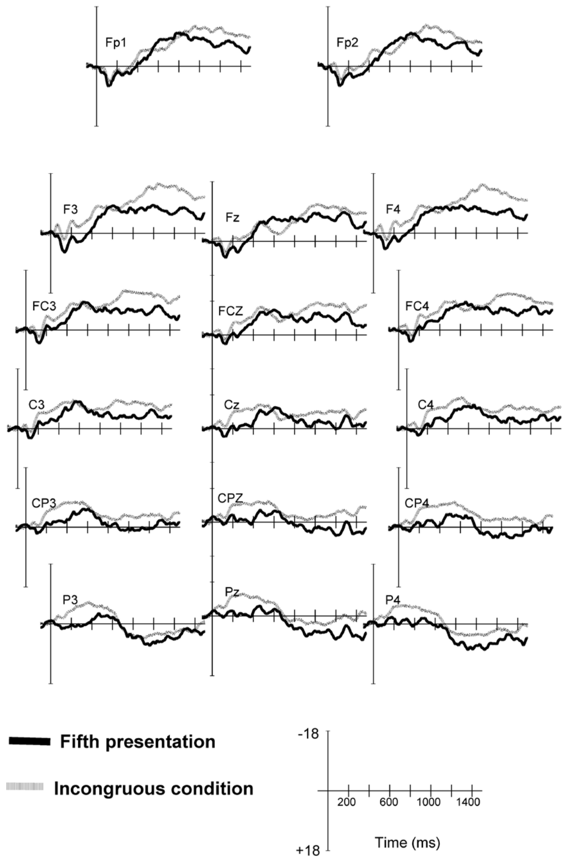


Figure 4. Novel words. Mean amplitude in the 200–400 ms interval for each of the five presentations at frontal electrodes. Negative is plotted down.

Figure 5. High production group. Novel words. Grand-average waveforms for the last presentation in the training phase and incongruity presentations where the trained association between word and picture was violated.



For the high production group, ERPs to the first presentation differed from ERPs to the third and subsequent presentations, while for the low production group only differences between the first and the fifth presentation were reliable. In other words, it appears that low producers needed two more repetitions than high producers in order to recognize the novel words.

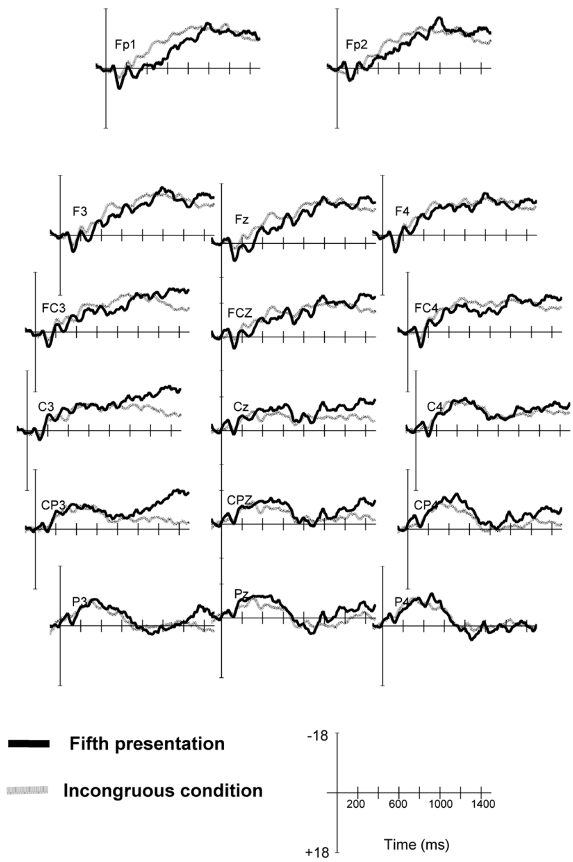
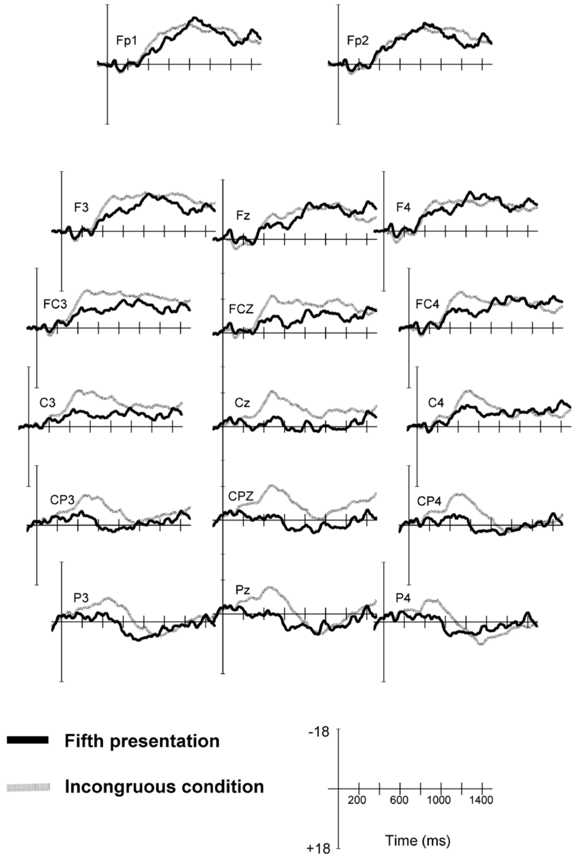


Figure 6. Low production group. Novel words. Grand-average waveforms for the last presentation in the training phase and incongruity presentations where the trained association between word and picture was violated.

4. Fast mapping

As can be seen in Figure 5, children in the high production group displayed a negativity on those trials where the trained association between novel words and pictures was broken (incongruous condition) compared to the last training trial (congruous condition) (for a more detailed description of results, see Torkildsen et al., 2008). The broad scalp distribution and the timing of this response was similar to the N400 incongruity effect which have been observed in previous studies of this age group in response to mismatches involving familiar words (Friedrich & Friederici, 2004; Mills et al., 2005; Torkildsen et al., 2006; Torkildsen et al., 2007b). This functional, temporal, and topographical similarity suggests that the observed negativity represents an N400-like effect. Children in the low production group did not show any difference between the congruous and incongruous condition (Figure 6).

Figure 7. High production group. Real words. Grand-average waveforms for the last presentation in the "training" phase and incongruity presentations.



A possible interpretation of these findings is that high producers, but not low producers, were capable of fast mapping between novel words and referents. An alternative explanation is that the low producers had made the fast mapping between words and referents, but that the mechanisms indexed by the N400 were not yet matured in this group. The latter explanation would be consistent with two previous studies of children with low vocabularies. Friedrich and Friederici (2006) retrospectively compared ERPs of 19-month-old children who were shown to have very delayed expressive language at age 30 months and children who displayed age-appropriate productive language at this age. They reported that 19-month-olds with low productive language scores 11 months later did not display an N400 response to incongruous picture-word pairs, while an age-matched control group did show an N400. As the low vocabulary children did exhibit an early lexical-phonological priming effect for congruous words, the authors

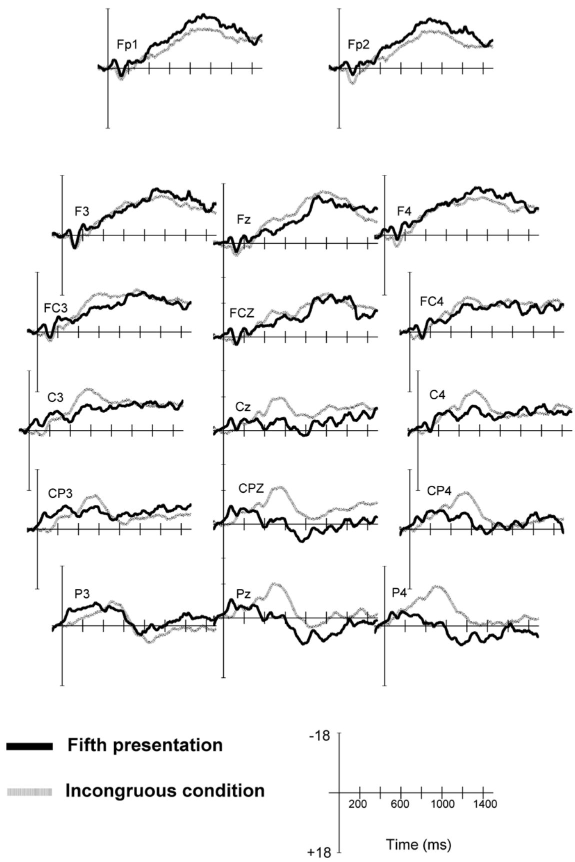


Figure 8. Low production group. Real words. Grand-average waveforms for the last presentation in the "training" phase and incongruity presentations.

argued that the lack of an N400 could not be due to missing lexical or semantic knowledge, and thus it appeared that the N400 mechanisms had not yet matured in this group. In a similar vein, Torkildsen et al. (2007a) found that picture-word mismatches elicited an N400-like incongruity effect in typically developing 20-month-olds, but not in a group of 20-month-olds at familial risk for dyslexia who also had lower productive vocabularies than the control group. Moreover, there was an enhanced early lexical priming effect in the at-risk group, which indicated that the at-risk children had acquired the vocabulary used in the test and were able to recognize correct referents for the pictures. However, an explanation in terms of lacking N400 mechanisms must be excluded in the present study, as children in both production groups did show an N400-like incongruity effect for real words which were paired with an incorrect picture (see Figures 7 and 8).

5. The productive vocabulary spurt and receptive word learning

Results of the present study indicate that there are substantial differences in receptive word learning abilities between children of the same age who have and have not reached a productive vocabulary spurt. Two major findings support this conclusion. First, 20-month-olds with high vocabularies appeared to have a faster rate of word encoding than children of the same age with small vocabularies. Second, only high producers showed evidence of having formed word-referent associations for novel words, i.e. fast mapping.

As mentioned in the introduction, a number of recent studies have found that children are able to fast map about half a year before the productive vocabulary spurt tends to take place (Schafer & Plunkett, 1998; Werker et al., 1998; Woodward et al., 1994). The latter findings have been used to suggest that receptive fast mapping skills are established already around 13–15 months of age (Werker et al., 1998), and consequently that the productive vocabulary spurt at the end of the second year may not result from general changes in language processing skills, but rather from changes specific to production, such as improved articulatory control or increased motivation to communicate (Ninio, 1995; Woodward et al., 1994). This argument has recently been strengthened by a preferential looking study which did not find a relation between fast mapping skills and productive vocabulary size in children around the age of the vocabulary spurt (Tan & Schafer, 2005).

Although 20-month-olds with low productive vocabularies did not show evidence of establishing word-referent associations in the present study, it is highly unlikely that these children were generally incapable of fast mapping: they were several months older than the children who have been shown to fast map in visual preference paradigms (Schafer & Plunkett, 1998; Werker et al., 1998; Woodward et al., 1994). A more reasonable interpretation of the results is that the fast mapping load in the present study, which used 30 novel words and fairly few repetitions, was too high for the low producers. Thus, findings of the current study indicate that receptive fast mapping is not an "all or none"-ability which is in place by e.g. 13–15 months. More specifically, there appears to be important improvements in fast mapping skills which are related to the attainment of the productive vocabulary spurt. In other words, findings of the present study point to the possibility that it is not only development in purely productive abilities, such as improvement of articulatory skills (Woodward et al., 1994) or gains in social motivation (Ninio, 1995) that may contribute to the dramatic changes in productive abilities that normally take place during the second half of the second year. We hypothesize that improvement in fast mapping abilities may be an important factor underlying the vocabulary spurt. This hypothesis is in accordance with studies showing a close relation between productive and receptive vocabularies in the second year of life (Harris et al., 1995; Reznick & Goldfield, 1992). A natural next step to investigate a possible causal relation between developments in productive vocabulary and receptive word learning would be a longitudinal study assessing whether changes in productive vocabulary go together with changes in receptive fast mapping abilities at different points in time.

As for word form recognition, results from the current experiment are compatible with recent Eye-Tracking studies which have shown that speed and accuracy of word recognition is positively related to productive vocabulary size in children in their second and third years of life (Fernald et al., 2006; Fernald et al., 2001; Zangl et al., 2001). These previous findings have been taken as evidence that there are significant changes in lexical processing skills around the vocabulary spurt. However, the above studies have all employed familiar words as stimuli, making it difficult to distinguish effects of experience with specific words from effects of general word processing skills. Results of the present study extends the previous results by showing effects of vocabulary size on *novel* word familiarization, thus eliminating explanations in terms of experience with specific lexical items.

In sum, findings of the current study indicate that the productive vocabulary spurt in the second year is related to both speed of word recognition and the receptive ability to fast map between words and referents. Future studies may determine whether the observed relation between vocabulary size and word learning ability results from a shift in language processing mechanisms which co-occurs with the vocabulary spurt or from cognitive differences which are stable across development.

6. Acknowledgements

A special thanks to Magnus Lindgren for numerous instructive discussions. Thanks are also due to Janne Mari Svangstu and Hanna Friis-Hanssen for their assistance with the data collection, and to Hanne Gram Simonsen, Inger Moen, and Lars Smith for helpful comments on earlier papers from this study.

References

Bates, E., Brethenton, I. & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge, UK: Cambridge University Press.

Bates, E., Thal, D., Finlay, B. & Clancy, B. (2002). Language development and its neural correlates. In I. Rapin & S. Segalowitz (Eds.), *Handbook of neuropsychology* (pp. 109–176). Amsterdam: Elsevier.

Benedict, H. E. (1979). Early lexical development. Comprehension and production. *Journal of Child Language, 6*, 183–200.

Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15,* 17–29.

Csibra, G., Kushnerenko, E. & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 247–262). Cambridge: MIT Press.

De Haan, M. (2007). Visual attention and recognition memory in infancy. In M. De Haan (Ed.), *Infant EEG and event-related potentials* (pp. 101–143). Psychology Press.

Fenson, L., Dale, P., Reznick, J.S., Bates, E., Thal, D. & Pethick, S. (1994). *Variability in early communicative development*. Chicago: University of Chicago Press.

Fenson, L., Dale, P., Reznick, J.S., Thal, D., Bates, E., Hartung, J., et al. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.

Fernald, A., Perfors, A. & Marchman, V.A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*, 98–116.

Fernald, A., Swingley, D. & Pinto, J.P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development, 72*, 1003–1015.

Friedrich, M. & Friederici, A.D. (2004). N400-like semantic incongruity effect in 19-month-olds: Processing known words in picture contexts. *Journal of Cognitive Neuroscience, 16*(8), 1465–1477.

Friedrich, M. & Friederici, A.D. (2006). Early N400 development and later language acquisition. *Psychophysiology, 43*, 1–12.

Gershkoff-Stowe, L. & Hahn, E.R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research, 50*, 682–697.

Goldfield, B.A. & Reznick, J.S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language, 17*, 171–183.

Harris, M., Yeles, C., Chasin, J. & Oakeley, Y. (1995). Symmetries and asymmetries in early lexical comprehension and production. *Journal of Child Language, 22*, 1–18.

Heibeck, T.H. & Markman, E.M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58,* 1021–1034.

Hill Karrer, J., Karrer, R., Bloom, D., Chaney, L. & Davis, R. (1998). Event-related brain potentials during an extended visual recognition memory task depict delayed development of cerebral inhibitory processes among 6-month-old infants with Down syndrome. *International Journal of Psychophysiology,* 29, 167–200.

MacShane, J. (1979). The development of naming. *Linguistics, 17*, 879–905.

Mills, D.L., Conboy, B. & Paton, C. (2005). Do changes in brain organization reflect shifts in symbolic functioning? In L. Namy (Ed.), *Symbol use and symbolic representation* (pp. 123–153). Mahwah, NJ: Lawrence Erlbaum Associates.

Nazzi, T. & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition? *Developmental Science, 6*, 136–142.

Ninio, A. (1995). Expression of communicative intents in the single-word period and the vocabulary spurt. In K.E. Nelson & Z. Reger (Eds.), *Children's language* (pp. 103–124). Hillsdale, NJ: Earlbaum.

Reznick, J.S. & Goldfield, B.A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology, 28*, 406–413.

Schafer, G. & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development, 69*, 309–320.

Tan, S.H. & Schafer, G. (2005). Toddlers' novel word learning: Effects of phonological representation, vocabulary size and parents' ostensive behaviour. *First Language, 25*, 131–155.

Torkildsen, J. v. K., Friis Hansen, H., Svangstu, J.M., Smith, L., Simonsen, H.G., Moen, I., et al. (in press). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*.

Torkildsen, J. v. K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H. G., Moen, I., et al. (2006). Semantic organization of basic level words in 20-month-olds: An ERP study. *Journal of Neurolinguistics, 19*, 431–454.

Torkildsen, J. v. K., Svangstu, J. M., Friis Hansen, H., Smith, L., Simonsen, H. G., Moen, I., et al. (2008). Productive vocabulary size predicts ERP correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience, 20*, 1266–1282.

Torkildsen, J. v. K., Syversen, G., Simonsen, H. G., Moen, I. & Lindgren, M. (2007a). Brain responses to lexical-semantic priming in children at-risk for dyslexia. *Brain and Language, 102*, 243–261.

Torkildsen, J. v. K., Syversen, G., Simonsen, H. G., Moen, I. & Lindgren, M. (2007b). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics, 20*, 332–351.

Webb, S. J. & Nelson, C. A. (2001). Perceptual priming for upright and inverted faces in infants and adults. *Journal of Experimental Child Psychology, 79*(1), 1–22.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M. & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34*, 1289–1309.

Wiebe, S., Cheatham, C. L., Lukowski, A. F., Haight, J. C., Muehleck, A. J. & Bauer, P. J. (2006). Infants' ERP responses to novel and familiar stimuli change over time: Implications for novelty detection and memory. *Infancy, 9*, 21–44.

Woodward, A. L., Markman, E. M. & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology, 30*, 553–556.

Zangl, R., Klarman, L., Thal, D. & Bates, E. (2001). *On-line word processing in infants and toddlers: The role of age, vocabulary and perceptual degradation*: Center for research in Language, University of California, San Diego.

Syntax

Restricted discrimination between local and global economy in agrammatic aphasia

Zoltán Bánréti

Research Institute for Linguistics, Hungarian Academy of Sciences

Abstract

We conducted sentence-picture matching tests with Hungarian speaking Broca's aphasics. The results showed that Principle A for binding was limited (not Principle B) in some complex syntactic structures. We will give a characterisation of the limited applicability of the following principles in the aphasic data:

We suggest a characterisation of the limited binding principles in agrammatic aphasia data in terms of global economy and local economy. In the case of local economy, the decision concerning the applicability of an operation depends on what pieces of information are available within the sentence representation at hand, irrespective of other sentence representations (cf. Principle A). Global economy constraints require comparison of several sentence representations in order for a decision to be made concerning the applicability of some operation (Principle B).

The discrimination between global economy and local economy in the judgement of pronominal categories was impaired in the performance of the subjects. This distinction is normally strictly made both in structure building and in interpretation but is lacking in the performance of our subjects.

1. Introduction

In the generative theory of grammar proposed by Chomsky (1997, 1998, 2004, 2005), the principles of Universal Grammar (UG) are revealed in each natural language. The set of economy constraints is a general property of the architecture of UG (i.e. their existence is not language specific). The formal economy principles of the Minimalist Program, namely, *shortest move, fewest steps in operation* and *cyclic application of operations in structure extension*, can be motivated in terms of a non-technical notion of economy, involving the notion of limited resources. According to Collins (2003, p. 45), the grammar tends to minimize whatever can be ranked along some scale: length of movements, number of operations, sets of features needed, number of syntactic objects in representations, etc. In the understanding of Wilder et al. (1997, pp. 205–226), what the economy principles represent within the architecture of grammar is the fact that a system of unlimited generative capacity is based on a limited set of mental computational resources.

Within the class of universal economy principles, local versus global economy constraints can be distinguished. In the case of local economy, decision concerning the applicability of a given operation depends on what pieces of information are available *within* the sentence representation at hand, irrespective of other possible sentence representations. Global economy constraints require comparison of several sentence representations in order for a decision to be made concerning the applicability of some operation.

The possibility arises that linguistic impairments due to some damage to cortical areas are linked with restricted functioning of principles or constraints of UG.

The economy constraints of the grammar are somehow restricted in the performance of agrammatic aphasic subjects. For instance, Bánréti (2003) showed that an impaired economy constraint prevents an aphasic patient from using of a type of economical ellipsis. Ruigendijk and Avrutin (2005) and Ruigendijk et al. (2006) demonstrated that syntactic operations are not the cheapest option for agrammatic speakers. In their study with Dutch agrammatic aphasics, the subjects were examined with a picture selection task that required interpretation of pronouns and reflexives. They investigated the subjects' ability to interpret pronominal elements in transitive clauses and Exceptional Case Marking constructions. The results were interpreted in the framework of the Primitives of Binding Theory (Reuland, 2001). Among other things, the results showed that the economy hierarchy for reference establishment in the case of agrammatic aphasia differs from that of normal speakers.

In what follows, some results of a test conducted with the participation of Hungarian speaking Broca's aphasic subjects will be shown. The results can be explained as due to some limitation of discrimination between the principles of local versus global economy.

2. The binding principles

The binding principles are among the universal principles of grammar. Binding depends on the properties of the syntactic structure concerned. The basic claim is this: α *binds* β *if* α *and* β *bear the same (referential) index and* α *c-commands* β.

Principle A:
An anaphor (reflexive pronoun) must be bound in its governing category (clause).

Principle B:
A personal pronoun must be locally free (non-bound).

Principle C:
An R-expression must be free (that is, its reference must be independent of that of other constituents).

Principle B states the condition of impossibility of binding, that is, where binding *cannot* be applied, and not the condition for binding to take place. The latter kind of condition is given by Principle A with respect to reflexive/reciprocal pronouns. As we will see, Principle B is to be applied *globally* in that it requires comparison of several possible sentence representations to be assessed. Principle B is constrained by global economy.

Principle A, on the other hand, works as a *local* principle since its applicability can be determined on the basis of the properties of a single sentence representation. Therefore Principle A must be based on local economy.

2.1. Is Principle A easier than Principle B?

The results of Grodzinsky et al. (1993) suggest that their agrammatic aphasics showed limitations in the tasks concerning the binding of personal pronouns as opposed to the binding of reflexive pronouns. They explained this by assuming that whenever the two types of binding can be chosen between in the case of ambiguous sentences, following Principle A (applied to reflexive pronouns) is simpler in that it does not require consideration of the context.

2.2. Does Principle B require a kind of global economy?

In Grodzinsky et al.'s experiments, the aphasic subjects generally made correct judgements for sentences containing quantified expressions plus personal pronouns where there was no alternative contentful antecedent in another clause. Hence, there was no need to compare alternative structures. But they performed at random when the antecedent of a personal pronoun was to be one of several available contentful referential expressions. Here they had to consider which of the alternative structures would fit the context. The point to consider is as follows. If the listener hears a sentence containing a personal pronoun, s/he has to decide if it is permitted for the personal pronoun to be coreferent with an antecedent *inside* the clause. In other words, s/he has to see if it is possible to replace the pronoun by a locally bound anaphor. For this, s/he has to construct an *alternative* bound representation. If this is *not* possible, the task has come to an end; coreference is permitted, not prohibited.

If there is a possibility of alternative binding, the listener has to construct two representations: one containing a possible binding relation within a local domain and another one that contains the alternative coreference reading with an antecedent from "outside" the local domain. Then s/he has to consider both representations against the context to see if they differ. If they do, coreference is permitted, otherwise it is rejected. Performing such a series of steps constitutes much more of a burden on computational resources than simply enforcing Principle A or Principle C. In addition, the alternative structural representations have to be assessed with respect to their compatibility with the context. In order to do that, at least two representations have to be accessed, and each of them has to be compared with the context so that the appropriate one can be selected. These operations require a kind of global economy.

3. The test material

The subjects were as follows: P., a 25-year-old right-handed man, with left fronto-temporo-parietal lesion of a traumatic origin, linguistic symptoms of agrammatic Broca's aphasia; and B. I., a 49-year-old right-handed man with left temporo-parietal insula of a traumatic origin, Broca's aphasia.

Five syntactic structures of varying complexity were selected following the logic of binding theory. For each construction, we compiled sentences, each of which had two versions constituting a minimal pair, involving a reflexive pronoun versus a personal pronoun. For each pair of sentences, two pictures were drawn, suggesting the meanings of the respective sentences. The members of sentence pairs/picture pairs were presented separately, in random order. The test material included 200 sentences and 200 pictures. In *yes/no* decision tasks, each sentence was heard twice once paired up with one of the relevant pictures, and once with the "wrong" picture. With the two aphasic subjects this yielded 400 grammaticality judgements. The subject saw a picture and heard a sentence and was asked to decide if what he heard corresponded to what he saw. Wherever we thought it was necessary, an introductory sentence providing context was also provided (which mainly served for identifying the persons in the picture and disambiguating the things or events that the picture was meant to represent). The total test material was presented in two sessions. The time that elapsed between the sessions was 2 weeks. In each test situation, a given sentence or a given picture only occurred once. Assuming that the distance of a pronoun from its antecedent may affect the assignment of referential relations, in one pair of sentences for each structure (a total of 10 sentences) the topicalised NP in sentence initial position was distanced from the pronoun by inserting *ami azt illeti* 'for that matter'.

We wanted to find out whether there are differences in assigning antecedents to the reflexive versus personal pronouns and if there are, how much they are affected by the structure of the sentences, by their complexity. The five types of structure are detailed below and illustrated by one example each in Figures 1–5.

Type I

This is a simple sentence structure ([1a–b], Figure 1). The first sentence is the version with a reflexive pronoun; the second sentence is that with a personal pronoun.

[1a]	A	bohóc	fejbevágja	<u>magát</u>
	the	clown	hits	himself
	'The clown hits <u>himself</u>'			

[1b]	A	bohóc	fejbevágja	<u>őt</u>
	the	clown	hits	him
	'The clown hits <u>him</u>'			

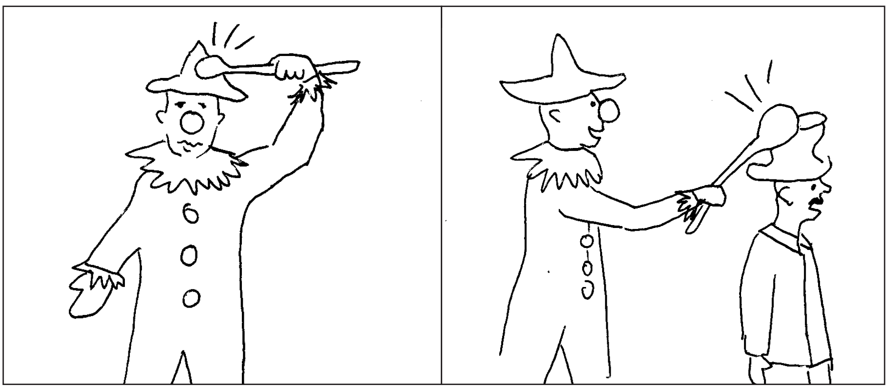


Figure 1. *The clown hits himself / him*. Simple sentence structure.

Type II

The reflexive pronoun or the personal/possessive pronoun occurs in a possessive construction in possessor position ([2a–b], Figure 2). With respect to Principle B, the possessive construction may work as a local domain, from outside which binding is not prohibited.

[2a]	[Mari	befonja	[$_{DP}$ a	maga	haját]].			
	Mary	plaits	the	herself	hair.poss.acc			

'Mary plaits her own hair.'

[2b]	[Mari	befonja	[$_{DP}$ az	ő	haját]].			
	Mary	plaits	the	herself	hair.poss.acc			

'Mary plaits someone else's hair.' / 'Mary plaits her own hair.'

It should be noted that the possessor position in a DP in Hungarian is one in which both types of pronouns can occur with the same reference. An overt possessive personal pronoun (ő) *may* also have a reading with *locally prohibited coreference* such that the hair concerned is not Mary's but somebody else's.

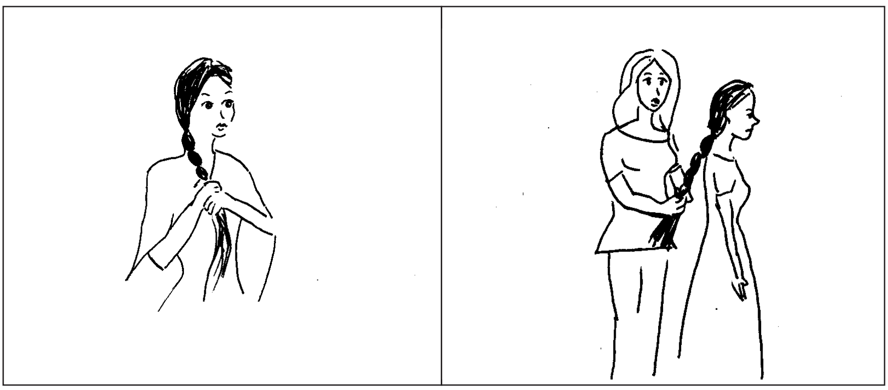


Figure 2. *Mary plaits her own hair / someone else's hair.* Pronoun in possessor position.

Type III

The pronoun is not in possessor position but is a modifier of the possessed noun ([3a–b], Figure 3). This restores complementary distribution between reflexive and personal pronouns. A factor that may make processing more difficult is that there are *two* possible antecedents (*lány, fiú*) of which the subject has to select one that suits the type of pronoun involved.

[3a]	A	lány	örül	[$_{DP}$ a	fiú	önmagáról	készült	fényképének]].
	the	girl	is glad	the	boy	on himself/herself	made	photograph

'The girl is glad about the boy's photograph of himself.'

[3b]	A	lány	örül	[$_{DP}$ a	fiú	őróla	készült	fényképének]].
	the	girl	is glad	the	boy	on him/her	made	photograph.dat

'The girl is glad about the boy's photograph of her.'

Figure 3. *The girl is glad about the boy's photograph of himself / her.* Pronoun is a modifier of the possessed noun.

Type IV

The infinitival construction (S_{inf}) has no lexical subject but does have a covert pronominal subject marked as PRO ([4a–b], Figure 4). The antecedent of the reflexive pronoun is PRO = the girl, whereas the personal pronoun has to find an antecedent *outside* the sentence.

[4a]	[A	lány	szereti	[$_{Sinf}$ PRO	nézegetni	<u>magát</u>	az	albumban]].
	The	girl	likes		look.at.inf	herself.acc	the	album

'The girl likes to look at <u>herself</u> in the album.'

[4b]	[A	lány	szereti	[$_{Sinf}$ PRO	nézegetni	<u>őt</u>	az	albumban]].
	The	girl	likes		look.at.inf	him/her	the	album.in

'The girl likes to look at <u>him/her</u> in the album.'

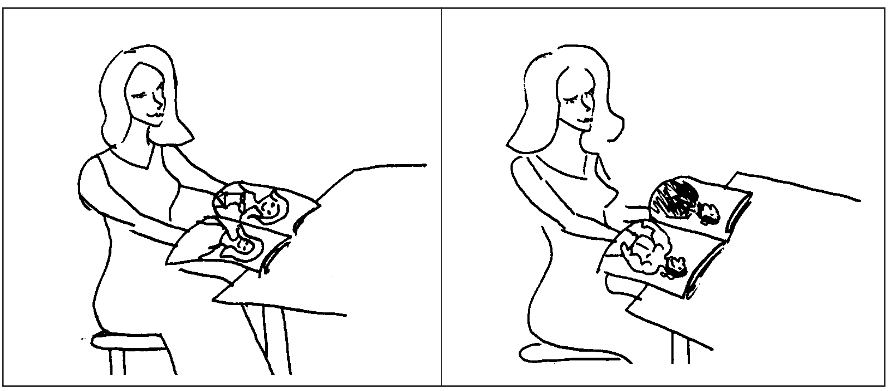


Figure 4. *The girl likes to look at herself / her in the album*. Infinitival construction (S_{inf}) with covert pronominal subject (PRO).

Type V
The infinitival construction (S_{inf}) has its "own" *lexical* subject that gets its accusative case feature from "outside", from the verb of the matrix clause (*látja*) but that is simultaneously the subject of *mutogatni* in the infinitival construction ([5a–b], Figure 5). Again, a factor making processing the sentences more difficult is that there are two possible antecedents (*asszony, valaki*) in the sentence to choose from in accordance with what the pronoun requires.

[5a]	Az	asszony	lát	[valakit	magára	mutogatni].
	The	woman	sees	someone.acc	him/herself.at	point.inf

'The woman sees someone pointing at himself/herself'.

[5b]	Az	asszony	lát	[valakit	őrá	mutogatni].
	The	woman	sees	someone.acc	him/her.at	point.inf

'The woman sees someone pointing at her.'

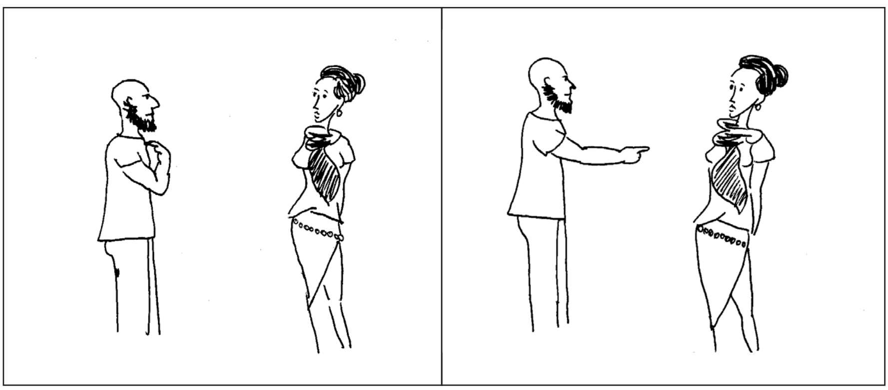


Figure 5. *The woman sees someone pointing at himself / her.* Infinitival construction (S_{inf}) with lexical subject.

4. Results and discussion

The 400 grammaticality judgements made in the two tests exhibit the distribution shown in Table 1.

In the case of type I, the subjects' assessments were faultless in the case of reflexive pronouns, whereas in assigning reference relations to personal pronouns the subjects made 8 wrong decisions in the context of the ungrammatical picture.

For type II, 4 wrong decisions were also made in the context of the ungrammatical picture, for examples containing the inserted material. Similarly, 4 wrong decisions were made with respect to personal pronouns, again in the ungrammatical picture context.

Table 1. Grammaticality judgements for the sentence-picture matching

Type of structure	Pronoun	Grammatical	Ungrammatical
I	Reflexive	40 (100%)	0 (0%)
	Personal	32 (80%)	8 (20%)
II	Reflexive	36 (90%)	4 (10%)
	Personal	36 (90%)	4 (10%)
III	Reflexive	8 (20%)	32 (80%)
	Personal	28 (70%)	12 (30%)
IV	Reflexive	32 (80%)	8 (20%)
	Personal	28 (70%)	12 (30%)
V	Reflexive	12 (30%)	28 (70%)
	Personal	24 (60%)	16 (40%)

Type III contained two contentful NPs. The pronoun occurred within the possessive construction as a modifier of the possessed noun. The subjects here consistently preferred the sentence-initial NP that was a lot further away from the pronoun than the other potential nominal antecedent. With respect to reflexive pronouns that require strictly local binding, this resulted in 32 wrong decisions of the 40, a very poor result. The same strategy led to 28 correct decisions in assessing the reference of personal pronouns since these are to be bound by a non-local antecedent.

In the case of type IV, both for reflexive and for personal pronouns, 8 and 12 wrong decisions were made in the context of the ungrammatical picture. The number of correct decisions was 32 and 28, respectively.

For type V sentences, we see a similar pattern as for type III. The sentences to be assessed contain two potential nominal antecedents, one sentence initially, and another one right before the pronoun. In assessing the reference of reflexive pronouns, 28 incorrect decisions were made along with 12 correct ones; the incorrect decisions all turned the sentence initial (non-local) nouns into antecedents although the reflexive pronoun would have to be locally bound by the noun immediately preceding it. The same strategy resulted in 24 correct decisions in the case of personal pronouns since the required non-local antecedent of a personal pronoun may well be the sentence initial NP.

5. Conclusions

Principle A, demanding reflexive pronouns to be locally bound, expressly refers to syntactic structure. Hence, the correctness of decisions concerning reflexive pronoun binding depends on the structural complexity of sentences. The structure of sentences of types I and II was simpler, whereas that of types III–V was more complex.

Principle B, referring to the binding of personal pronouns, is based on under what circumstances that binding *would* be impossible. The application of that principle requires the listener to construct alternative syntactic structures/representations, compare them to one another, and relate them to the context, too. Due to the role of context, it is not only the UG principle but also language specific preferences and personal problem solving strategies that may play a role in the listener's decision. In the case of the simplest, type I structures, the subjects made correct discriminations between reflexive and personal pronouns, as they did not resort to guessing for any of them. As the structural complexity of the sentences grew, the number of correct decisions with respect to personal pronouns decreased somewhat. Two points can be made in that respect. Incorrect decisions were mainly made in the context of the non-matching picture, especially if the personal pronoun was in a relatively complex construction and/or was preceded by inserted material. The deteriorative role of insertion is not surprising: it shows that wherever one of the possible antecedents is distanced from the pronoun, the task becomes more difficult.

An especially interesting result is what we got in the case of types III and V, for reflexive pronouns. The subjects wrongly took the sentence-initial NP to be the antecedent of the pronoun in 16 and 14 cases, respectively, as opposed to the actual, local antecedent that immediately preceded the pronoun. The limited range of our data obviously only allows us to draw tentative conclusions. Structures III and V are syntactically complex but not in the same way as sentences involving inserted material. What matters is not the mere distance of the critical items but rather the structural position of the pronoun. The pronoun is either at the "deepest" point of the possessive construction, in the possessed NP (type III) or it is a constituent of an infinitival clause within the sentence. We have to assume that the structural complexity of sentences elicits *alternative structural analyses* and their assessment, similarly to what we said about decisions with respect to the local binding of personal pronouns. And the net result is that the subjects wrongly assumed bindings for reflexive pronouns that would have been grammatical local bindings for personal pronouns.

The scheme of the incorrect decisions showed a kind of restricted economy. For type III the subjects incorrectly accepted [6a] in the context of a picture that showed a photograph of the boy, not of the man. According to the wrong decision, *himself* = *the man*. The binding relation in the incorrect decision is the following:

[6a] *A férfi$_m$ örül a fiú önmagáról$_m$ készített fényképének.
*'The man$_m$ is glad about the boy's photograph of himself$_m$.'

On the other hand, the grammatical binding relation (*himself* = *the boy*) would have been as in [6b] in a correct decision:

[6b] A férfi örül a fiú$_m$ önmagáról$_m$ készített fényképének.
'The man is glad about the boy's$_m$ photograph of himself$_m$.'

For type V the subjects incorrectly accepted [7a] in the context of a picture that showed a man pointing at a little boy, not at himself. According to the wrong decision *himself = the little boy:*

[7a] *A kisfiú$_w$ látja a férfit mutogatni önmagára$_w$.
*'The little boy$_w$ sees the man point at himself$_w$.'

The correct decision would have been based on a grammatical binding relation in which *the man* binds *himself* [7b]:

[7b] A kisfiú látja a férfit$_w$ mutogatni önmagára$_w$.
'The little boy sees the man$_w$ point at himself$_w$.'

The assessment of alternative structural analyses elicited by more complex syntactic structures resulted in better decisions with respect to personal pronouns: for type III and V the subjects made correct decisions in 28, respectively 24, cases out of 40.

As witnessed by the simpler structures (types I and II), the subjects did possess the ability to distinguish binding relations of reflexive from those of personal pronouns. With more complex structures (types III and V) they did *not* react to increasing complexity by trying to resort to some structure that was simpler or shorter. On the contrary: in their *incorrect* decisions they used a more costly mechanism, incorrectly. The distinction between local economy and global economy was *not* properly accessible for them. Instead of assessing the operations to be employed in a less costly manner, restricting their attention to local structural relations and ignoring other, non-local structural representations, they tried to do the opposite: they attempted to make a decision on the applicability of some structural operation by comparing alternative structural relations to one another. But the correct decisions simply needed analysing local structural relations within a local domain. This option was avoided. We suggest that such distribution of performance can be attributed to the subject's limited ability to tell local and global economy from each other.

6. Acknowledgements

This work is part of the OTKA (Hungarian National Research Foundation) funded project (grant number: NK 72461).

References

Avrutin, S. (2006). Weak syntax. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's Region* (pp. 49–62). Oxford University Press, New York.

Bánréti, Z. (2003). Economy principles of grammar in agrammatic aphasia. *Brain and Language, 87*(1), 65–66.

Chomsky, N. (1997). Language and mind: Current thoughts on ancient problems, *Pesquisa Linguistica*, Universidade de Brasilia 3–4, 21.

Chomsky, N. (1998). Some observations on economy in generative grammar. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax* (pp. 115–127). MIT Press, Cambridge MA.

Chomsky, N. (2004). *Biolinguistics and the human capacity*. Talk at the Hungarian Academy of Sciences May 17, 2004. Budapest, Hungary.
Chomsky, N. (2005). Three factors in language design, *Linguistic Inquiry 36*, 1–22.
Collins, C. (2003). Economy conditions in syntax. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 45–61). Blackwell, Oxford/Malden.
Grodzinsky, Y., Wexler, K., Chien, Y., C., Marakovitz, S. & Solomon, J. (1993). The breakdown of binding relations. *Brain and Language 45*, 369–422.
Reuland, E (2001). Primitives of binding. *Linguistic Inquiry 32*, 439–492.
Reuland, E. & Avrutin, S. (2005). Binding and beyond: Issues in backward anaphora. In A. Branco et al. (Eds.), *Anaphora processing* (pp. 139–162). John Benjamins, Amsterdam.
Ruigendijk, E., Vasić, N. & Avrutin, S. (2006). Reference assignment: Using language breakdown to choose between theoretical approaches. *Brain and Language*, *96*, 302–317.
Wilder, Ch., Gaertner, H. M. & Bierwisch. M. (Eds.) (1997). The role of economy principles in linguistic theory. *Studia Grammatica 40*, 205–226. Akademie Verlag, Berlin.

Negative polarity items and complementizer agreement in Basque

Leticia Pablos
University of Reading
University of the Basque Country

Douglas Saddy
School of Psychology and Clinical Language Sciences, University of Reading

Abstract

This paper examines a specific agreement requirement with the complementizer in Negative Polarity Item (NPI) constructions in Basque. The realization of this agreement is done through a partitive morphology at the complementizer that can only be licensed under the presence of an NPI within the same clause and a licensor such as negation in the main clause. Our aims were two-fold. First we looked at this agreement requirement to determine whether it is the semantic properties of individual licensors or the semantic properties of the whole proposition that are relevant for the licensing of this agreement. Second, we examined the interaction of the syntactic and semantic components involved in NPI licensing when there is an agreement realization required besides the basic NPI licensing requirements. In order to examine these questions we conducted a self-paced reading experiment in Basque.

1. Introduction

In this study we examine an agreement phenomenon with the complementizer in Basque in Negative Polarity Item (NPI) constructions. In particular, we are interested in examining whether it is the semantic properties of individual licensors and the proximity of an NPI, or the semantic properties of the whole proposition that are relevant for the licensing of agreement in these constructions. In other words, we want to explore contexts where the presence of a licensor and its accessibility is not enough to license the NPI, and where there needs to be a specific morphological agreement realization at the complementizer. For example, in Basque the complementizer must have a partitive agreement realization in the clause where the NPI is contained in order for the NPI to be licensed. To address these questions we tested contexts in which the partitive marked complementizer might be licensed: presence or absence of negative matrix verb; and presence or absence of an NPI in the embedded clause.

1.1. NPI Licensing: General requirements

Negative Polarity Items (NPIs) are lexical elements that are licensed in specific semantic environments (among many others, Ladusaw, 1980; Linebarger, 1980; Krifka, 1995; von Fintel, 1999; Giannakidou, 2002). One of the basic requirements that an NPI must fulfill is that there needs to be a licensor, Negation being one of the most common licensors. There are, however, other contexts where NPIs are licensed, such as conditionals or interrogatives, which share some semantic properties with Negation. As shown in the contrast in [1] and [2], when the licensor is absent, the NPI cannot be licensed and therefore the sentence becomes unacceptable.

[1] No child would ever eat herring.

[2] *A child would ever eat herring.

[3] *The teacher who liked no child would ever eat herring.

In addition to the specific semantic environment, NPIs must have a specific structural relation with the licensors, i.e. the NPI needs to be c-commanded by its licensor. This is shown in the contrast in [2] and [3]. The NPI *ever* in [3] cannot be licensed by negation because the negation is within the relative clause and it cannot c-command the NPI, which is part of the main clause.

In sum, if any of these basic requirements is violated, then the licensing conditions for the NPI are violated and it results in ungrammaticality.

1.2. NPI Licensing in English and German

Research done in German and English on the processing of NPI licensing has investigated the licensing conditions of NPIs (Saddy et al., 2004; Drenhaus et al., 2005; Vasishth et al., 2008; Xiang et al., 2009).

In these studies some of the basic requirements for licensing NPI constructions were manipulated to examine their processing: whether there was a licensor in the previous context; whether this licensor was accessible to license the NPI when the licensor was contained within an embedded clause; whether there were effects created by this embedded licensor, which could interfere in the licensing of the NPI; and whether all kinds of licensors behaved equally when licensing the NPI during real-time processing.

To investigate these questions, previous studies examined the following paradigm summarized in [4] to [6]. The examples in (a) are the sentences used by Xiang et al. (2006) to test English NPIs and those in (b) are the sentences used by Saddy et al. (2004) and Drenhaus et al. (2005, 2006) to test German NPIs. In [4a–b], the licensor is accessible; the negation c-commands the NPI. In [5a–b], on the other hand, the potential licensor is inaccessible: it does not c-command the NPI because it is buried within a relative clause. Finally, in [6a–b], there is no licensor in the previous context to license the NPI.

[4a] No bills that the Democratic senators have supported will ever become law.

[4b] Kein Mann, der einen Bart hatte, war jemals glücklich.
'No man who had a beard was ever happy'

[5a] *The bills that no Democratic senators have supported will ever become law.

[5b] *Ein Mann, der keinen Bart hatte, war jemals glücklich.
'A man who had no beard was ever happy'

[6a] *The bills that the democratic senators had supported will ever become law.

[6b] *Ein Mann, der einen Bart hatte, war jemals glücklich.
'A man who had a beard was ever happy'

The studies on NPI licensing in German and English revealed that the parser does not always respect the licensing of structural conditions of NPIs and that, surprisingly, speakers sometimes consider constructions where the licensor is structurally inaccessible for the NPI, such as those in [5], as acceptable. This effect has been referred to as the "spurious licensing effect" in the literature.

Drenhaus et al. (2005) and Vasishth et al. (2005) found a spurious licensing effect for the condition with an inaccessible licensor in [5b]. In an ERP experiment, Drenhaus et al. (2005) found that sentences with an inaccessible licensor generated an N400 and a P600 effect, ERP components that reflect semantic and syntactic incongruencies respectively. In a speeded-grammatical judgement task, Vasisth et al. (2005) showed that participants considered sentences with inaccessible licensors as grammatical.

Subsequent results from behavioural data by Xiang et al. (2006) and Vasishth et al. (2006), on the other hand, pointed to a contrast between off-line versus on-line tests with respect to the spurious licensing effect found in these constructions. They concluded that participants considered the sentences with an inaccessible licensor as grammatical when they processed the sentences in real-time, but that the contrary occurred when they were asked to judge the grammaticality of the sentences off-line.

Overall results from the NPI studies in German and English raise the issue of whether or not the observed intrusive negation effects could be accounted for as a processing strategy rather than as a side-effect of a failure to meet the structural conditions for the licensing of the polarity item.

Furthermore, they also bring up the question of what kind of information the parser uses when constructions that require more semantic composition than pure syntactic dependencies are processed such as wh-dependencies (Crain & Fodor, 1985; Stowe, 1986; Aoshima et al., 2004; Lee, 2004), backward anaphora dependencies (van Gompel & Liversedge, 2003; Sturt et al., 2004; Kazanina et al., 2007), or topic–clitic dependencies (Pablos, 2006; Pablos & Phillips, 2006). Besides the specific syntactic conditions required to form the structural dependency consisted of the licensor (negation) and the dependent element (NPI), NPI licensing imposes certain semantic constraints on the parser that are in principle harder to disentangle in languages such as German and English because of their internal grammatical configuration. Results from the German and English studies suggest that NPI licensing requires more than structural dependence alone and that semantic aspects of the sentences need to be considered when licensing these constructions on-line. Nevertheless, it is difficult to identify what specific information the parser is using to license these negative elements and how these two components of the grammar interplay when processing the sentence in real-time.

In light of previous research, the current study exploits an agreement requirement with the complementizer in Basque, to look at a dependency that imposes requirements both on the structure and on the semantic context and to test if this agreement phenomenon can shed further light onto NPI licensing.

2. NPI licensing in Basque: Partitive agreement

In this section we turn to Basque, a language that has a specific agreement realization for NPI licensing at the complementizer in order to explore an agreement phenomenon present in NPI constructions. Specifically, we examine how this is reflected in on-line comprehension of sentences containing NPIs when the licensing conditions for polarity items are mediated through an agreement requirement present in embedded structures that involve NPIs. This agreement requirement in NPI constructions in Basque is reflected in the appearance of a partitive complementizer. It is therefore an agreement that is realized at the complementizer and that should in principle be more complex to process than other kinds of agreement phenomena such as subject–verb or object–verb agreement because it involves the clausal level. This agreement could be checked locally (within the embedded clause/with the NPI in the embedded clause) or long-distance (with the whole proposition/with the main verb).

Basque has a partitive suffix (*-rik*) that requires licensing by some polar element and that can be licensed in several syntactic environments, negation being among them as shown in [7] (Etxepare, 2003).

[7]	Pello-k	ez	du	liburu-rik	irakurri.
	Pello-ERG	not	aux	book-PRTT	read
	'Pello hasn't read any book'				

Interestingly, this partitive morpheme *-ik* can be attached to finite dependents headed by the complementizer *-en* (a complementizer used in interrogative or relative clauses). The complementizer *-enik* resulting from this combination is only present in contexts where the partitive is triggered (i.e. negation, conditionals, interrogatives) and for propositional complements of negative verbs as shown in [8] and [9].

[8]	Pello-k	ez	du	uste	etorri-ko	d-en-ik
	Pello-ERG	not	aux	think	come-FUT	aux-COMP-PRTT
	'Pello doesn't think she will come'					

[9]	Pello-k	ukatu	du	etorri-ko	d-en-ik
	Pello-ERG	deny	aux	come-FUT	aux-COMP-PRTT
	'Pello denied that he/she would come'				

Interestingly for our purposes, NPIs such as *inor* 'anybody' can be licensed in dependent clauses by this partitive complementizer (Laka, 1993).

[10]	Pello-k	ez	du	uste	[inor-∅	etorri-ko	d-en-ik]
	Pello-ERG	neg	aux	think	anyone-ABS	come-FUT	aux-COMP-PRTT
	'Pello doesn't think that anybody will come'						

In sum, the properties of the semantic context in which the partitive complementizer is obligatorily required in Basque NPI constructions are the following (Laka, 1990; Uribe-Etxebarria, 1994; Hualde & Ortiz de Urbina, 2003):

1. when there is an NPI in the preceding context and this NPI is a clause-mate of the Complementizer;
2. when there is negation in the main clause (overt, or covert through a negative verb such as *deny, reject*);
3. when the semantics of the verb scoping over the NPI license it.

These requirements are exemplified in [11] and [12], which illustrate how the semantics of the main verb and the presence of an NPI in the embedded clause affect the requirement to have a partitive complementizer. The partitive complementizer is only required when there is negation in the main clause and an NPI in the embedded clause, as shown in [11a] and [11b]. When there is no negation in the main clause the partitive complementizer is not possible, as in [11c]. The declarative complementizer, on the other hand, is not licensed in [11a] through [11c].

[11a] Pello-k ez du uste [inor-∅ etorri-ko ✓d-en-ik/*d-ela]
Pello-ERG neg aux think anyone-ABS come-FUT aux-COMP-PRTT/aux-DECL
'Pello doesn't think that anybody would come'

[11b] Pello-k ukatu du [inor-∅ etorri-ko ✓d-en-ik /*d-ela]
Pello-ERG deny aux anyone-ABS come-FUT aux-COMP-PRTT/aux-DECL
'Pello has denied that anybody would come'

[11c] Pello-k onartu du [inor-∅ etorri-ko *d-en-ik /*d-ela]
Pello-ERG confirm aux anyone-ABS come-FUT aux-COMP-PRTT/aux-DECL
'Pello has confirmed that anybody would come'

Furthermore, when the NPI is not present in the embedded clause and there is a proper noun instead, then the partitive complementizer becomes optional as in [12a], or impossible, as illustrated in [12b] and [12c]. The only possible complementizer is the declarative complementizer *-ela*.

[12a] Pello-k ez du uste [Itxasne-∅ etorri-ko ✓d-en-ik/✓d-ela]
Pello-ERG neg aux think Itxasne-ABS come-FUT aux-COMP-PRTT/aux-DECL
'Pello doesn't think that Itxasne will come'

[12b] Pello-k ukatu du [Itxasne-∅ etorri-ko *d-en-ik /✓d-ela]
Pello-ERG deny aux Itxasne-ABS come-FUT aux-COMP-PRTT/aux-DECL
'Pello has denied that Itxasne would come'

[12c] Pello-k onartu du [Itxasne-∅ etorri-ko *d-en-ik /✓d-ela]
Pello-ERG confirm aux Itxasne-ABS come-FUT aux-COMP-PRTT/aux-DECL
'Pello has confirmed that Itxasne would come'

Whether or not the parser is affected by an incorrect agreement could provide more information about the role of this type of agreement related to complementizers in comprehension and about whether NPIs contribute in the processing of this agreement. In addition, if the presence of an NPI is manipulated in the sentence, more things could be learnt about how long-distance agreement is parsed in NPI constructions and about the effects that surface when this agreement information comes later in the sentence and when it depends on the appearance of previous elements (i.e. NPI) in order to be correctly interpreted.

In the current study, the factors that we manipulated were the presence of the specific agreement at the complementizer, and the presence of an NPI or an NP in the embedded clause. This was all included within a semantic context that had no negation, thus no licensor. Therefore, it is important to note that in contrast to the previous NPI-dependency studies, in the current study there is one more element that needs to be taken into consideration in the dependency, and that is agreement. Besides the licensor and the NPI, agreement is the third component that needs to be present, all within the appropriate semantic context. Furthermore, our paradigm is designed to take advantage of the fact that the processing of NPIs can be modulated by the existence of potential but not adequate licensors (the spurious licensing effect).

3. On-line experiment: The Partitive Complementizer in Basque

This section presents a self-paced reading experiment conducted to investigate the processing of NPI constructions that have an agreement realization at the complementizer and how this agreement gets assigned and interpreted when it is intrinsically related to the presence of a negative context and a preceding NPI. Relevantly, we tested whether the assignment of this agreement is done locally within the embedded clause or whether it is done at the clause level and it requires the whole proposition to have the right context.

We used a semantic context where negation was *not* present in the main verb scoping over the embedded clause and where the NPI could not be licensed, resulting in ungrammatical sentences. In particular, we examined whether agreement information is relevant when trying to decide if these constructions are licensed.

In order to test how this kind of agreement is checked, we manipulated whether there was no negation present in the semantics of the main verb; whether there was an NPI or a NP in subject position within the embedded sentence, and whether the agreement reflected at the complementizer had partitive morphology (and therefore it would be licensed under NPI licensing) or declarative morphology (where it can only be licensed if there was an NP in the preceding context).

3.1. Procedure, participants, materials

A self-paced reading experiment was conducted using a word-by-word window paradigm. We tested 32 participants who were native speakers of Basque at the University of Deusto, Bilbao, Spain. The experimental sentences we used are shown in [13a-d] (NPI partitive, NPI declarative, NP partitive, and NP declarative, respectively).

[13a] *Pello-k onartu du festa-ra inor-∅
Pello-ERG accept aux party-to nobody-ABS
mozkortuta etorri d-en-ik gaur gauean
drunk come aux-COMP-PRTT tonight
'Pello has admitted that nobody has come drunk to the party tonight'

[13b] *Pello-k onartu du festa-ra inor-∅
Pello-ERG accept aux party-to nobody-ABS
mozkortuta etorri d-ela gaur gauean
drunk come aux-DECL tonight
'Pello has admitted that nobody has come drunk to the party tonight'

[13c] *Pello-k onartu du festa-ra Itxasne-∅
Pello-ERG accept aux party-to Itxasne-ABS
mozkortuta etorri d-en-ik gaur gauean
drunk come aux-COMP-PRTT tonight
'Pello has admitted that Itxasne has come drunk to the party tonight'

[13d] *Pello-k onartu du festa-ra Itxasne-∅
Pello-ERG accept aux party-to Itxasne-ABS
mozkortuta etorri d-ela gaur gauean
drunk come aux-DECL tonight
'Pello has admitted that Itxasne has come drunk to the party tonight'

3.2. Analysis

All sentences for which the corresponding comprehension question was answered incorrectly were excluded from the RT analyses. Furthermore, in the RT analyses we excluded those reading times that were below 200 and above 1800 ms. The overall subject comprehension accuracy was 90%. No participants were excluded from the experiment. Figure 1 shows mean reading times for the NP and NPI conditions.

3.3. Results

While there were no significant differences before region 5 (NPI *inor* and NP *Itxasne*), there was an increase of reading time for all conditions at region 4 (adverbial *festara*). This reading time increase was partly expected because this adverbial cannot be interpreted in the main clause and when constructing the experimental items it had been used to signal a clause boundary.

In region 5, there was a significant difference for the "embedded subject" factor, where reading times for the NP conditions in [8c] and [8d] were overall slower than for the NPI conditions in [8a] and [8b]. The NP conditions were read on average 98 ms slower than the NPI conditions at this region. The pairwise comparisons on the other hand were not significant for this region.

In region 6 (adverbial *mozkortuta*) and region 7 (intransitive verb *etorri*) there were no significant differences.

In region 8 (auxiliaries *denik* and *dela*), there was a significant effect within the NP conditions [8c] and [8d], where the NP declarative condition was read on average 66 ms slower than the NP partitive condition. There was additionally a significant effect for the factor embedded subject, were the NPI partitive condition was read 100 ms slower than the NP declarative condition.

In region 9 (adverbial *gaur*) there was a significant effect for the NP conditions [8c] and [8d], where the NP partitive condition was read slower than the NP indicative condition (30 ms). This effect remained significant for region 10 (adverbial *gauean*) (115 ms).

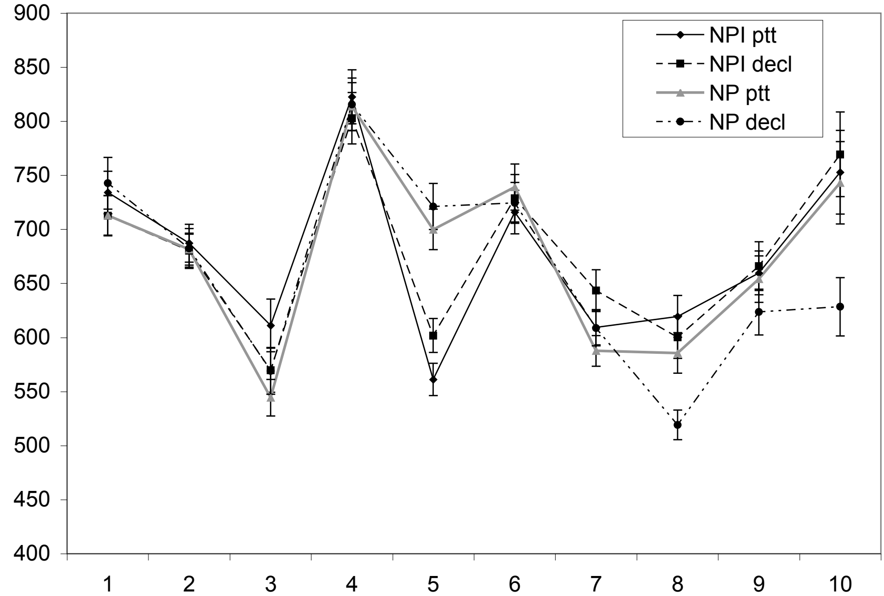


Figure 1. Mean reading times in ms for each region for the NPI and NP conditions. The x-axis refers to word position of ([13a-d]), i.e.

Pello-k$_1$ onartu$_2$ du$_3$ festa-ra$_4$ inor-∅/Itxasne-∅$_5$
Pello-ERG$_1$ accept$_2$ aux$_3$ party-to$_4$ nobody-ABS/Itxasne-ABS$_5$
mozkortuta$_6$ etorri$_7$ d-en-ik/d-ela$_8$ gaur$_9$ gauean$_{10}$
drunk$_6$ come$_7$ aux-COMP-PRTT/aux-DECL$_8$ tonight$_{9–10}$

3.4. Discussion

The most prominent effect in this experiment was the effect at the auxiliaries that carried the agreement for different complementizers (region 8), where the NP partitive *denik* in [14c] was read slower than the NP declarative *dela* in [14d], and the NPI partitive *denik* in [14a] was read significantly slower than the NP declarative *dela* in [14d]. This effect showed that there was a distinction made between grammatical and ungrammatical conditions after the whole sentence has been processed and that the parser is sensitive to agreement distinctions at the complementizer.

Furthermore, the fact that this effect is maintained across the following two regions (9 and 10) indicates that this is a robust effect of grammaticality.

On the other hand, the fact that there are no significant differences at the complementizer position in the pairwise comparison between the NPI conditions suggests that there is a tension between local licensing and long-distance licensing. If licensing proceeded locally between the NPI and the complementizer, the preference should be to have a partitive complementizer. On the other hand, if licensing proceeded long-distance between the main verb and the complementizer, the preference should be to have a declarative complementizer. The lack of effect at the complementizers in the NPI conditions suggests these two competing licensing alternatives are cancelling each other out and are obviating the effect.

The second significant effect was the overall slowdown for the NP position *Itxasne* (region 5) in the NP conditions in [14c] and [14d] in comparison to the NPI *inor* position in the NPI conditions in [14a] and [14b], which were overall faster. This effect seems to suggest that there is an additional cost for processing proper names and that the parser is not doing a definitive grammaticality check in NPI constructions until the critical complementizer region is encountered. There are several reasons why the NP conditions were read overall slower than the NPI conditions at this position. One of them is the high number of NPs in the experimental items and fillers (17%), which might have contributed to participants paying more attention to these cases in order to answer the comprehension questions correctly. Additionally, proper names have been found to require a higher cost in processing as illustrated by studies that have shown effects associated with processing more than one proper name in the discourse (Gordon & Hendrick, 1997; Gordon et al., 1999; Ledoux et al., 2007). The other reason is the lack of grammatical sentences with NPIs in the experiment, which could have made participants adopt a strategy to think that all the NPI sentences were ungrammatical. This last explanation is dispreferred because it would predict that there should not be any difference in reading time at the regions following the NPI if this was the case. Even if the regions following the NPI do not show significant differences within the pair-wise comparisons for the NPI conditions, the changes in reading times between regions suggest that participants are paying attention to sentences that contain NPIs and that the reasons for this effect might be different.

4. Conclusions and further work

Results from the behavioral experiment in Basque show that the parser is sensitive to agreement information in the sentence; and that the agreement information at the complementizer in Basque is used to license different constructions. Reading time differences surfaced at the complementizers when the agreement was wrong and when there was no NPI in the preceding context. Moreover, these differences were maintained over several regions.

Overall, agreement seems to provide more insight into how the timing in the licensing of these constructions proceeds and it seems to be a more reliable cue to detect unlicensed constructions than the presence of an NPI.

Basque speakers showed sensitivity to the agreement processes associated with the embedded complementizer but the presence of an NPI in the embedded clause obviated this effect. In order to license the NPI, the whole structure and semantics of the sentence need to be taken into consideration. Accordingly, when the NPI preceded the agreement information, the ungrammaticality did not seem to be as rapidly recognized at the NPI as when there was a proper name. This might be related to the fact that proper names were more costly to process in this experiment than NPIs.

Taken as a whole, this study provides further insight into the kinds of cues that are used in NPI licensing, when, besides the NPI and the licensor, licensing of the whole sentence requires a specific agreement to be present within the clause where the NPI is contained.

All in all, further work needs to be done with respect to several factors in our experimental design. Among these factors are: to include NPIs that appear in grammatical sentences, to include referential NPs that are less costly than proper nouns, and to test contexts where there is negation in the main clause so that there is a licensor. Furthermore, we expect future ERP results from Basque will shed further insight into the kind of components generated in NPI licensing and will be able to make further distinctions with respect to what particular semantic or tsyntactic process is involved in these dependencies and what the specific role of the complementizer agreement is in the licensing of these constructions.

5. Acknowledgements

This work is part of a British Academy-funded project (project number: H5027100) and a post-doctoral research grant awarded to the first author (LP) by the Basque Government. We would like to thank the organizers of Brain Talk for providing us with the opportunity to present this work and Jon Ortiz de Urbina, Itziar Laka, Myriam Uribe-Etxebarria, Nina Kazanina, Masaya Yoshida, Manolo Carreiras, Patrick Sturt, Peter beim Graben, Theo Marinis, Itziar San Martin and Urtzi Etxeberria for their helpful comments and their valuable help at various stages of this project. All the errors, of course, remain ours.

References

Aoshima, S. Phillips, C. & Weinberg, A. (2004). Processing long-distance dependencies in a head-final language. *Journal of Memory and Language*, *51*, 23–54.

Crain, S. & Fodor, J. (1985). How can grammars help parsers. In Dowty, D., Karttunen, L. & Zwicky, A. M. (Eds.), *Natural language parsing. Psychological, computational, and theoretical perspectives* (pp. 94–128). Cambridge: Cambridge University Press.

Drenhaus, H., Saddy, D. & Frisch, St. (2005). Processing negative polarity items: When negation comes through the backdoor. In S. Kepser & M. Reis, (Eds.), *Linguistic evidence – empirical, theoretical, and computational perspectives* (pp. 145–165). Berlin: Mouton de Gruyter.

Drenhaus, H., beim Graben, P., Frisch, St. & Saddy, D. (2006). Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain and Language, 96,* 255–268.

Etxepare, R. (2003). Negation. In Hualde, J. I. & Ortiz de Urbina, J. (Eds.), *A grammar of Basque*. Berlin: Mouton de Gruyter.

von Fintel, K. (1999). NPI-licensing, Strawson-entailment, and context dependency. *Journal of Semantics 16*, 97–148.

Giannakidou, A. (2002). Licensing and sensitivity in polarity items: From downward entailment to (non)veridicality. In M. Andronis, A. Pycha & K. Yoshimura (Eds.), *CLS 38: Papers from the 38th Annual Meeting of the Chicago Linguistic Society*.

Van Gompel, R. G. & Liversedge, S. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 128–139.

Gordon, P. C. & Hendrick (1997). Intuitive knowledge of linguistics coreference. *Cognition, 62*, 325–370.

Gordon, P. C., Hendrick, R., Ledoux, K & Yang, CL. (1999). Processing reference and the structure of language: An analysis of complex noun phrases. *Language and Cognitive Processes, 14*, 353–379.

Hualde, J. I. & Ortiz de Urbina, J. (2003). *A grammar of Basque*. Berlin: Mouton de Gruyter.

Kazanina, N., Lau, E., Liberman, M., Yoshida, M. & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56*, 384–409.

Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis, 25*, 209–257.

Ladusaw, W. A. (1980). *Polarity sensitivity as inherent scope relations*. New York: Garland Publishing.

Laka, I. (1990). *Negation in syntax: On the nature of functional categories and projections.* PhD Dissertation, MIT.

Laka, I. (1993). Negation in syntax: The view from Basque. *Rivista di Linguistica*, *5*(2), 245–273.

Lee, M. W. (2004). Another look at the role of empty categories in sentence processing and grammar. *Journal of Psycholinguistic Research, 33*, 51–73.

Ledoux, K., Gordon, P. C., Camblin, C. & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms and discourse integration. *Memory and Cognition, 35*(4), 801–815.

Linebarger, M. (1980). *The grammar of negative polarity*. PhD Dissertation, MIT.

Pablos, L. (2006). *Pre-verbal structure building in Romance languages and Basque*. PhD Dissertation, University of Maryland.

Pablos, L. & Phillips, C. (2006). The effects of an active-search for clitic pronouns in left-dislocated constructions in Spanish. Ms. University of Maryland.

Saddy, D., Drenhaus, H. & Frisch, St. (2004). Processing polarity items: Contrastive licensing costs. *Brain and Language, 90*, 495–502.

Stowe, L. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes 1*, 227–245.

Sturt, P., Lombardo, V. & Betancort, M. (2004). Evidence against even mild delay models of structure building. *AMLaP Conference*, Aix-en-Provence.

Uribe-Etxebarria, M. (1994). *Interface licensing conditions on negative polarity items: A theory of polarity and tense interactions*. PhD dissertation, University of Connecticut.

Vasishth, S., Drenhaus, H., Saddy, D. & Lewis, R. (2005). Processing negative polarity. *Paper presented at the 18th Annual CUNY Conference on Human Sentence Processing,* Tucson, AZ.

Vasishth, S., Drenhaus, H., Lewis, R. & Saddy, D. (2006). Processing constraints on negative and positive polarity. *Poster presented at the 19th Annual CUNY Conference on Human Sentence Processing,* New York.

Vasishth, S., Bruessow, S., Lewis, R. L. & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science 32* (4).
Xiang, M., Dillon, B. & Phillips, C. (2006). Testing the effect of the spurious licensing effect for negative polarity items. *Paper presented at the 19th Annual CUNY Conference on Human Sentence Processing,* New York.
Xiang, M., Dillon, B. & Phillips, C. (2009). Illusory licensing effects across dependency types. *Brain and Language, 108(1)*.

When semantic P600s turn into N400s

On cross-linguistic differences in online verb-argument linking

Matthias Schlesewsky

Department of English and Linguistics, Johannes Gutenberg-University Mainz

Ina Bornkessel-Schlesewsky

Research Group Neurotypology, MPI for Human Cognitive and Brain Sciences
Department of Germanic Linguistics, University of Marburg

Abstract

The literature on the neurocognition of language has recently seen a debate about "semantic P600" effects, which have been used to challenge established views about both the functional interpretation of language-related ERP components and the architecture of the human language comprehension system. Here, we show that the semantic P600 phenomenon appears to be language-specific: whereas semantic P600s have been consistently observed in English and Dutch, the present study revealed an N400 for a comparable manipulation in German. On the basis of this finding, we argue that semantic P600s are specific to languages with a fixed word order, in which verb-argument linking is solely conditioned by the linear order of the arguments. In languages in which linking draws upon a richer variety of different information types, by contrast, N400 effects are observable instead. We show how this data pattern can be derived in a neurocognitive model of cross-linguistic language comprehension, the extended Argument Dependency Model (eADM).

1. Introduction

In recent years, one of the "hot topics" in the cognitive neuroscience of language has lain in the examination of so-called "semantic P600" effects. The P600, a centro-parietal positivity with a peak latency of approximately 600 ms, is one of the best known electrophysiological correlates of language comprehension. Following its first observation in the early 1990s (Hagoort et al., 1993; Osterhout & Holcomb, 1992, 1993), it was typically described as an index of syntactic processing. The empirical basis for this functional interpretation was twofold: on the one hand, the P600 was first observed in response to a range of syntactic anomalies and for the dispreferred resolution of syntactic ambiguities ("garden paths"); on the other, it was qualitatively distinct from the classic electrophysiological correlate of semantic processing, the so-called N400 (a centro-parietal negativity with a peak latency of approximately 400 ms, which is engendered by semantically anomalous or unexpected words (Kutas & Hillyard, 1980; see also Kutas & Federmeier, 2000)).

The assumption of a clear-cut functional subdivision between the N400 and P600 components ("semantics" versus "syntax") is contradicted by a number of more recent event-related potential (ERP) studies from English and Dutch, which have reported P600 effects in response to seemingly semantic manipulations. For example, Kim and Osterhout (2005) observed a P600 at the position of the verb in sentences like *The hearty meals were devouring …* (note that here and in the following, the critical sentence positions at which ERP effects of interest were elicited are underlined). Here, the syntactic structure unambiguously calls for the hearty meals to be analysed as the subject of the sentence and, hence, as the Agent of *devour*. Yet the implausibility of this interpretation did not engender an N400, but a P600. Since a follow-up experiment revealed an N400 for similar sentences in which there was no semantic relation between the subject and the verb (e.g. *The dusty tabletops were devouring …*), Kim and Osterhout argued that a strong semantic attraction between a predicate and an argument may lead the processing system to analyse a syntactically well-formed sentence as syntactically ill-formed, thereby eliciting a P600. They thus concluded that "at least under certain conditions, semantics (and not syntax) is 'in control' of how words are combined during on-line sentence processing" (Kim & Osterhout, 2005, p. 216).

While other researchers have proposed alternative functional explanations for semantic P600 effects (see Section 1.1 for further details), there has been a certain consensus that these effects challenge established accounts of sentence comprehension and particularly the idea of an initial stage of syntactic structure building (e.g. Frazier, 1978; Frazier & Rayner, 1982; Frazier & Clifton, 1996; Friederici, 2002). In the present paper, we will argue that this assumption does not hold – at least not across the board. We will begin by illustrating (in Section 1.2) how semantic P600 effects can be derived within a neurocognitive comprehension model with a syntax-first architecture, the extended Argument Dependency Model (eADM). Notably, the eADM differs from classic syntax-first models with respect to a number of key architectural considerations, most of which derive from the adoption of a cross-linguistic perspective on the comprehension process. As we shall discuss in Section 1.3, one of the crucial predictions arising from these assumptions is that semantic P600 effects are a language-specific phenomenon which should be restricted to fixed word order languages such as English and Dutch. In Section 2, we show that this prediction is indeed borne out by demonstrating that semantic P600-type manipulations engender N400 effects in German. Finally, Section 3 discusses these results and their consequences for the language comprehension architecture.

1.1. The semantic P600: A challenge to established language processing models?

The findings by Kim and Osterhout (2005) which were described above in fact form part of a larger body of results on semantic P600 effects. The relevant findings are summarised in Table 1. As is apparent from the table, semantic P600 effects are consistently elicited when a plausibility mismatch occurs between a verb and an argument which would have made for a highly plausible filler of a different role assigned by the same verb (e.g. when the syntactic structure forces a subject – rather than an object – reading of *eggs* in the context of *eat* or of *meals* in the context of *devour*). In line with Kim and Osterhout's (2005) explanation of semantic P600 effects (see above), this basic result could be taken as evidence that online sentence interpretation is primarily semantically-driven, with a P600 resulting when "semantic attraction" leads to an ungrammatical analysis.

A somewhat related proposal, though involving a different functional interpretation of the semantic P600, was put forward by Kolk and colleagues (Kolk et al., 2003; van Herten et al., 2005, 2006; Vissers et al., 2006; Vissers et al., 2008). These authors argue for a plausibility heuristic which serves to construct the most plausible combination of sentence constituents and which acts in parallel with (algorithmic) syntactic analysis. A conflict between the two analyses is thought to engender a P600. From this perspective, the P600 is thus not envisaged as a correlate of a syntactic processing problem, but of a more general process of "conflict monitoring". The conflict monitoring perspective is supported by several empirical findings. Firstly, van Herten et al. (2005) showed that the semantic P600 effects observed in Dutch by Kolk et al. (2003) are not likely due to a syntactic mismatch. Secondly, as demonstrated by van Herten et al. (2006), the likelihood for an implausible sentence to engender a P600 rather than an N400 increases when it contains a plausible chunk (e.g. *climb a ladder* or *prune a tree*). Under these circumstances, conflicting response tendencies ("the sentence is plausible" versus "the sentence is implausible") are likely to emerge. Finally, as predicted by the conflict monitoring account, P600 effects are also elicited by conflicts in other linguistic domains, for example in the processing of orthographic errors (Vissers et al., 2006) or in sentence-picture mismatches (Vissers et al., 2008).

An alternative way of approaching semantic P600 effects has lain in the emphasis of thematic processing aspects (Kuperberg et al., 2003; Hoeks et al., 2004; Kuperberg et al., 2007). Here, semantic P600s are interpreted as evidence for an animacy-based assignment of thematic roles, which may contradict the role assignments required by the syntax (Kuperberg et al., 2003; Kuperberg et al., 2007). Arguably the strongest piece of evidence in favour of such an account was presented by Kuperberg et al. (2007), who

Table 1. Summary of semantic P600 findings in the literature.

Reference	Language	Examples	Observed effect
Kolk et al. (2003)	Dutch	*De vos die op de stropers joeg …* the fox that at the poachers hunted *De bomen die in het park speelden …* the trees that in the park played	P600 N400, P600
Kuperberg et al. (2003)	English	*For breakfast, the eggs would only eat …* *For breakfast, the boys would only bury …*	P600 N400
Hoeks et al. (2004)	Dutch	*De speer heeft de atleten geworpen.* the javelin has the athletes thrown *De speer heeft de atleten opgesomd.* the javelin has the athletes summarised	P600 N400, P600
Kim and Osterhout (2005)	English	*The hearty meals were devouring …* *The dusty tabletops were devouring …*	P600 N400
van Herten et al. (2005)	Dutch	*De vos die op de stropers joeg…* the fox that at the poachers hunted *De vos die op de stroper joeg…* the fox that at the poacher hunted	P600 P600
Kuperberg et al. (2006)	English	*For breakfast, the eggs would eat …* *For breakfast, the boys would eats …*	P600 P600
van Herten et al. (2006)	Dutch	*…dat de olifanten de bomen snoeiden…* …that the elephants the trees pruned… '…that the elephants pruned the trees…' *…dat de olifanten de bomen verwenden…* …that the elephants the trees caressed… '…that the elephants caressed the trees…'	P600 N400, small P600
Kuperberg et al. (2007)	English	*For breakfast, the eggs would eat …* *For breakfast, the eggs would plant …*	P600 P600

Key findings
P600 for an implausible relation between subject and object (no animacy mismatch) with a semantically related verb; N400 and (task-dependent) P600 for an animacy mismatch betwen a subject and a semantically unrelated verb
P600 for an animacy mismatch between a subject and a semantically related verb; N400 for a "pragmatically" unexpected (semantically unrelated) verb without an animacy mismatch
P600 for an implausible relation between subject and object (animacy mismatch) with a semantically related verb; N400 and P600 for an animacy mismatch (subject and object) with a semantically unrelated verb
P600 for an animacy mismatch between a subject and a semantically related verb; N400 for an animacy mismatch betwen a subject and a semantically unrelated verb
Replication of Kolk et al. (2003) P600 was also observable when the two arguments did not differ in terms of number, thus suggesting that the effect is not due to a morpho-syntactic mismatch
Replication of Kuperberg et al. (2006) The semantic P600 showed a similar distribution but smaller amplitude than the P600 elicited by a morphosyntactic violation (number agreement mismatch); it was also affected by verb transitivity and the plausibility of the passivised version of the exp. sentence
P600 for implausible sentences containing a highly plausible fragment (e.g. prune the trees) N400 (and small P600) for sentences without such a fragment
P600 for an animacy mismatch between a subject and a semantically related verb (replication of earlier findings); same effect with a semantically unrelated verb.

showed a P600 effect for subject-verb animacy mismatches even in the absence of a semantically related verb. This finding appears difficult to derive via a plausibility-based approach and would thereby seem to call for a thematically-based account involving animacy as a crucial factor. Conversely, however, an account along these lines is challenged by a number of the findings obtained in the context of the conflict monitoring approach such as the observation of P600 effects in implausible sentences containing highly plausible chunks or in domains that are clearly external to thematic processing (e.g. orthography).

To accommodate these challenges, Kuperberg (2007) proposed an account of semantic P600 effects which assumes the interaction of multiple parallel processing streams: syntactic processing, thematic processing and associative memory-based processing. From this perspective, the P600 is thought to reflect a "continued combinatory analysis" (2007, p. 37) which is required in the case of a conflict between the different streams. The likelihood that such processes will be initiated is further thought to be influenced by the discourse context and the task performed by participants, both of which may render an in-depth analysis of plausibility and, hence, additional processing if required, more likely.

In spite of the differences between the various accounts of semantic P600 effects (for a more detailed recent review, see Bornkessel-Schlesewsky & Schlesewsky, 2008), they all have in common that they view semantic P600 effects as a challenge to established architectures of sentence comprehension. For example, Kuperberg (2007, p. 44) argues that "the outcome of a semantic memory-based analysis and possibly a semantically-driven combinatorial thematic analysis can, under some circumstances, temporarily dominate online comprehension, even in simple, canonical unambiguous sentences." Similarly, van Herten et al. (2006, p. 1194) conclude that their results are "incompatible not only with syntax-first but also with constraint-based" processing models and that they provide "compelling evidence for the existence of a plausibility strategy".

The logic behind arguments such as these is as follows: though syntax-first and constraint-based processing models offer different views on the relative timecourse of when syntactic and semantic information influences the comprehension process, both types of approaches assume that different sentence-level interpretations result from different syntactic structures. Thus, the classic view in sentence processing is that syntax determines interpretation, though it is controversial whether the choice of the preferred structure can be affected by non-syntactic information types (as assumed in contraint-based models, e.g. MacDonald et al., 1994) or not (as assumed in syntax-first models, e.g. Frazier, 1987). With respect to ERP correlates of sentence processing, a vast number of findings from the 1980s and 1990s suggested that when a particular syntactic structure unambiguously forces an implausible reading, an N400 results (e.g. in *He spread the warm bread with socks*, from Kutas & Hillyard, 1980). Hence, the observation that implausibilities in simple, unambiguous sentences engender P600 effects under certain circumstances appears to challenge the standardly held assumption that syntactic structure determines sentence-level interpretation. A popular way of approaching this issue has been to assume an alternative ("heuristic") route to interpretation, which is typically construed as an efficient, but possibly inaccurate (i.e. "quick and dirty") means of attaining a sentence-level meaning which applies in lieu of a full, algorithmic syntactic analysis (for a recent overview, see Ferreira & Patson, 2007).

1.2. Deriving semantic P600 effects in a "syntax-first" model: The eADM

We recently argued (Bornkessel-Schlesewsky & Schlesewsky, 2008) that semantic P600 effects can also be derived in an independently motivated sentence processing architecture of a somewhat different kind, the extended Argument Dependency Model (eADM; Bornkessel, 2002; Schlesewsky & Bornkessel, 2004; Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2009b). The eADM posits a syntax-first architecture, but differs from classic syntax-first models in that the syntactic structure does not determine sentence interpretation (cf. Figure 1). This assumption is based on cross-linguistic considerations: the diversity of the languages of the world calls into question whether syntactic features indeed universally determine sentence interpretation. Consider the question of role identification, i.e. how the language comprehension system may determine "who is acting on whom". Whereas this issue is settled via information types that are traditionally considered "syntactic" in most of the languages that have been subjected to extensive psycholinguistic investigation (e.g. via word order in languages like English and case marking in languages like German), it is driven by non-syntactic features such as animacy and person in others (for a detailed discussion from a psycholinguistic perspective, see Bornkessel-Schlesewsky & Schlesewsky, 2009b). On the basis of this observation, the eADM proposes that all of these information types play a functionally equivalent role during the online interpretation of core sentence constituents (i.e. nouns and verbs). This assumption is summarised in the following hypothesis (Bornkessel-Schlesewsky & Schlesewsky, 2009b):

The interface hypothesis of incremental argument interpretation (IH)

> Incremental argument interpretation (i.e. role identification and assessment of role prototypicality) is accomplished by the syntax-semantics interface, i.e. with reference to a cross-linguistically defined set of prominence scales and their language-specific weighting. The relevant prominence scales are:
>
> - morphological case marking (nominative > accusative / ergative > nominative)
> - argument order: (argument 1 > argument 2)
> - animacy (+animate > –animate)
> - definiteness/specificity (+definite/specific > –definite/specific)
> - person (+1st/2nd person > –1st/2nd person)

Within the eADM, argument interpretation is accomplished via the assignment of the generalised semantic roles "Actor" and "Undergoer", which are based on the agent and patient prototypes, respectively. Which argument is assigned which role is determined on the basis of the prominence scales in the IH and their language specific weighting. Hence, the interpretation of an argument is logically independent of that argument's position in the syntactic structure. Moreover, the degree of competition be-

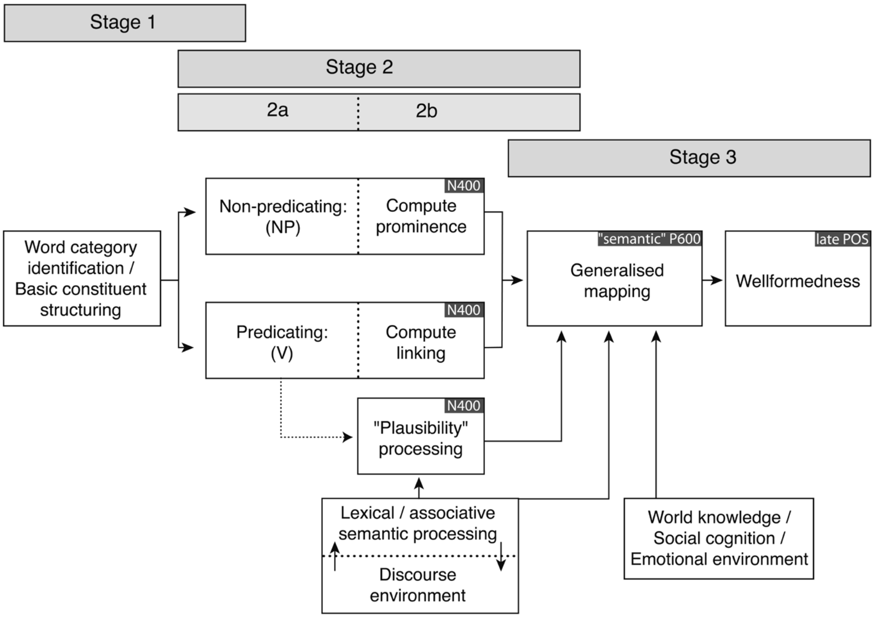


Figure 1. The architecture of the extended Argument Dependency Model (from Bornkessel-Schlesewsky & Schlesewsky, 2008).

tween the roles is assessed according to how many overlapping features the arguments bear, with complete feature distinctness amounting to a prototypical transitive relation. When the verb is encountered, the Actor and Undergoer arguments are mapped onto the variables in a decomposed semantic representation (the "logical structure", e.g. `do'(x, CAUSE(x, BECOME(broken'(y))))` for *break*, see Van Valin, 2005) according to a mapping algorithm (cf. Bornkessel-Schlesewsky & Schlesewsky, 2009b, for details). These two steps of Actor/Undergoer role assignments and argument-verb linking lie at the heart of sentence-level interpretation in the eADM and are captured by the Compute Prominence and Compute Linking steps in stage 2 of the model (cf. Figure 1), respectively.

In a second processing stream of stage 2 which operates in parallel to, but separately from, Compute Prominence and Compute Linking, the comprehension system undertakes a combination of words on the basis of their lexical properties. This processing step was dubbed "Plausibility processing" by Bornkessel-Schlesewsky and Schlesewsky (2008) for want of a more fitting label (for earlier and much less detailed discussions of an analogous step in the processing architecture, see Bornkessel, 2002; Schlesewsky & Bornkessel, 2004). It essentially takes major categories as an input and determines the way in which they might be combined most plausibly. This decision may be based on lexical-semantic associations or on specific lexical information such as qualia properties

(Pustejovsky, 1995). In this way, the eADM's Plausibility processing step is also conceptually related to the "thematic processor" of Rayner et al. (1983).

The interpretive steps in stage two of processing are preceded by a stage involving word category identification and basic constituent structuring. (Note that the hierarchically ordered processing steps within the eADM are organised in a cascaded rather than in a strictly serial manner; for details, cf. Bornkessel-Schlesewsky & Schlesewsky, 2009a, 2009b). On the one hand, this first stage of processing serves to build up a basic phrase structure which determines how and in what order word categories (and phrasal categories) may be combined with one another. On the other, it provides information as to whether the input element currently being processed is predicating (typically a verb) or non-predicating (typically a noun phrase), thereby helping to determine whether the Compute Prominence or Compute Linking step is applicable in stage 2. Hence, due to this initial word category-based processing stage, the eADM is a "syntax-first" model. However, as already discussed in detail above, it contrasts with typical syntax-first architectures in that the syntactic structure built up in stage 1 of processing does not determine sentence-level interpretation.

Finally, in stage 3 of the eADM, the outputs of the different processing streams from phase 2 (Compute Prominence or Compute Linking and Plausibility Processing) are mapped onto one another and, at the same time, integrated with a range of further information types (e.g. emotional and social aspects of the current discourse environment). This is accomplished by the Generalised Mapping step which, in turn, is followed by a task- and environment-dependent evaluation of the current utterance's well-formedness (Wellformedness Processing).

Crucially, due to the eADM's hierarchical architecture, a fatal problem in an earlier processing step can block the application of a later step. Thus, just as a failure of stage 1 processing leads to a blocking of stage 2 (see Bornkessel & Schlesewsky, 2006, for a discussion of this issue in general terms; and Bornkessel-Schlesewsky & Schlesewsky, 2009b, for a discussion of the role of the cascaded architecture in this process), a general failure of either Compute Linking (or Compute Prominence) or Plausibility Processing in stage 2 leads to a blocking of Generalised Mapping in stage 3. The only processing step which is "exempt" from this general rule is the Wellformedness check, since this, by definition, must take into account the success and/or failure of previous processing stages.

Now consider how these components of the eADM's processing architecture account for semantic P600 effects. For purposes of illustration, we will use examples from Hoeks et al. (2004), which are repeated in [1] for convenience.

[1a]	De	speer	heeft	de	atleten	<u>geworpen</u>.
	the	javelin	has	the	athletes	thrown
	'The javelin has thrown the athletes.'					

[1b]	De	speer	heeft	de	atleten	<u>opgesomd</u>.
	the	javelin	has	the	athletes	summarised
	'The javelin has summarised the athletes.'					

Before the verb is encountered, the two arguments *speer* 'javelin' and *atleten* 'athletes' are assigned the Actor and Undergoer roles, respectively, via their relative linear position (which, as in English, is the dominant factor for Compute Prominence in Dutch). When the verb is reached in [1a], Compute Linking maps the Actor role onto the x-argument and the Undergoer role onto the y-argument in the logical structure (LS) of the verb throw `(do'(x,throw'(x,y)))`. Simultaneously, the Plausibility Processing step registers that both *javelin* and *athletes* are plausible arguments of *throw* and that the most plausible combination of these items involves an interpretation of *athletes* as the Actor and *javelin* as the Undergoer of the action described. However, when the outputs of Compute Linking and Plausibility Processing are mapped onto one another in the Generalised Mapping step, a mismatch arises because of the conflicting role assignments reached by the two processing steps. Hence, Generalised Mapping registers a processing conflict, which is reflected in a P600.

Now consider what happens in the case of [1b]. Initially, Actor and Undergoer roles are assigned as described above. Compute Linking also proceeds without problems when the verb is reached, since linear order is the sole determinant of linking in a language of this type. In contrast to [1a], however, Plausibility Processing fails because *javelin* is not a plausible argument of *summarise*. Hence, an N400 is engendered and Generalised Mapping is blocked. Any late positivities engendered in sentences of this type are thus due to the Wellformedness check, as supported by the observation that such positivities are task-dependent and more likely to surface in the context of an acceptability judgement task (Kolk et al., 2003).

In this way, the processing architecture of the eADM – which was proposed independently of the semantic P600 debate – straightforwardly derives the basic finding of a semantic P600 in the presence of a semantic association, as opposed to an N400 in the absence of such an association. For a detailed discussion of further examples and of how the eADM accounts for the absence of an N400 in Kuperberg et al.'s (2007) semantically unrelated sentences with an animacy mismatch, see Bornkessel-Schlesewsky and Schlesewsky (2008).

1.3. Cross-linguistic predictions

Crucially for present purposes, the eADM's account of semantic P600 effects makes an interesting cross-linguistic prediction. Recall first of all the two key assumptions in the explanation of these effects, namely that there is no problem in linking but a problem in plausibility processing, thereby leading to a conflict in Generalised Mapping, which is reflected in a P600. However, the assumption that an animacy mismatch does not lead to a linking problem is predicted to be language-specific: animacy is not linking-relevant in English or Dutch, since the strict word order completely dominates the linking process. By contrast, animacy is predicted to play a role in a closely related language, namely German. Consider [2a-d]:

[2a]	… dass	dem Arzt	der Abt	gefällt.
	… that	[the doctor]:DAT	[the abbot]:NOM	is.pleasing.to
	'… that the doctor finds the abbot pleasing.'			

[2b]	… dass	der Abt	dem Arzt	droht.	
	… that	[the abbot]:NOM	[the doctor]:DAT	threaten	
	'… that the abbot threatens the doctor.'				

[2c]	… dass	dem Arzt	der Bericht	gefällt.	
	… that	[the doctor]:DAT	[the report]:NOM	is.pleasing.to	
	'… that the doctor finds the report pleasing.'				

[2d]	… dass	der Bericht	dem Arzt	gefällt	/ #droht.
	… that	[the report]:NOM	[the doctor]:DAT	is.pleasing.to	/ #threatens
	'… that the doctor finds the report pleasing / the report threatens the doctor.'				

As illustrated in [2], in German, animacy influences linking preferences in bivalent constructions with a nominative and a dative argument. In sentences with two animate arguments ([2a–b]), the nominative argument could either be the Undergoer [2a] or the Actor [2b]. In this case, the word order preference (dative-before-nominative in [2a]; nominative-before-dative in [2b]) depends on how this ambiguity is resolved by the particular verb encountered. By contrast, when the nominative argument is inanimate and the dative argument is animate, the nominative must be the Undergoer, as illustrated by the impossibility of completing [2d] with the nominative-Actor verb threaten (indicated by "#"). Hence, under certain circumstances, animacy steps in to mediate the linking process in German.

Assuming, then, that linking involves an evaluation of animacy features in German, we predict a fundamentally different pattern of results for semantic P600 manipulations involving an animacy violation. In contrast to what happens in English and Dutch, German sentences analogous to [1] should already lead to a conflict in the Compute Linking step. According to the eADM, this should elicit an N400 (see Figure 1). The aim of the present study was to test this prediction.

2. The present study

2.1. Experimental design

As outlined in Section 1.3 above, the aim of the present ERP study was to test the hypothesis that "semantic P600 manipulations" should lead to N400 effects in German. To this end, we employed the experimental design shown in Table 2.

The first two conditions in Table 2 (US/UO) correspond to the classic semantic P600 sentence types used previously in English and Dutch (see Section 1.2 and Table 1 for details). Both include an inanimate and an animate argument followed by a verb requiring an animate subject (and for which the inanimate NP is a plausible object). Condition US (i.e. "unambiguous subject-first") is implausible because the nominative case marking on the inanimate argument forces a subject (or Actor) reading for this argument. Condition UO (i.e. "unambiguous object-first") is the plausible control

Table 2. Examples of the four critical sentence conditions used in the present ERP study as well as mean acceptability/plausibility ratings and reaction times for each condition (by-participant standard deviations are given in parentheses). Hash marks ("#") indicate implausible sentences. NOM: nominative case; ACC: accusative case; U: unambiguous; A: ambiguous; S: subject-initial; O: object-initial; APR: acceptability/plausibility rating; RT: reaction time.

Cond.	Example	Mean APR (%)	Mean RT (ms)
US	… dass der Schalter den Techniker bedient. … that [the switch]:NOM [the technician]:ACC operates # '… that the switch operates the technician.'	17.89 (15.38)	622.02 (227.04)
UO	… dass den Schalter der Techniker bedient. … that [the switch]:ACC [the technician]:NOM operates '… that the technician operates the switch.'	78.62 (14.11)	601.20 (212.52)
AS	… dass Techniker Schalter bedienen. … that technicians switches operate '… that technicians operate switches.'	85.51 (10.99)	552.26 (192.02)
AO	… dass Schalter Techniker bedienen. … that switches technicians operate '… that technicians operate switches.'	37.79 (27.70)	617.03 (226.45)

condition, in which the inanimate argument is marked for accusative case and thereby identified as an object / Undergoer.

Two further conditions (AS/AO) were included into the experimental design, in order to manipulate the strength of the grammatical evidence in favour of an implausible reading. Even though German allows object-initial orders and these are even preferred over their subject-initial counterparts under certain circumstances (Haupt et al., 2008; Grewe et al., 2006), it is well know that, in the absence of unambiguous case marking, the language processing system initially adopts a subject-initial analysis (the so-called "subject-first preference", e.g. Schriefers et al., 1995; Bader & Meng, 1999; Schlesewsky et al., 2000). Thus, whereas condition AS (i.e. "ambiguous subject-initial") should lead to a plausible reading, condition AO (i.e. "ambiguous object-initial") should induce a mismatch at the position of the verb because the subject-first preference calls for an Actor reading of NP1, whereas the animacy information calls for an Actor reading of NP2. By including these "locally implausible" sentences, we can therefore investigate whether the ERP response to the "globally implausible" sentence condition US is due to an initial linking mismatch or whether it is conditioned by the fact that the conflict is irresolvable. In the first case, we should expect to observe very similar ERP responses to US and AO. In the second, the two conditions should differ from one another.

2.2. Materials and methods

Participants

Forty-one native speakers of German participated in the present study (28 women, 13 men; mean age: 24.3 years; age range: 19–30 years). All participants had been raised monolingually, were right handed (according to an adapted German version of the Edinburgh handedness inventory, Oldfield, 1971) and had normal or corrected-to-normal vision. A further seven participants were excluded from the final data analysis due to excessive EEG artefacts.

Materials

Sixty sets of the four critical conditions (cf. Table 2) were constructed. Sentences always began with a matrix clause of the form Adverb–Auxiliary–Participle (e.g. *Heute wurde berichtet* …, 'Today (it) was reported …'). The 240 sentences thus resulting were subsequently distributed over two lists of 120 sentences each, with each list containing 30 items per condition and two items from each lexical set. The items on each list were pseudo-randomly interspersed with a further 180 items, 120 of which instantiated a different experimental manipulation (involving two animate NPs and plausible and implausible sentences), and 60 of which were plausible filler sentences with inanimate subjects and animate objects. These fillers served to ensure that participants could not decide whether a sentence was plausible or implausible prior to reading the verb.

Procedure

Participants were seated in a dimly lit, acoustically shielded room, approximately 1 m in front of a 17-inch computer screen. Sentences were presented in the centre of the screen in a phrase-by-phrase manner (i.e. determiners and nouns were presented together; for all other items, the presentation was word-by-word). Each trial began with the presentation of a fixation asterisk (presentation time: 300 ms; inter-stimulus-interval, ISI: 200 ms) preceding the sentence proper. Noun phrases were presented for 500 ms, while other elements were presented for 400 ms. The ISI was always 100 ms. Following 1000 ms of blank screen after the sentence-final word, three question marks appeared in the centre of the screen. These signalled to participants that they were required to judge whether the sentence that they had just read was an acceptable and plausible sentence of German. Participants responded using two hand-held push-buttons; the assignment of acceptable/plausible and unacceptable/implausible to the left and right buttons was counterbalanced across participants. After a participant had responded or the maximal reaction time of 2000 ms had expired, there was a 1500 ms inter-trial-interval before the beginning of the next trial. Participants were asked to refrain from moving and from blinking their eyes during the presentation of a sentence (i.e. between the presentation of the asterisk and the onset of the acceptability judgement task).

The 300 experimental sentences were presented in 6 blocks of 50 sentences each. Between blocks, participants took short breaks. Prior to the experiment proper, participants completed a short practice run comprising 10 sentences. The entire experimental session lasted approximately 2.5 hours including electrode preparation.

EEG recording and preprocessing

The EEG was recorded by means of 25 Ag/AgCl-electrodes (ground: AFZ). Recordings were referenced to the left mastoid (rereferenced to linked mastoids offline). The electro-oculogram (EOG) was monitored via electrodes at the outer canthus of each eye (horizontal EOG) and above and below the participant's right eye (vertical EOG). Electrode impedances were kept below 5 kΩ. EEG and EOG channels were amplified using a Twente Medical Systems DC amplifier (Enschede, The Netherlands) and recorded with a digitization rate of 250 Hz.

In order to eliminate slow signal drifts, a 0.3–20 Hz band-pass filter was applied to the raw EEG data. Subsequently, average ERPs were calculated per condition per participant from the onset of the clause-final verb to 1000 ms post onset, before grand-averages were computed over all participants. Trials containing EEG or EOG artefacts were excluded from the averaging procedure.

Data analysis

For the behavioural task mean acceptability/plausibility ratings and reaction times were analysed by means of repeated-measures analyses of variance (ANOVAs) involving the condition factors word order (ORDER: subject-initial versus object-initial) and ambiguity (AMB: ambiguous versus unambiguous) and the random factors participants (F1) and items (F2).

For the statistical analysis of the ERP data, repeated-measures ANOVAs were calculated for mean amplitude values per time window per condition. In order to examine the topographical distribution of the effects, the topographical factor region of interest (ROI) was included in the analyses. ROIs were defined as follows: left-anterior (F3, F7, FC1, FC5); left-posterior (CP1, CP5, P3, P7); right-anterior (F4, F8, FC2, FC6); right-posterior (CP2, CP6, P4, P8). For analyses involving more than one degree of freedom in the numerator, significance values were corrected when sphericity was violated (Huynh & Feldt, 1970).

2.3. Results

Judgement task

Mean acceptability/plausibility ratings and reaction times for the four critical conditions employed in the present study are shown in Table 2.

The statistical analysis of the acceptability/plausibility ratings revealed a main effect of AMB ($F_1(1, 40) = 5.12$, $p < .03$; $F_2(1, 59) = 6.16$, $p < .02$) and an interaction AMB x ORDER ($F_1(1, 40) = 87.36$, $p < .001$; $F_2(1, 59) = 193.45$, $p < .001$). The main effect of ORDER only reached significance in the analysis by items ($F_2(1, 59) = 5.15$, $p < .03$). Resolving the interaction by AMB revealed effects of ORDER for both unambiguous ($F_1(1, 40) = 86.86$, $p < .001$; $F_2(1, 59) = 115.53$, $p < .001$) and ambiguous sentences ($F_1(1, 40) = 23.87$, $p < .001$; $F_2(1, 59) = 78.10$, $p < .001$). For the unambiguous conditions, this effect was due to a higher acceptability of object-initial sentences, whereas it resulted from a higher acceptability of subject-initial sentences in the ambiguous conditions. The analysis of the reaction times revealed no significant main effects of interactions.

ERP data

Grand average ERPs at the position of the critical clause-final verb are shown in Figures 2 and 3 for unambiguous and ambiguous sentences, respectively. Visual inspection of the figures suggests that both unambiguous and ambiguous sentences engendered an N400 (for the globally implausible subject-initial condition in the unambiguous case and for the locally implausible object-initial condition in the ambiguous case). In the unambiguous conditions, the subject-initial sentences additionally elicited a late positivity in comparison to their object-initial counterparts. We thus chose two time windows for the statistical analysis: 350–650 ms for the N400 and 700–1000 ms for the late positivity.

For the earlier time window (350–650 ms), a repeated measures ANOVA revealed a main effect of AMB ($F(1, 40) = 8.40$, $p < .001$) and interactions of ROI x AMB ($F(3, 120) = 31.73$, $p < .001$) and AMB x ORDER ($F(1, 40) = 24.80$, $p < .001$). Resolving the interaction ROI x AMB by ROI revealed significant or marginally significant effects of AMB in each region, though the ambiguity effect was most pronounced over the left hemisphere (maximal $F(1, 40) = 36.99$, $p < .001$ in the left-anterior ROI; minimal $F(1, 40) = 3.90$, $p < .06$ in the right-posterior ROI). This effect was due to a (left-lateralised) negativity for ambiguous versus unambiguous sentences. The interaction AMB x ORDER was resolved by AMB, thus revealing effects of ORDER for both unambiguous ($F(1, 40) = 6.11$, $p < .02$) and ambiguous sentences ($F(1, 40) = 23.71$, $p < .001$). For the unambiguous sentences, the order effect resulted from a negativity for subject- versus object-initial sentences, whereas it was due to a negativity for object- versus subject-initial sentences in the ambiguous sentences.

The statistical analysis of the later time window (700–1000 ms) showed a main effect of AMB ($F(1, 40) = 18.67$, $p < .001$) and an interaction AMB x ORDER ($F(1, 40) = 4.48$, $p < .05$). Resolving the interaction by AMB revealed a significant effect of ORDER for unambiguous sentences ($F(1, 40) = 9.20$, $p < .01$), but not for their ambiguous counterparts ($F < 1$). Thus, only unambiguous sentences showed a late positivity (for the implausible subject-initial condition in comparison to the object-initial control).

In order to examine the N400 effect in more detail, we conducted an additional analysis: we computed difference waves for the ambiguous and unambiguous conditions by subtracting the plausible control condition from the critical implausible condition in each case (i.e. subtracting the subject-initial condition from the locally implausible object-initial condition for the ambiguous sentences and the object-initial condition from the globally implausible subject-initial condition in the unambiguous case). These difference waves, which are depicted in Figure 4, allow us to undertake a direct comparison of the animacy mismatch in the ambiguous and unambiguous sentences without having to take the factor word order into account (i.e. whether the mismatch was induced in a subject- or an object-initial structure).

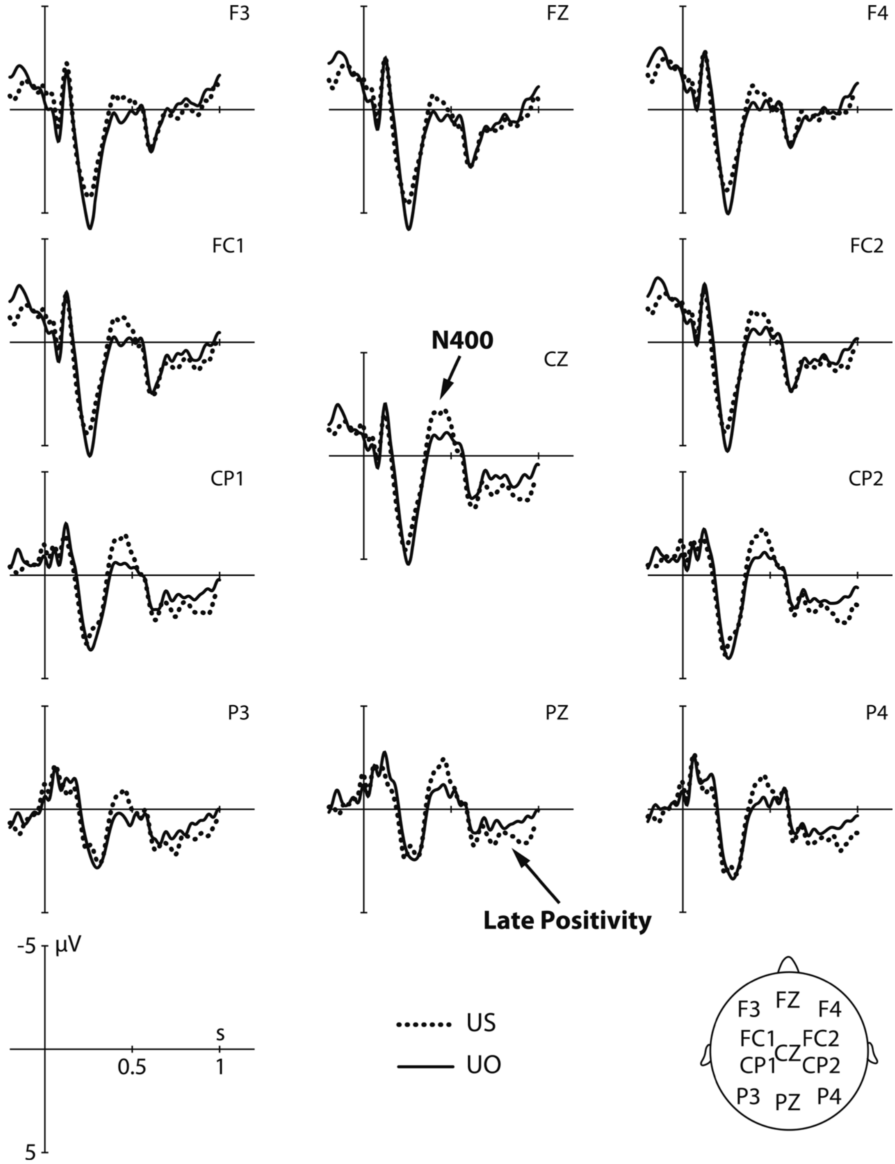


Figure 2. Grand average ERPs at the position of the clause-final verb (onset at the vertical bar) in the object-initial (solid line) and subject-initial (dotted line) unambiguous conditions. Negativity is plotted upwards.

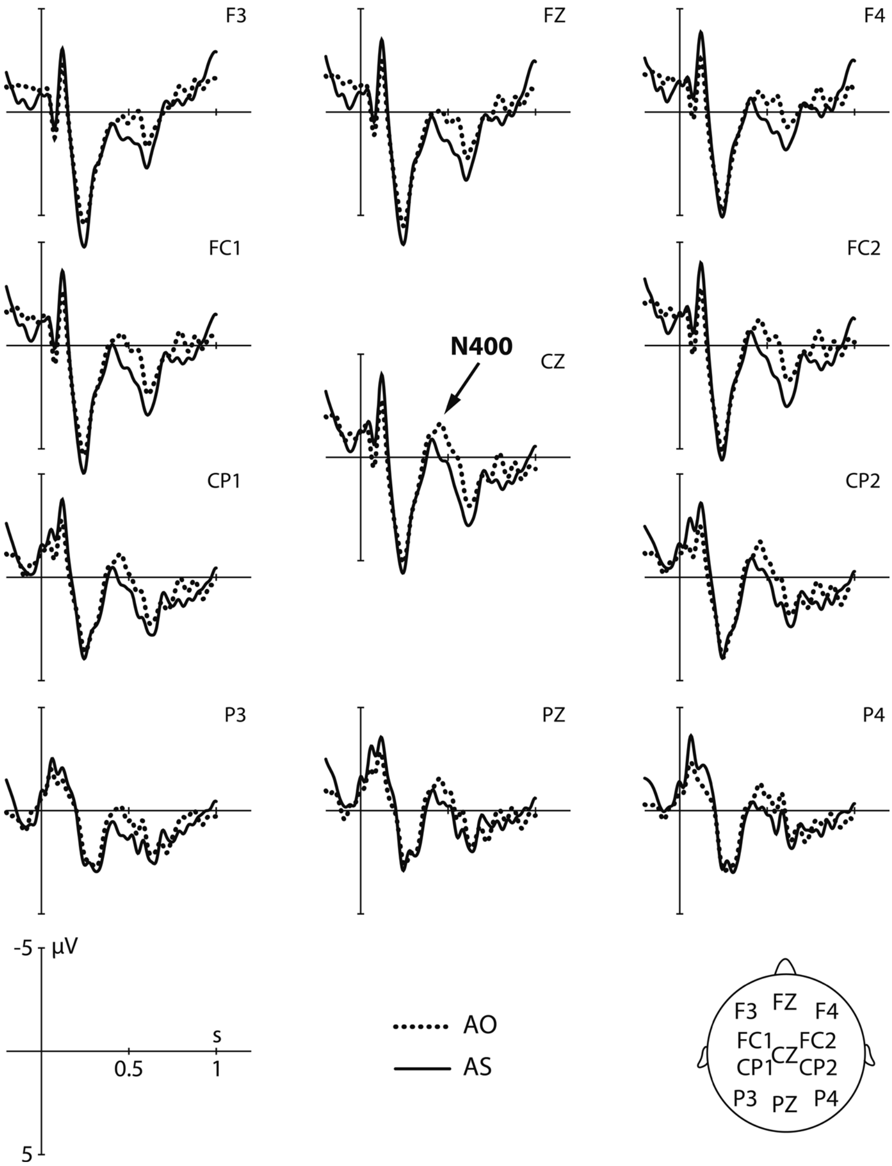


Figure 3. Grand average ERPs at the position of the clause-final verb (onset at the vertical bar) in the subject-initial (solid line) and object-initial (dotted line) ambiguous conditions. Negativity is plotted upwards.

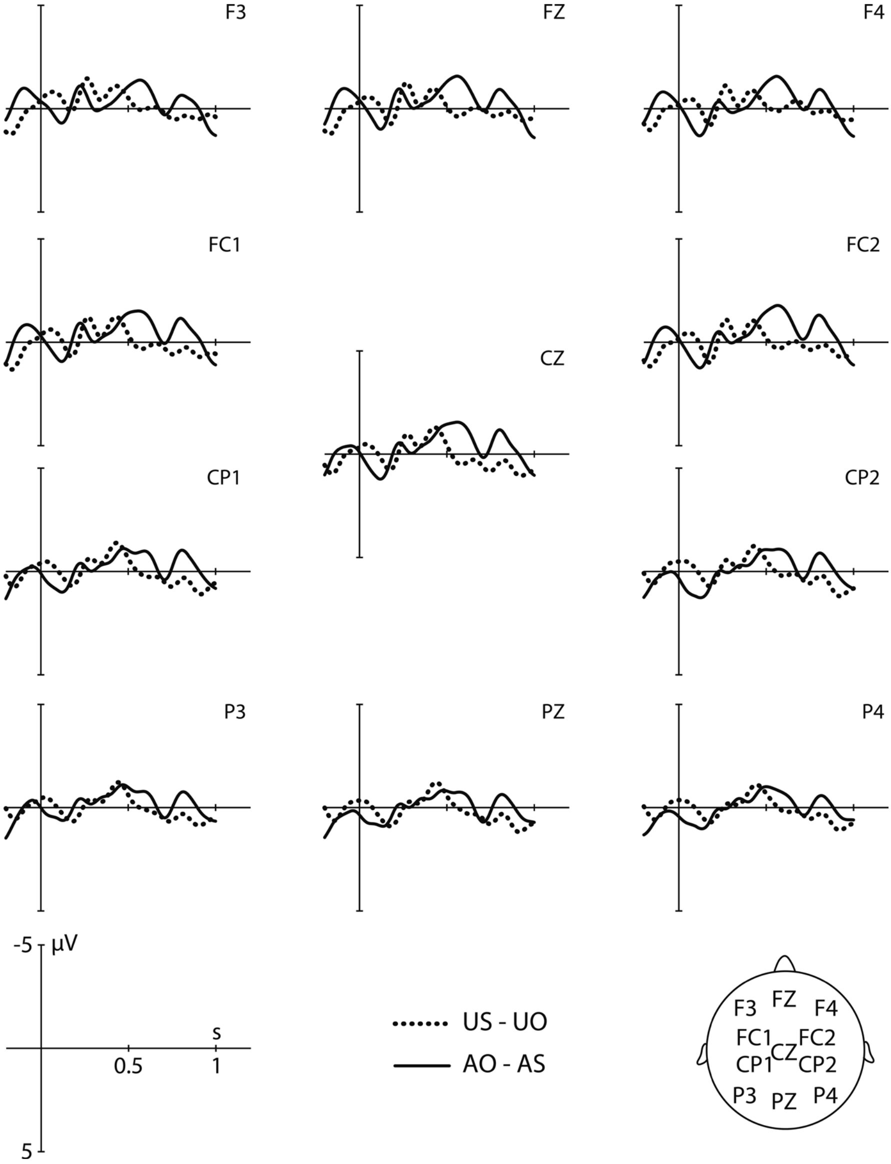


Figure 4. Difference waves for the ambiguous (solid line) and unambiguous (dotted line) conditions. In both cases, the difference wave was constructed by subtracting the (locally) implausible condition from the plausible control condition (ambiguous: subject-initial – object-initial; unambiguous: object-initial – subject-initial). Negativity is plotted upwards.

Visual inspection of Figure 4 suggests that the negativity in the ambiguous sentences differs from that in unambiguous sentences in two respects: it appears to have (a) a longer duration, and (b) a broader topographical distribution (i.e. the negativity for the unambiguous sentences is most readily apparent at centro-parietal sites, whereas the effect for the ambiguous sentences is broadly distributed). These observations were confirmed by a statistical analysis involving the factor AMB (difference between the unambiguous conditions versus difference between the ambiguous conditions) in three consecutive 100 ms time windows: 350–450 ms; 450–550 ms; 550–650 ms. In the first time window (350–450 ms), the two difference wave conditions did not differ from one another (i.e. there was neither a main effect of AMB nor an interaction ROI x AMB). The analysis of the second time window (450–550 ms) showed a marginal interaction of ROI x AMB ($F(3, 120) = 2.46$, $p < .09$), which resulted from a significant effect of AMB in the right-anterior region ($F(1, 40) = 6.07$, $p < .02$; ambiguous structures more negative than unambiguous structure). Finally, in the third time window, there was a significant main effect of AMB ($F(1, 40) = 10.81$, $p < .01$), which resulted from a negativity for ambiguous versus unambiguous conditions. In summary, the analysis of the difference waves confirmed both the longer duration of the negativity for the ambiguous sentences and its broader topographical distribution.

3. Discussion

The present ERP study was the first to employ a semantic P600 manipulation in German, i.e. it examined the processing of simple, unambiguous, transitive sentences which were rendered implausible via an animacy mismatch at the position of the clause-final verb. In contrast to previous studies from Dutch and English, which consistently reported P600 effects for sentences of this type, the present experiment revealed a biphasic N400-late positivity pattern. Using a second set of conditions, we further examined the electrophysiological processing correlates of a "local" animacy-induced plausibility mismatch, i.e. of an animacy conflict which arises in the preferred subject-initial word order, but which can be resolved via a reanalysis towards an object-initial reading. For these ambiguous (i.e. locally implausible) sentences, we also observed an N400 effect, though this effect showed a longer duration and a broader topographical distribution than that for the unambiguous (i.e. globally implausible) sentences. In contrast to the unambiguous sentences, however, the ambiguous conditions did not show a late positivity.

In the following, we will first discuss the findings for the unambiguous conditions and the consequences of these results for a cross-linguistic account of semantic P600 effects (Section 3.1), before comparing the results of the unambiguous and ambiguous conditions (Section 3.2).

3.1. Cross-linguistic differences in semantic P600 effects

The present results confirm the predictions of the eADM by demonstrating that a typical "semantic P600 manipulation" can lead to N400 effects when a language other than English or Dutch is examined. Recall that, in a language of the English type, verb–argument linking is entirely determined by linear order: under no circumstances can a sequence of NP-V-NP be interpreted as Undergoer-V-Actor (cf. the implausibility of a sentence such as *The eggs ate Steve*).[1] Hence, sentences employing a typical semantic P600 manipulation (e.g. *The hearty meals were devouring …*) do not lead to a linking problem at the position of the verb, since the linear order information suffices for linking to proceed successfully. Likewise, there is no problem in the Plausibility Processing step, since *hearty meals* are a plausible argument of *devour*. However, a conflict subsequently arises in Generalised Mapping, where is becomes clear that Compute Linking and Plausibility Processing yielded conflicting outputs. In German, by contrast, animacy can influence argument linking (see Section 1.3). Hence, a sentence such as *… dass der Schalter den Techniker bedient* ('… that the switch operates the technician') already leads to a conflict in the Compute Linking step because animacy and case marking contradict one another. As predicted by the eADM (see Figure 1), this linking conflict engenders an N400.

Why do the implausible unambiguous sentences additionally yield a late positivity? We propose that this effect is a correlate of the irresolvable processing conflict arising in these sentences and the concomitant decrease in Wellformedness. (It cannot be a correlate of Generalised Mapping because this step is blocked whenever one of its input processing steps does not yield an output; see Section 1.2.) This interpretation is supported by the observation that the ambiguous sentences did not give rise to a late positivity (see below for further discussion). It further leads to the testable prediction that the late positivity in the unambiguous sentences should be task-dependent, surfacing primarily in the context of a judgement task. In this way, it should behave analogously to the task-dependent late positivity observed by Kolk et al. (2003). For a detailed discussion of the task-dependent nature of late positivities engendered by the Wellformedness check and how they are distinct from semantic P600s, see Bornkessel-Schlesewsky and Schlesewsky (2008).

The present findings thus demonstrate that the semantic P600 phenomenon is specific to languages of a particular type: it only arises when verb-argument linking is determined entirely by word order, thus eliminating the influence of other features such as animacy.

3.2. Global versus local conflicts: The role of ambiguity

Like the animacy conflict in unambiguous sentences, the local implausibility (required reanalysis towards an object-initial order) in the ambiguous conditions elicited an

[1] Note that this does not mean that the sentence-initial argument must always be interpreted as an Actor in English. Such a generalisation would clearly yield the wrong result for passive sentences. However, the linking regularities in an English sentence are always predictable on the basis of an argument's position in combination with voice information.

N400. This observation is consistent with a number of recent results demonstrating that word order reanalyses in German generally correlate with N400 effects (Bornkessel et al., 2004; Schlesewsky & Bornkessel, 2006; Haupt et al., 2008). In terms of the eADM, this "reanalysis N400" is interpreted in the same way as the N400 for the unambiguous conditions in the present study, namely as a correlate of Compute Linking. It arises because the Actor/Undergoer assignments indicated by the word order of the sentence do not match those required by the animacy of the arguments.

Interestingly, the N400 effect for the ambiguous object-initial condition shows similarities to and differences from that for the unambiguous subject-initial sentences. Between 350 and 450 ms, the two effects are very similar. In the two subsequent time windows (450–550 ms and 550–650 ms), by contrast, the effect for the ambiguous conditions shows a broader distribution and a longer duration. We will discuss both of these properties in turn.

Firstly, the longer duration could be due to the difference between an irresolvable conflict – and hence the abortion of processing – in the unambiguous conditions and a resolvable conflict – and hence a required additional effort in conflict resolution – in the ambiguous conditions. A similar distinction was proposed by Roehm et al. (2004) in order to explain the different frequency characteristics of N400 effects elicited by case violations (an irresolvable conflict) and N400 effects for inanimate subjects (a conflict that does not preclude sentence interpretation).

Secondly, the broader distribution may reflect the additional prominence violation in the object-initial ambiguous sentences, i.e. the required reanalysis not only towards an Undergoer-initial order, but also towards an inanimate-initial order. On the one hand, functional neuroimaging results from German have demonstrated that inanimate-before-animate orders are more costly to process than animate-before-inanimate orders (Grewe et al., 2006). On the other, Haupt et al. (2008) observed a reanalysis N400 with a broader distribution (i.e. significant at anterior and posterior sites) when the first argument is less prominent than the second. Though Haupt and colleagues' findings were based on the definiteness/specificity scale rather than the animacy scale, the results for 450–550 ms time window in the analysis of the difference waves show a remarkable similarity to these previous observations.

Taken together, these findings thus point to important points of overlap between the N400 effects in the unambiguous and ambiguous conditions, but also to crucial differences between them. They suggest that the Compute Linking step first checks the compatibility of all information types, with an initial mismatch in this step processed similarly in resolvable and irresolvable conflicts. If the conflict is found to be irresolvable, Linking is aborted. By contrast, if it appears resolvable, the representations constructed prior to the verb must be reanalysed and Compute Linking is consequently prolonged.

3.3. Summary and conclusions

The present study suggests that semantic P600 effects are language-specific and only arise in fixed word order languages like English and Dutch. These findings thus bear out the cross-linguistic predictions of the extended Argument Dependency Model (eADM). From a cross-linguistic perspective, it thus seems apparent that semantic P600s should not be taken as predictors for the relation between syntactic and semantic processing. Rather, they may provide an interesting diagnostic for the information types that condition linking in a particular language.

4. Acknowledgements

The research reported here was conducted in collaboration with the Clinic for Audiology and Phoniatry (Prof. Manfred Gross) of the Charité Berlin. We are grateful to Katja Bruening for invaluable assistance in data acquisition. We would also like to thank Alena Witzlack-Makarevich for the preparation of the sentence materials and Markus Philipp and Jona Sassenhagen for assistance in the preprocessing and analysis of the ERP data.

References

Bader, M. & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, *28*, 121–143.

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*, 55–73.

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2009a). *Processing syntax and morphology: A neurocognitive perspective.* Oxford: Oxford University Press.

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2009b). The role of prominence information in the real time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, *3,* 19–58.

Bornkessel, I. (2002). *The Argument Dependency Model: A neurocognitive approach to incremental interpretation* (28). Leipzig: MPI Series in Cognitive Neuroscience.

Bornkessel, I., McElree, B., Schlesewsky, M. & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*, 495–522.

Bornkessel, I. & Schlesewsky, M. (2006). The Extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*, 787–821.

Ferreira, F. & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Connecticut.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance, vol. 12: The psychology of reading*. (pp. 559–586). Hove: Erlbaum.

Frazier, L. & Clifton, C., Jr. (1996). *Construal.* Cambridge, MA: MIT Press.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y. & Schlesewsky, M. (2006). Linguistic prominence and Broca's area: The influence of animacy as a linearization principle. *Neuroimage*, *32*, 1395–1402.

Hagoort, P., Brown, C. & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D. & Bornkessel-Schlesewsky, I. (2008). The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, *59*, 54–96.

Hoeks, J. C. J., Stowe, L. A. & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59–73.

Huynh, H. & Feldt, L. S. (1970). Conditions under which the mean-square ratios in repeated measurement designs have exact F-distributions. *Journal of the American Statistical Association*, *65*, 1582–1589.

Kim, A. & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.

Kolk, H. H. J., Chwilla, D. J., van Herten, M. & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*, 1–36.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M. & Holcomb, P. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes*, *21*, 489–530.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N. & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentence: Evidence from event-related potentials. *Brain and Language*, *100*, 223–237.

Kuperberg, G. R., Sitnikova, T., Caplan, D. & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–469.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Osterhout, L. & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L. & Holcomb, P. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, *8*, 413–437.

Pustejovsky, J. (1995). *The generative lexicon.* Cambridge, MA: MIT-Press.

Rayner, K., Carlson, G. & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 657–673.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S. & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *Neuroreport*, *15*, 409–412.

Schlesewsky, M. & Bornkessel, I. (2004). On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua*, *114*, 1213–1234.

Schlesewsky, M. & Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolution: Electrophysiological evidence from subject-object ambiguities in language comprehension. *Brain Research*, *1098*, 139–152.

Schlesewsky, M., Fanselow, G., Kliegl, R. & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 65–93). Dordrecht: Kluwer.

Schriefers, H., Friederici, A. D. & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*, 499–520.

van Herten, M., Chwilla, D. J. & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*, 1181–1197.

van Herten, M., Kolk, H. H. J. & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*, 241–255.

Van Valin, R. D., Jr. (2005). *Exploring the syntax-semantics interface.* Cambridge: Cambridge University Press.

Vissers, C. T. W. M., Chwilla, D. J. & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1106*, 150–163.

Vissers, C. T. W. M., Kolk, H. H. J., van de Meerendonk, N. & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, *46*, 967–982.

Pragmatics

On-line studies of repeated name and pronominal coreference

Kerry Ledoux

Cognitive Neurology, Johns Hopkins University

Abstract

Coreference (the process by which anaphors and their referents are understood to refer to the same entities in the world) is a ubiquitous and essential feature of natural language processing. Much attention has been paid to processing differences in the establishment of coreference using different referential forms, such as pronouns and repeated names. In a series of experiments, using both Eye-Tracking and event-related potentials (ERPs), we further explored this question by manipulating the discourse context in which different referential forms were used to establish coreference. Specifically, we examined the effect of the discourse prominence of the antecedent on the processing of coreferential repeated names. Additionally, we examined the interaction of discourse prominence and implicit causality during the processing of pronouns and coreferential repeated names. The results of these experiments suggest that the structure of the discourse context has direct effects on the processes engaged during the establishment of coreference. These effects can, in some cases, override effects that would be expected based on lexical or lexical-semantic processing alone.

1. Introduction

Two linguistic expressions are said to corefer when they refer to the same entity in the world. Coreference is ubiquitous to natural language processing and indeed is essential to the coherence of any discourse. As exemplified in [1] (where coreferential words are underlined), coreference can be established using different linguistic forms, such as pronouns and repeated names.

[1] Daniel and Amanda moved the desk because Daniel need room for his filing cabinet.

The choice of one form of referring expression over another in language has been shown to be influenced, at least in part, by the focus of the antecedent expression in the mental representation of the discourse: pronouns tend to be preferred with focused antecedents, while repeated expressions (names or NPs) are commonly used to refer back to a nonfocused antecedent. Focus can be established in any number of ways (semantic, pragmatic, prosodic); my colleagues and I have been particularly interested in a focusing mechanism, discourse prominence, that is derived from discourse structure. A formal definition of discourse prominence can be found in Gordon and Hendrick (1998); informally, prominence is related to height in a syntactic tree, and thus entities that appear in subject position or are not deeply embedded syntactically are more prominent than entities that appear in object position or are more syntactically embedded. In four experiments, we examined the processing of different anaphoric expressions that referred to antecedents that varied along the dimension of discourse prominence, and attempted to address two research questions: first, to what extent does discourse prominence influence the processing of individual words in a discourse? Second, to what extent does discourse prominence interact with other focusing mechanisms?

2. Experiments

2.1. The processing consequences of antecedent prominence

We addressed the first of these research questions in three experiments (one using Eye-Tracking, one using visual ERPs, and one using auditory ERPs) using the same stimuli ([2]). In these stimuli, we took advantage of one of the features of coreferential repeated names, that being that these words function as both repetitions of lexical form and as referring expressions. We thus examined the contribution to the processing of these words of two different influences, that of repetition and that of the discourse contextual factor of prominence.

[2] At the office Daniel (and Amanda) moved the desk because Daniel/Robert need room for the filing cabinet.

We manipulated the discourse factor of prominence by including one or two names in the first noun phrase. When the initial name was a singular NP, it was more prominent than when it was embedded in a conjoined first NP. We manipulated the lexical factor of repetition by including a control for the repeated names, that is, sentences that introduced a new entity to the discourse model in the second clause. (Sentences were written so that the repeated name and new name continuations were judged equally acceptable in pretesting.) Discourse prominence should hinder the establishment of coreference between a repeated name and a prominent antecedent (relative to a case in which the antecedent is nonprominent). Repetition priming should confer a processing advantage to repeated lexical forms, relative to new ones. Alternatively, these factors might interact to determine the processing of the repeated names in our stimuli.

Experiment 1: Eye-Tracking

Previous Eye-Tracking experiments have shown a temporal dissociation between lexical repetition and discourse context effects. Specifically, repetition priming effects (faster reading times to repeated relative to new words) tend to be observed in early processing measures (like first fixation duration and gaze duration), while effects of sentential or discourse context tend to be observed on later processing measures (like total reading time) or on aggregate measures (like rereading). Additionally, context effects have often been observed on words downstream from the critical word itself, in the spillover region.

Our results showed similar patterns of reading times (Ledoux et al., 2007). At the critical word (underlined in [2]), on the early processing measures of first fixation duration and gaze duration, we observed evidence of repetition priming (faster reading times for repeated forms relative to new ones). At the verb region following the critical word, the later measure of total reading time and the aggregate measure of rereading showed evidence of an effect of antecedent prominence, with slower reading times for coreferential repeated names that followed prominent antecedents, relative to those that followed nonprominent antecedents.

Experiment 2: Visual ERPs

Previous ERP experiments have shown that lexical repetition (in lists or in sentence or discourse contexts) leads to a reduction in the amplitude of the N400 component. This same component has also been shown to be sensitive to the ease of integrating a word into its immediate sentence or discourse context.

Our results suggested the two factors of lexical repetition and discourse prominence interacted to determine the amplitude of the N400 to the critical word (Ledoux et al., 2007). We found a typical and large repetition priming effect on the N400 (with a reduced N400 to repeated relative to new words) when the antecedent of the repeated name was not prominent in the discourse representation. When the antecedent of the repeated name *was* prominent, this repetition priming effect was absent. Similarly, when we looked at the ERPs to new names that followed a prominent or nonprominent first NP, we found no effect of the prominence manipulation. For repeated names, on the other hand, we found a large difference at the N400: the amplitude of the N400 was reduced to repeated names that followed a nonprominent antecedent relative to those that followed a prominent antecedent.

Experiment 3: Auditory ERPs
Finally, we wanted to demonstrate that these effects were not unique to reading. We presented the same stimuli to a new set of participants in an auditory ERP experiment (Camblin et al., 2007). Here, we found essentially the same results as in the reading experiment: a reduction of the N400 to repeated names relative to new names following a nonprominent antecedent (but not following a prominent antecedent), and a difference in the amplitude of the N400 as a function of prominence for repeated names (but not for new names).

Across two ERP experiments, then, we found evidence of repetition priming for repeated names that were used in the service of establishing coreference, but only for sentences in which the antecedent of the repeated name was not prominent in the discourse representation. We also found evidence from the N400 that repeated names were more difficult to process following a prominent antecedent than they were following a nonprominent antecedent, a finding that has been termed a *repeated name penalty* in previous behavioural work (see, for example, Gordon et al., 1999). In general, across the Eye-Tracking and ERP experiments, we saw an influence of the discourse factor of antecedent prominence on the processing of words in a discourse.

2.2. Interaction of discourse prominence and implicit causality

In our fourth experiment (using visual ERPs), we chose to investigate prominence along with another well-known lexical-semantic focusing mechanism known as *implicit causality*. Implicit causality is a feature of certain interpersonal verbs by which information about the cause of events described by a verb is conveyed implicitly as part of the verb's meaning. Such verbs are often described as NP1-biasing verbs or NP2-biasing verbs, depending on the noun phrase that is generally identified as the cause of the events described by the verb ([3–4]).

[3a] Yesterday evening Ronald amused Alison because <u>he/Ronald</u> told a very funny joke. (NP1 bias – consistent)

[3b] Yesterday evening Ronald amused Alison because <u>she/Alison</u> needed cheering up. (NP1 bias – inconsistent)

[4a] At the museum Amy thanked Joe because <u>he/Joe</u> had explained the paintings so patiently. (NP2 bias – consistent)

[4b] At the museum Amy thanked Joe because <u>she/Amy</u> was trying to practice good manners. (NP2 bias – inconsistent)

In our experiment (Ledoux et al., 2008), coreference was established by means of either a pronoun or a repeated name. This anaphor was either consistent or inconsistent with the implicit causality bias of the verb. We also investigated the influence of antecedent prominence in these stimuli: for all sentences, the entity in subject position (*Ronald* or *Amy*) is more prominent in the discourse representation than the entity in object position (*Alison* or *Joe*).

For pronoun sentences, at the critical word, we replicated the previous results of Van Berkum et al. (2007): between 500 and 800 ms post-stimulus onset, there was a greater positivity to pronouns that were inconsistent with the implicit causality bias of the verb, relative to pronouns that were consistent with this bias, and this factor did not interact with antecedent prominence.

For repeated names, we did observe an interaction of implicit causality with discourse prominence, such that the effect of violating the implicit causality bias of the verb differed, depending on the prominence of the antecedent. When the antecedent was not prominent in the discourse representation (a situation in which repeated name coreference was expected to be felicitous), we observed an effect of implicit causality that was similar to that seen for the pronouns: a larger positivity between 500 and 800 ms to repeated names that violated the implicit causality bias of the verbs relative to those that did not. On the other hand, when the antecedent of the repeated name was prominent in the discourse representation, and repeated name coreference was expected to be infelicitous, we observed an effect of verb implicit causality instead on the N400, the amplitude of which was reduced to names that were consistent with the bias of the verb.

We would like to interpret this result as suggesting that focusing mechanisms in discourse processing can have interactive effects: at least for the repeated names in our experiment (for which an effect of discourse prominence would be strongest), we saw that the effects of violating the implicit causality bias of the verb differed depending on the prominence of the antecedent.

3. Conclusions

In four experiments, using Eye-Tracking and event-related potentials, we have attempted to examine the influence of a particular discourse contextual factor specific to coreference (the prominence of the antecedent in the mental representation of the discourse) on the processing of individual words in that discourse. We have shown that this feature of the discourse can influence the processing of individual words, and can interact with other mechanisms used to establish discourse focus.

4. Acknowledgements

I am grateful to have had the opportunity to present the work I have conducted with several collaborators (Megan Boudewyn, Chrissy Camblin, Barry Gordon, Peter C. Gordon, Mikael Roll, Tamara Swaab, and Corey Ziemba). Many thanks to the conference organizers and to the Birgit Rausing Language Program for this wonderful opportunity.

References

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C. & Swaab, T. Y. (2007). Processing new and repeated names: Effects of coreference on repetition priming with speech and fast RSVP. *Brain Research, 1146*, 172–184.

Gordon, P. C. & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22*, 389–424.

Gordon, P. C., Hendrick, R., Ledoux, K. & Yang, C.-L. (1999). Processing of reference and the structure of language: An analysis of complex noun phrases. *Language & Cognitive Processes, 14*, 353–379.

Ledoux, K., Gordon, B., Roll, M. & Swaab, T. Y. (2008). The influence of implicit causality on the establishment of coreference: An event-related potential study. *Manuscript in preparation*.

Ledoux, K., Gordon, P. C., Camblin, C. C. & Swaab, T. Y. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & Cognition, 35*, 801–815.

Van Berkum, J. J. A., Koornneef, A. W., Otten, M. & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171.

Can context affect gender processing?

ERP differences between definitional and stereotypical gender

Hamutal Kreiner, Sibylle Mohr,
Klaus Kessler, Simon Garrod
Department of Psychology, University of Glasgow

Abstract

Previous research using Eye-Tracking and ERP has shown that readers experience processing difficulty when an anaphor (*herself*) refers to a gender-mismatching antecedent (*minister*). The mismatch-effect is due to a clash between the gender of the pronoun and the gender of the co-referential noun. We report two EEG experiments using anaphora (Experiment 1) and cataphora (Experiment 2) sentences, designed to investigate the processing differences between stereotypical (*minister*) and definitional (*king*) gender role nouns. Consistent with previous findings (Osterhout et al., 1997), our results reveal similar mismatching effects for these noun types in anaphora. Critically, however, in cataphora, where the pronoun precedes its co-referring noun, diverging ERP signatures are revealed. Differences in an early interval suggest fast "gender-coercing" for stereotypical nouns while effects in a later interval are likely to reflect gender mismatch processing in definitional nouns. These findings suggest that discourse constraints modulate the processing of these two noun types in different ways.

1. Introduction

Although previous research has shown that gender information is represented at different linguistic levels such as morphology, syntax, and pragmatics, gender agreement is typically considered as a grammatical process. In English, however, there is no regular morphological marking for gender and the computation of gender-agreement may rely on non-grammatical information. For example, in words such as *boy* gender is part of the lexical entry whereas words like *soldier* are assumed to be definitionally neutral but their interpretation is stereotypically biased by world knowledge towards either male or female. There are two theoretical accounts that make different assumptions about the processing of definitional and stereotypical gender role nouns. According to the lexical view, gender is a lexical feature for both noun types, whereas the inferential view assumes that stereotypical gender has to be inferred from world knowledge or discourse context (Carreiras et al., 1996; Garnham et al., 2002). In this paper, we address these contrasting views and investigate whether the processing of gender is controlled by different processes in these two noun types.

Several Eye-Tracking (ET) studies on gender agreement demonstrate that readers immediately slow down when an anaphor (*herself*) refers to an antecedent (*minister*) that mismatches its stereotypical gender (Carreiras et al., 1996; Kennison, 2003; Sturt, 2003; Duffy & Keir, 2004). A corresponding mismatch-effect was also shown in ERP (Osterhout et al., 1997). These effects have been attributed to a clash between the gender of the pronoun and the conflicting gender of the co-referring noun. According to the inferential account, inferences are made immediately when a stereotypical noun is processed, leading to processing difficulty when the anaphor mismatches the inferred gender. According to a lexical view, stereotypical gender is a lexico-semantic feature, just like definitional gender, and the mismatch-cost reflects a feature clash. Most of the previous research trying to clarify this issue used anaphora sentences, where the antecedent precedes the referring pronoun (but cf. Banaji & Hardin, 1996). This allows time for inferences to take place before the referring expression is encountered, making it impossible to say whether the processing difficulty occurs at the discourse level due to an inference conflict or at the grammatical level due to a feature clash.

However, a recent ET study by Kreiner et al. (2008) compared anaphora [1a–1d] and cataphora [2a–2d] sentences contrasting the processing of definitional (*king*) versus stereotypical (*minister*) role nouns. Note that in cataphora, the reflexive pronoun precedes the co-referring role noun and unambiguously determines the gender of the relevant discourse referent before the critical noun is actually processed. Crucially, the lexical and the inferential views make different predictions for the processing of stereotypical and definitional role nouns in *cataphora* sentences: according to the inferential view, gender is lexically underspecified and acquired via inferences for stereotypical nouns. Consequently, in cataphora, where the pronoun has already assigned a gender to the discourse referent before the relevant role noun is encountered, the gender inference is constrained and the agreement violation is prevented. Hence, stereotypical violations of gender agreement in cataphora are not expected to exhibit a mismatch effect. By contrast, the lexical view presumes no qualitative difference in the processing of stereotypical and definitional role nouns and consequently predicts similar mismatch effects in violations of gender agreement for both noun types. The results of the ET study by Kreiner et al. (2008) showed indeed that in cataphora, where a gender-marked pronoun precedes its co-referential noun, the mismatch-cost for stereotypical gender is eliminated [2d vs. 2c], suggesting that unlike definitional-gender [2a, 2b] the processing of stereotypical gender is modulated by discourse constraints.

These findings are inconsistent with previous findings that showed that both noun types elicited similar ERP components (P600) in anaphora sentences, which suggested a syntactic rather than a discourse clash between the pronoun and the antecedent (Osterhout et al., 1997). To resolve these conflicting findings, we designed two EEG experiments using anaphora (Experiment 1, [1a–1d]) and cataphora sentences (Experiment 2, [2a–2d]) that contrast stereotypical and definitional role nouns bound by matching or mismatching reflexive pronouns. In these experiments, we examined the hypotheses that discourse constraints may modulate the processing of stereotypical gender whereas definitional gender that is lexically determined would be insensitive to such constraints. If distinct processes control gender agreement with stereotypical compared to definitional role nouns, we would expect the different nouns to elicit distinct ERP signatures during the processing of gender agreement.

2. Experiments

2.1. Method

Participants. For each experiment, we recruited twenty native speakers of English from the Glasgow University community who were paid for participation. Only participants that had not taken part in Experiment 1 were allowed to participate in Experiment 2.

Materials. Experiment 1 used 160 anaphora sentences such as [1a–1d] that employed a role noun as antecedent of a matching or mismatching co-referring pronoun (*himself/herself*; pronoun and role noun are both underlined in the examples) that served as the target word in the sentence. Each item had four versions, representing the four experimental conditions of a 2x2 design manipulating Gender Type (definitional [1a, 1b] versus stereotypical [1c, 1d]) and Matching (match [1a, 1c] versus mismatch [1b, 1d]). The selection of stereotypical role nouns was based on a norming study with 50 participants from Glasgow University who did not take part in the other experiments.

[1a] The king left London after reminding himself about the letter.

[1b] The king left London after reminding herself about the letter

[1c] The minister left London after reminding himself about the letter

[1d] The minister left London after reminding herself about the letter

Experiment 2 was similar to Experiment 1, except for the experimental sentences. We used 160 cataphora sentences such as [2a–2d] which were derived from the materials used in Experiment 1. The linear order of pronoun and noun was reversed, i.e. the referring pronoun occurred prior to the critical noun, hence the noun served as the target word. As in Experiment 1, each item appeared in four versions, representing the four experimental conditions.

[2a] After reminding himself about the letter, the king left London.

[2b] After reminding herself about the letter, the king left London.

[2c] After reminding himself about the letter, the minister left London.

[2d] After reminding herself about the letter, the minister left London.

Procedure. EEG was recorded while participants read silently sentences presented word by word on a computer screen. A BIOSEMI Active-Two amplifier system was used for continuous recording of EEG activity from 72 electrodes at a sampling rate of 256 Hz. Each trial started with a sentence title including 2–5 words presented on the screen until the participant was ready and pressed a key to continue. This was followed by a 500 ms blank screen and then by a fixation cross in the centre of the screen. The sentence was then presented word-by-word. The words appeared in the centre of the screen for a duration of 300 ms each, and with an inter-stimulus interval of 200 ms. In addition to the 160 experimental items 160 filler sentences were included in each experiment. Finally, 25% of the stimuli (both experimental and fillers) were followed by a simple comprehension question that required a yes/no reply via button press.

Data analysis. BESA Version 5.2 was used to filter frequency bands of 0.3–25.0 Hz. and correct or reject ERP artifacts. Participants for which more than 12.5% of the trials of a particular experimental condition were rejected were excluded from the analysis (all together 4 participants in each experiment). EEG was time-locked to the onset of the target word (i.e., the pronoun in Experiment 1 and the role noun in Experiment 2). We computed average waveform in this critical time window for each participant in each experimental condition, after normalizing the waveforms of the individual trials relative to a 150 ms pre-stimulus baseline interval preceding the critical word using the electrodes' average as reference.

For the purpose of statistical analysis, the electrodes were grouped into five Regions Of Interest (ROI): Left Anterior (LA: Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, T7); Left Posterior (LP: PO9, O1, PO7, P7, CP3, CP1, TP7, CP5, P1, P3, P5, PO3); the corresponding Right Anterior (RA) and Right Posterior (RP) regions; and the Central region (Pz, CPz, Fz, FCz, Cz). ERPs were averaged over electrodes within each ROI. At each relevant epoch, we performed hierarchical statistical analysis, starting with a within subject 2x2x2x2 design reflecting (2 hemispheres) x (2 anterior-posterior ROIs) x (2 types of role nouns – definitional versus stereotypical) x (2 matching conditions – match versus mismatch). The Central ROI was analyzed

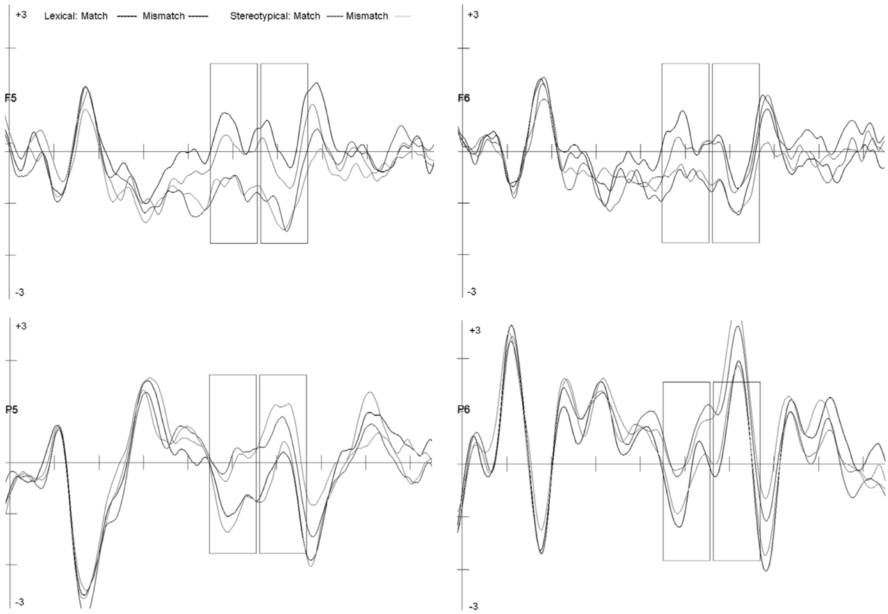


Figure 1. Grand averaged ERPs time-locked to the onset of the target word (the reflexive pronoun *herself/himself*). The darkest lines represent the matching conditions with the darkest representing the Definitional-Matching and the slightly lighter line the Stereotypical-Matching. The lighter lines represent the mismatching conditions with the slightly darker representing the Definitional-Mismatching and the lightest line the Stereotypical-Mismatching. Four example electrodes that represent the four different ROIs and show reliable effects are presented.

separately. Subsequently, for ROIs in which reliable effects were observed, we report repeated measures ANOVA testing the effects of experimental factors for each electrode.

2.2. Experiment 1: Results and discussion

Figure 1 presents grand averages for four example electrodes from the LA, LP, RA, and RP ROIs. As can be clearly seen in this figure, matching and mismatching conditions show diverging patterns from around 450 ms after stimulus onset. Based on this visual observation, we performed statistical analysis on the epochs of 450–550 ms and 550–650 ms from pronoun onset.

In the epoch of 550–650 ms from the target (pronoun) onset, we found a reliable effect for Matching ($F(1, 19) = 26.24$, $p < .0001$) modulated by a reliable interaction with Hemisphere ($F(1, 19) = 5.07$, $p < .05$) and with the Anterior-Posterior factor ($F(1, 19) = 35.83$, $p < .0001$), as well as a reliable 3-way interaction between Matching, Noun Type and the Anterior-Posterior factor ($F(1, 19) = 7.29$, $p < .05$). Subsequent 2x2 analysis of each ROI shows a reliable Matching effect in all ROIs except the Central

(LA: $F(1, 19) = 45.89$, $p < .0001$; LP: $F(1, 19) = 6.63$, $p < .05$; RA: $F(1, 19) = 9.13$, $p < .005$; RP: $F(1, 19) = 39.29$, $p < .0001$). This was modulated by an interaction with the Noun Type reliable only in the Left ROIs (LA: $F(1, 19) = 6.63$, $p < .05$; LP: $F(1, 19) = 7.91$, $p < .05$). Inspections of the epoch mean amplitudes at each region reveals a positivity Matching effect for both LP (definitional match –0.13 vs. mismatch 0.80; stereotypical match 0.09 vs. mismatch 0.57) and RP (definitional match 0.47 vs. mismatch 1.63; stereotypical match 0.68 vs. mismatch 1.38). The reverse pattern, namely a negativity effect, was observed for LA (definitional match –0.15 vs. mismatch –1.12; stereotypical match –0.41 vs. mismatch –0.87) and RA (definitional match –0.04 vs. mismatch –0.51; stereotypical match –0.12 vs. mismatch –0.42). Although the interactive pattern, showing a slightly smaller difference between the matching and mismatching conditions of stereotypical compared to definitional gender, is only reliable for the posterior ROIs, a similar trend is observed in the Anterior ones. Detailed analysis of the single electrode data reveals a consistent pattern with reliable Matching effect in several electrodes in LA (Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3), LP (CP5, CP3, P1, P3, P5, P7, PO7, PO3, O1), RA (Fp2, AF8, AF4, F2, F4, F6, F8, FT8, FC6), RP (TP8, CP6, CP4, CP2, P2, P4, P6, P8, PO8, PO4, O2) and in the Central ROI (Iz, Oz, POz, Pz, CPz, Fpz, AFz, Fz). Only a few electrodes showed reliable interaction (La: Fp1, F7, FC5, FC3; LP: PO7, O1; RP: P8; Central: Pz).

Analysis of the earlier epoch of 450–550 ms from the target onset reveals that the Matching effect is reliable at this epoch. Thus, a reliable effect is shown for Matching ($F(1, 19) = 18.65$, $p < .0005$) modulated by a reliable interaction with the Anterior-Posterior factor ($F(1, 19) = 17.16$, $p < .001$). Subsequent 2x2 analysis of each ROI shows a reliable Matching effect in all ROIs (LA: $F(1, 19) = 15.64$, $p < .001$; LP: $F(1, 19) = 13.05$, $p < .01$; RA: $F(1, 19) = 7.38$, $p < .01$; RP: $F(1, 19) = 17.50$, $p < .0001$; Central: $F(1, 19) = 26.83$, $p < .0001$). In this epoch, unlike in the 550–650 epoch, this effect was modulated by an interaction with the Noun Type only at the Central ROI ($F(1, 19) = 14.28$, $p < .001$). As can be seen in the data from the example electrodes (Figure 1) at this epoch, the Matching effect is more negative for the posterior ROIs and relatively positive for the anterior ROIs. Detailed analysis of the single electrode data reveals that the Matching effect is reliable in several electrodes in LA (Fp1, AF7, AF3, F3, F5, F7, FT7, FC5, FC3, FC1), LP (CP5, CP3, CP1, P1, P3, P5, P7, PO7, PO3, O1), RA (Fp2, AF8, AF4, F6, F8, FT8, FC6, FC2, C2, C4), RP (TP8, CP6, CP4, CP2, P2, P4, P6, P8, PO8, PO4, O2) and in the Central ROI (POz, Pz, CPz, Fpz, AFz, FCz, Cz).

In general, the results are consistent with previous findings that showed a P600-like Matching effect for both noun types modulated by an interaction with the Noun Type (Osterhout et al., 1997) and with EM findings showing a corresponding mismatch-cost for both Noun Types.

2.3. Experiment 2: Results and discussion

Figure 2a and 2b presents grand averages for four example electrodes from the LA, LP, RA, and RP ROIs. A brief look at the figure reveals that the ERP patterns elicited by the cataphora sentences are very different from those elicited by anaphora. The most striking difference is that the Gender Matching factor does not seem to result in a very clear Matching effect across the two Noun Types as demonstrated in Experiment 1. Based on our previous eye-movement findings that showed that the gender matching effect for cataphora is delayed compared to anaphora, and on visual inspection of the wave forms, we performed statistical analysis on the epochs of 650–800 ms and 250–400 ms from pronoun onset.

In the epoch of 650–800 ms from the target (role noun) onset, the general 4-way analysis did not reveal any reliable effects. However, subsequent 2x2 ANOVAs for the individual ROIs revealed that the effect of Noun Type reached significance in the RA ROI ($F(1, 19) = 5.35, p < .05$). Whereas none of the ROIs exhibited a reliable effect of Matching, the interaction between Matching and Noun Type was reliable in both LA ($F(1, 19) = 8.49, p < .01$) and RP ($F(1, 19) = 6.98, p < .05$) ROIs. Inspections of the epoch mean amplitudes clarifies the pattern of simple effects underlying this interaction: a reliable Matching effects for the definitional (RP: match –0.08 vs. mismatch –0.14; LA: match –0.20 vs. mismatch 0.06) but not for the stereotypical (RP: match 0.04 vs. mismatch 0.09; LA: match –0.10 vs. mismatch –0.15) conditions. Detailed analysis of the single electrode data in ROIs that showed reliable effects reveals a reliable interaction in several electrodes in the LA ROI (F1, F3, F5, FC5, FC3, FC1) and only in one electrode (P6) in the RP ROI. The effect of Noun Type reached significance in 3 RA electrodes (F8, FT8, T8). While this pattern of results is not as clear as the results shown for anaphora sentences, it is generally consistent with our previous EM findings for cataphora sentences that indicated an interaction between Noun Type and Matching whereby violation of definitional gender agreement, unlike stereotypical, results in a processing disruption.

In the 450–550 ms and 550–650 ms from the target onset, we see some reliable effects for the Noun Type. Such effects may reflect differences between these noun types (e.g. morphology) that are not the main focus of this paper (as long as they do not modulate the Matching effect). Hence, the related analyses are not reported here. Surprisingly, in the earlier epoch of 250–400 ms from target onset, the general 4-way analysis reveals a reliable 3-way interaction between the Matching, the Noun Type and the Hemispheres ($F(1, 19) = 5.10, p < .05$). None of the main effects or other interactions has reached significance. Subsequent 2x2 ANOVAs for the individual ROIs revealed that the effect of Noun Type reached significance in the RP ROI ($F(1, 19) = 5.04, p < .05$). Whereas the Matching effect was not reliable, the interaction between Matching and Noun Type was reliable in both LP ($F(1, 19) = 8.24, p < .01$) and RA ($F(1, 19) = 8.34, p < .01$) ROIs. Inspections of the epoch mean amplitudes clarifies the pattern of simple effects underlying this interaction, showing Matching effects for the stereotypical (RP: match 0.07 vs. mismatch 0.11; LA: match –0.08 vs. mismatch –0.13) but not for the definitional (RP: match –0.08 vs. mismatch 0.005; LA: match –0.01 vs. mismatch –0.05) conditions. Detailed analysis of the single electrode data in ROIs that showed reliable effects reveals a reliable interaction in several electrodes in the LP (P5, P7, PO7, O1) and RP (FC6, C6, T8) ROIs. These relatively early effects for stereotypical but not definitional gender agreement do not correspond to our previous EM findings for cata-

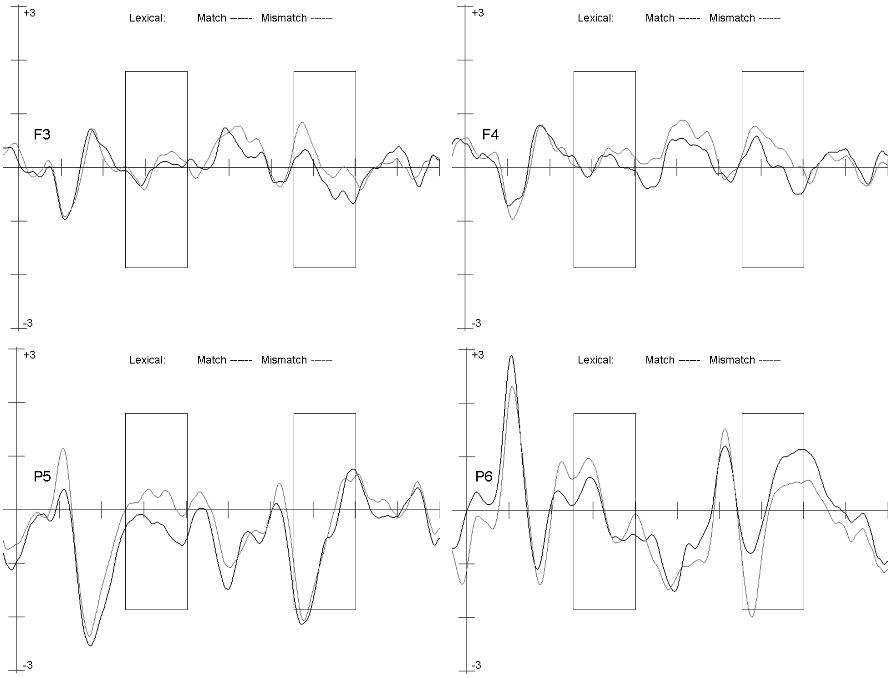


Figure 2a. Grand averaged ERPs time-locked to the onset of the target word (the role noun *king/ minister*) in the Definitional condition. The darker line represents the matching condition and the slightly lighter line the Mismatching condition. Four example electrodes that represent the four different ROIs and show reliable effects are presented.

phora sentences. It seems, therefore that the ERP effects at this epoch reflect processes that are not captured by the EM measures. One possible interpretation of this effect is that it reflects the process underlying the coercion of stereotypical gender to discourse constraints (we discuss this interpretation in more detail in the Conclusion).

2.4. Oscillatory brain activity

Recent research suggests that transforming the EEG signal into the time-frequency domain may provide a more sensitive analysis, since changes in oscillatory activity are largely cancelled out due to signal averaging during ERP analysis. Therefore we also looked for Event-related Phase Synchronization in a frequency range of 2–60 Hz to determine whether stimulus-specific oscillatory modulations are linked to gender-agreement violations during the anaphoric processing of definitional and stereotypical gender nouns. Power estimates for each subject and each condition were obtained from wavelet-based time-frequency (TF) representations using the BESA® 5.1 software package. To isolate induced-activity oscillations, we subtracted evoked activity oscillations

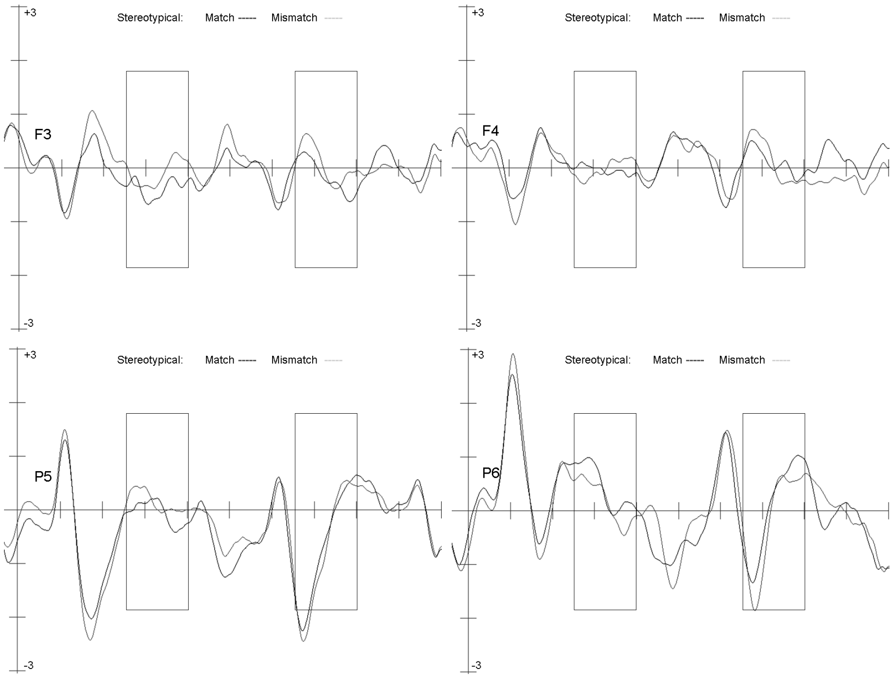


Figure 2b. Grand averaged ERPs time-locked to the onset of the target word (the role noun *king/minister*) in the Stereotypical condition. Legend as for Figure 2a.

by means of regression (cf. Tallon-Baudry, 2003). Using FieldTrip (a Matlab® toolbox developed at the FC Donders Centre for Cognitive Neuroscience, Nijmegen) the TF representations were then averaged over trials for each subject and each condition. To normalize for inter-individual differences, we set a baseline interval –5.7 s to –5.2 s prior to stimulus onset. We analyzed the time window of 0–1000 ms from stimulus onset applying a randomization procedure, a cluster-based permutation test (Monte Carlo method, 1000 randomizations, $\alpha = 0.05$) with a cluster growing approach that corrects for the multiple comparisons problem (Maris et al., 2007). This method identifies time-frequency-channel clusters, which show reliable power changes between the mismatching and matching conditions. This analysis revealed significant differences only for the alpha and gamma frequency bands.

At the alpha frequency range (8–12 Hz) the comparison between mismatch vs. match conditions shows a decrease of power in both experiments for both definitional and stereotypical role nouns (one negative cluster each, $p < .05$, left posterior electrodes for both experiments, e.g. Pz, PO7, PO3, PO1, P7, P5, P3, O1, 0–1000 ms). This result indicates that while the alpha band activity is sensitive to gender matching, it does not reveal any distinction between the different noun types we have examined.

In contrast, the gamma frequency range (30–60 Hz for all channels) appears to be more sensitive to the processing differences between these two noun-types. In the cataphora sentences (Experiment 2) it reveals a positive cluster ($p = 0.02$) indicating a strikingly larger power increase for the mismatch compared with match conditions for definitional but not for stereotypical role nouns. The anaphora sentences of Experiment 1 do not reveal such a pattern. Subsequent comparison of hemispheres showed that the effect is confined to central electrodes of the right hemisphere (F6, F8, FC2, FC4, FC6, FT8, C2, C4, C6, T8, CP4, CP6, TP8) and peaks around 300–900 ms.

Currently, little is known about the neuro-cognitive functions underlying oscillatory brain activity during language comprehension. Therefore, we can only propose preliminary and rudimentary interpretations of these findings. Modulations in the alpha frequency range have been interpreted as reflecting attention (Klimesch, 1999), working memory (Jensen et al., 2002), semantic processing (Rohm et al., 2001), and visual processing involved in reading words (Bastiaansen & Hagoort, 2003). The differentiated alpha modulations in matching compared to mismatching conditions observed in both experiments are consistent with higher level functions such as working memory or semantic processing. Gamma rhythms have been linked to attention and alertness (Tallon-Baudry, 2003) and may reflect underlying feature-binding and memory processes ("binding" gamma rhythm; Engel & Singer, 2001). Although this feature-binding interpretation was proposed in the context of attentional-perceptual integration, it may have implications for linguistic integration. The widespread gamma modulation for the definitional mismatch condition in cataphora, which is hardly observable in anaphora, may suggest differences in the process of linguistic binding underlying reference resolution in anaphora compared with cataphora (see Kazanina et al. (2007) for a discussion of the processing differences between anaphora and cataphora). Future research will shed more light on these somewhat speculative interpretations and on the potential contributions of lexical syntactic and discourse factors to such gamma modulations.

3. Conclusion

The results of Experiment 1 clearly show a mismatch effect in anaphora sentences for both stereotypical and definitional gender role-nouns, and this effect is quantitavely modulated by an interaction with the noun type. This pattern replicates Osterhout et al.'s (1997) findings, and is seemingly consistent with the lexical view that the representation and processing of gender in these two noun types are qualitatively similar. The quantitative differences between them may reflect graded differences in the gender representation rather than distinct processing mechanisms. However, as argued above, the critical test for the hypothesis that the gender processing in these noun types is controlled by different processes is the cataphora experiment.

The cataphora sentences used in Experiment 2 elicited strikingly different waveforms and ERP components, plausibly related to general differences (discussed below) between these structures. However, the critical finding in addressing the definitional-stereotypical processing difference hypothesis is that in cataphora, unlike anaphora, these noun types elicit distinct ERP signatures. Thus, while in anaphora both noun types exhibited a similar mismatch-effect, in cataphora differentiated patterns were observed: initially, at the 250–400 ms, a mismatch effect was observed only for stereotypical role nouns; later at the 650–800 ms epoch, a mismatch-effect was observed only for definitional and not for stereotypical gender. Similarly, the TF analysis at the gamma band revealed a mismatch effect only for definitional gender. The latter definitional mismatch-effects correspond to the ET findings of Kreiner et al. (2008), indicating that violation of gender agreement in cataphora sentences results in reading difficulty only for definitional gender. Both the inferential and the lexical views assume that definitional gender is part of the lexical entry. Thereby, it is accessed by default, leading to a gender clash when the dependency is resolved and agreement violation is detected both in anaphora and in cataphora. Crucially, however, the inferential view, unlike the lexical view, predicts a different pattern in cataphora. Based on the inferential view, we initially predicted that in cataphora stereotypical role nouns would not exhibit mismatch effects because the discourse constraints set by the preceding gender marked pronoun make gender inferences unnecessary and may therefore prevent the gender clash. The stereotypical mismatch effect revealed in the relatively early (250–400 ms) epoch is inconsistent with this prediction. To account for this finding we propose a slightly different interpretation of the inferential view. Namely, stereotypical gender inferences may not be prevented by discourse constraints. Rather, they are automatically activated but coerced by discourse constraints in case of conflict. According to this account, the early and short-lived mismatch-effect exhibited by the stereotypical role nouns may be interpreted as reflecting the process of coercion, i.e. shifting from the world-knowledge based gender bias to the discourse constrained bias. Since such coercion cannot occur in the case of definitional role nouns, this early mismatch-effect is exhibited selectively by stereotypical role nouns. The notion of automatic activation of gender stereotype is consistent with previous findings both from anaphora resolution (Oakhill et al., 2005; Reynolds et al., 2006) and from priming (Banaji & Hardin, 1996) studies.

Owing to the different linear order of the cataphora sentences, there is another major difference between the stimuli used in the two experiments. In cataphora, the earliest point at which agreement can be computed is when the role noun is encountered; inevitably then the target word is the role noun. Thus, whereas the target words in anaphora are high frequency function words and morphologically marked for gender (*himself/herself*), in cataphora the targets are of lower frequency and not marked with a regular morphological gender marker.[1] It is plausible that the different wave forms and ERP components observed reflect the different processes involved in processing pronouns compared with role nouns. However, this divergence cannot account for the differences in the mismatch effect in stereotypical compared to definitional nouns.

Taken together, the findings from the two experiments reported here are consistent with the inferential view. In showing distinct ERP signatures for reference resolution

[1] Some of the definitional role nouns have affixes such as *-ess* in *waitress*. However, since these are not regular gender markers in English, we tend to consider them as part of the lexical entry rather than a morpho-grammatical feature.

with stereotypical compared with definitional role nouns, they support the view that the underlying mechanisms of gender processing in these noun types are different.

4. Acknowledgements

This work is part of the ESRC-funded Gender project (grant number: RES-000-22-1924).

References

Banaji, M. R. & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7*(3), 136–141.

Bastiaansen, M. & Hagoort, P. (2003). Event-induced theta responses as a window on the dynamics of memory. *Cortex, 39*, 967–992.

Carreiras, M., Garnham, A., Oakhill, J. & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology, 49*, 639–663.

Duffy, S. A. & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition, 32*(4), 551–559.

Engel, A. K. & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*(1), 16–25.

Garnham, A. Oakhill, J. & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition, 30*(3), 439–446.

Jensen, O., Gelfand, J., Kounios, J. & Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex, 12*(8), 877–882.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M. & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56*, 384–409.

Kennison, S. M. & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research, 32*(3), 355–378.

Klimesch, W. (1999). EEG alpha and thete oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*, 169–195.

Kreiner, H., Sturt, P. & Garrod, S. C. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language, 58*, 239–261.

Maris, E., Schoffelen, J.-M., Fries, P. (2007). Nonparametric statistical testing of coherence differences. *Journal of Neuroscience Methods, 163*, 161–175.

Oakhill, J. V, Garnham, A. & Reynolds, D. J. (2005). Immediate activation of stereotypical gender information in reading. *Memory & Cognition, 33*(6), 972–983.

Osterhout, L., Bersick, M. & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition, 25*, 273–285.

Reynolds, D. J., Garnham, A. & Oakhill, J. (2006). Evidence of immediate activation of gender information from a social role name. *The Quarterly Journal of Experimental Psychology, 59*(5), 886–903.

Rohm, D., Klimesch, W., Haider, H. & Doppelmayr, M. (2001). The role of theta and alpha oscillations for language comprehension in the human electroencophalogram. *Neuroscience Letters, 310*(2–3), 137–140.
Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*(3), 542–562.
Tallon-Baudry, C. (2003). Oscillatory synchrony and human visual cognition. *J. Physiol. Paris* 97 (2–3), 355–363.

Prosody

Electrophysiological studies of prosody

Mireille Besson, Céline Marie
Institut de Neurosciences Cognitives de la Méditerranée, CNRS-Marseille Universités

Corinne Astésano
Laboratoire Jacques-Lordat, Université de Toulouse II – Le Mirail

Cyrille Magne
Psychology Department, Middle Tennessee State University

Abstract

This paper reviews a series of experiments that aimed at investigating the role of prosody in speech comprehension. These studies addressed different aspects of the linguistic function of prosody, in relation to the semantic and pragmatic levels of language processing. Specifically, modality, focus, and meter segmentation (through final syllable lengthening) were examined. Behavioral and electrophysiological measures (Event-Related Potentials, ERPs) were analyzed in all experiments.

Results revealed that prosodic cues are processed on-line and influenced ongoing sentence comprehension. Moreover, they also showed interactive effects of modality and semantics, of prosodic focus and pragmatic constraints, and of meter and semantics. These results are discussed in relation to models of spoken language comprehension and in the context of the specific or general nature of the processes involved in language processing.

1. Introduction

When you arrive home in the evening, you can easily guess from the intonation of the voice of the person that welcomes you whether s/he is happy, sad, or angry and whether s/he had a good or a bad day. The intonation patterns thereby convey the most intuitive form of prosody called emotional prosody that is a central feature in language acquisition (Cutler et al., 1997; Jusczyck & Hohne, 1997) and may be one of the primary features of human language (Molino, 2000). How these emotional prosodic patterns are perceived and processed by adult listeners, whether they are processed in real time, by which neurophysiological underlying mechanisms and how they influence spoken language comprehension were questions first addressed by Annett Schirmer, Sonja Kotz and collaborators from the Max Planck Institute (MPI) for Cognitive Neuroscience in Leipzig (Schirmer et al., 2002; see also Besson et al., 2002, for a review). But prosody also has several essential linguistic functions. Through pauses, meter, accents, and intonation, prosody has a structural function and allows the segmentation of a continuous flow of auditory information into meaningful units (words and prosodic phrases; Darwin, 1975; Cutler et al., 1997; Vaissière, 2005). In this respect, Steinhauer and collaborators (1999), also from the MPI in Leipzig, demonstrated in a seminal study that the presence of prosodic boundaries is associated with a specific pattern of brain waves that they called the “Closure Positive Shift”. This was the first piece of evidence that linguistic cues are processed on-line to favor speech segmentation into relevant prosodic and syntactic units. Around the same time, Böcker et al. (1999) evidenced the lexical function of prosody by studying the role of linguistic accents in speech comprehension.

In this paper, we will report the results of further studies conducted in our laboratory that aimed at examining other functions of linguistic prosody such as modality, focus, and meter. However, before describing these experiments, it is important to note that prosody in speech is expressed through a set of well-defined acoustic parameters: *pitch*, the percept derived from variations in frequency; *duration*, the perception of time; *loudness*, the percept derived from variations in intensity; and *timbre*, the most complex aspect (often defined by what it is not), that comprises spectral characteristics allowing one to recognize familiar voices and differentiate speech sounds. Thus, the different emotional and linguistic functions of prosody are expressed through variations in these fundamental acoustic parameters. As a consequence, studying prosody is an interdisciplinary enterprise that requires close collaboration between acousticians, phoneticians, linguists, psycholinguists, and neurolinguists if one is interested, as we are, in the neurophysiological mechanisms that underlie the perception of prosody. To track down the changes in brain activity elicited by the perception of some aspects of linguistic prosody, we used the Event-Related Potentials (ERPs) method that we will quickly describe before going into the details of the experiments.

2. The Event-Related Potential method

The brain is an extremely active device and its activity is reflected in metabolic, electric and associated magnetic changes. Here, we are interested in the changes in the brain electrical background activity, the electroencephalogram (EEG, Berger, 1929), that are elicited by the occurrence of a specific event or stimulus. Precisely time-locked to the presentation of a stimulus, small variations of electric currents occur in the EEG that are difficult to see because of poor signal to noise ratio. In order to increase this ratio, Dawson (1951) developed the averaging method: by synchronizing EEG recordings to the presentation of different stimuli and by averaging these different EEG epochs, he was able to increase the signal to noise ratio so that the variations in brain electrical activity that were directly linked to stimulus presentation emerged from the background EEG noise while random variations cancelled out. Since then, this method has become very popular mainly because of its ease of use and of its excellent temporal resolution. It is indeed possible to examine the time-course of stimulus processing, as reflected by variations in electric currents, as it unfolds over time with a temporal resolution that is only dependent upon the sampling rate (usually set from 250 Hz to 1000 Hz). The drawbacks of this method are its poor spatial resolution, so that it remains difficult to determine where the electrical variations recorded at the surface of the scalp are generated in the brain (called the inverse problem), and the difficulty linked to the interpretation of the functional significance of the successive variations, also called components. ERP components are defined by their polarity (negative or positive relative to the baseline of brain electrical activity before stimulus presentation), their amplitude (in μV), the latency of maximum amplitude (in ms) and their topographic distribution on the scalp.

2.1. The ERP components

In the auditory modality, which is of particular interest here, successive components reflect transmission of neural input from the cochlea and through the auditory pathway to primary and secondary auditory cortices. The first components develop between 0 and 10 ms in the brainstem. They are followed by mid-latency components between 10 and 80 ms and by late components between 80 and 1000 ms that are of particular interest in our experiments. The P1–N1–P2 complex (between 80 and 220 ms) reflects processing of acoustic parameters such as pitch, duration, intensity, and timbre. The amplitude of this complex is modulated by attention. The later components, such as the N2–P3 complex depend upon cognitive processing and stimulus relevance for the task to be performed. The amplitude and latency of the N2–P3 complex are modulated by categorization and decision processes. The P3 is one of the most studied ERP components and is classically considered as comprising a frontal P3a, reflecting detection and processing of surprising events, and a parietal P3b, reflecting categorization and decision processes. Finally, long-lasting components may develop, such as the "Negative Slow Wave" over frontal regions, or the "Contingent Negative Variation" that develops in the time interval between two stimuli when the first one signals the occurrence of the second one. The functional significance of late components is difficult to determine because processing variability increases from stimulus onset.

2.2. The N400 component

The use of the ERP method to study language processing has largely benefited from the discovery of the N400 component by two researchers from the University of California at San Diego, Marta Kutas and Steven Hillyard (1980). This negative component, with maximum amplitude around 400 ms after stimulus onset, is elicited by the presentation of semantically unexpected words within a sentence context. Further research has demonstrated that its amplitude is modulated by semantic expectancy so that the less expected a word within a given semantic context is, the larger the N400 amplitude (Kutas et al., 1984). While these pioneering experiments were conducted in the visual modality, further experiments have demonstrated that N400 components can also be elicited in the auditory modality (although with somewhat different latency and scalp topography; McCallum et al., 1984; Holcomb & Neville, 1990) and in sign language (Kutas et al., 1987). Thus, most researchers in the field would agree that the N400 component is tightly linked to semantic processing, independently of modality. However, the precise functional interpretation of the N400 remains a matter of debate. The occurrence of an N400 may reflect the mismatch between the word expected on the basis of the context (be it a sentence or a single word) and the word actually presented. In this case, the N400 component would reflect semantic expectancy and anticipation processes. However, it may be that the meaning of words occurring with a low probability within a context is more difficult to integrate than the meaning of words occurring with high probability. In this case, the N400 would reflect *a posteriori* integration of the meaning of a word within the preceding semantic context. Finally, and more recently, the N400 has been considered in light of a phenomenological approach in which the occurrence of an N400 component would reflect the process through which a perceived object becomes an object for consciousness (Piotrovski, 2008). In any event and whatever the most correct functional interpretation of the N400 may be, this component remains an excellent tool for examining semantic processing.

3. The modality function of prosody

The general aim of this first experiment was to study the modality function of prosody, that is the function of expressing speech acts such as asserting, questioning, or commanding through intonation (Astésano et al., 2004). Interestingly, in most languages, intonation contours are sufficient to convey the interrogative mode in otherwise syntactically declarative sentences (Garding, 1998; Hirst, 1998). This phenomenon is particularly frequent in French (Delattre, 1966) where sentences comtaining exactly the same words and syntactic structure, as illustrated by [1a] and [1b], are clearly perceived as a Statement ([1a]) or as a Question ([1b]) on the basis of just the intonation:

[1a] Michel fume un cigare.
'Michel is smoking a cigar.'

[1b] Michel fume un cigare?
'Michel is smoking a cigar?'

Results of experiments using behavioral measures have shown that intonation contours convey the speakers' communicative intentions mainly through variations in fundamental frequency (F0). Thus, while statements are characterized by falling intonation contours (F0-fall), questions are characterized by a declination phenomenon (i.e. downstepping of F0) followed by a final intonation rise on the last syllable (F0-rise). Based on this knowledge about modality and intonation contours, we created prosodic mismatches by cross-splicing the beginning of Statements with the end of Questions (*Statements) and vice-versa (*Questions; see Figure 1). Moreover, we also created semantic mismatches by replacing the sentence final congruous words by semantically unexpected incongruous words. Thus, we used a 2x2 design with four experimental conditions in which spoken sentences were: (a) both prosodically and semantically congruous: *Le chauffeur conduisait* 'The chauffeur was driving'; (b) prosodically incongruous and semantically congruous: *Le cycliste pédalait?* 'The cyclist was pedaling?' (*Statement); (c) prosodically congruous and semantically incongruous: *Le chauffeur clignotait* 'The chauffeur was flashing'; and (d) both prosodically and semantically incongruous: *Le cycliste consultait?* 'The cyclist was consulting?' (*Question).

This design allowed us to address three related questions. First, we were interested in finding out whether modality is processed in real time sand thus whether sentences with incongruous modality elicit different patterns of brain waves than sentences with congruous modality. Second, it was of interest to directly compare the time course of modality and semantic processing. In spoken language, the N400 component elicited by semantic incongruities is known to develop within the first 200 ms following word onset and to reach its maximum amplitude around 400 ms. Thus, we were interested in finding out whether prosodic modality incongruities would elicit an earlier or a later onset effect compared to the N400. Third, the linguistic functions of prosody are often considered to be intimately linked to other linguistic components such as syntax, semantics, or pragmatics. Here, we investigated the link between the modality function of prosody and semantics by testing whether their processing, as manipulated in the present experiment, is independent or interactive. In case of interactive effects, it was important to determine when such an interaction would occur. Finally, we also manipulated the participants' attention by using two different tasks in separate blocks of trials. They were asked to focus attention either on the prosody of the sentence, to decide whether the intonation contour was congruous or incongruous, or on the semantics to decide whether the sentences were semantically congruous or incongruous. We analyzed both the percentage of correct responses and the electrophysiological data time locked to the onset of the sentence.

Figure 1. F0 contours of the six-syllable statement and question sentences that we presented in the experiment for the prosodically congruous and incongruous sentences (adapted from Astésano et al., 2004).

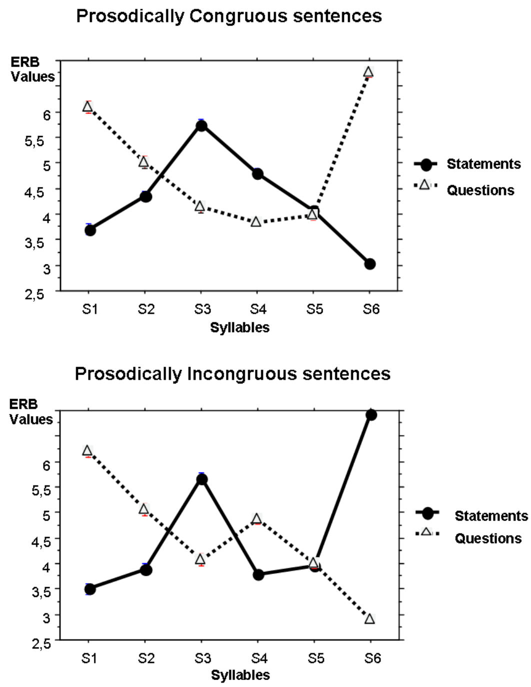
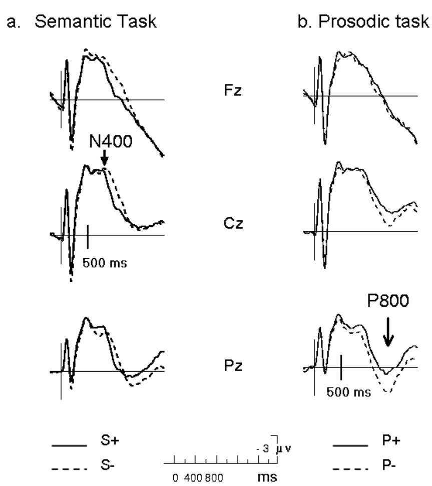


Figure 2. Event-Related Potentials (ERPs), time-locked to the first word of the sentence, and recorded at Frontal (Fz), Central (Cz) and Parietal (Pz) locations while participants performed the semantic and prosodic tasks. The overlapped ERPs are elicited by semantically (a) or prosodically (b) congruous (solid line) and incongruous words (dotted line). In this and subsequent figures, amplitude (µV) is in the ordinate (negative is up) and time (ms) is in the abscissa (adapted from Astésano et al., 2004).



The main outcomes in terms of accuracy were that the level of performance was higher in the semantic task than in the prosodic modality task and that, in this case, participants made more errors for sentences with prosodic incongruities. Thus, the prosodic task was clearly more difficult than the semantic task and prosodic modality incongruities were not always perceived accurately. In terms of ERPs, results in the semantic task showed that, as expected on the basis of previous literature, an N400 component is associated with semantic incongruities (Figure 2a). More interesting is the finding that, in the prosodic task, a P800 component is elicited by prosodic incongruities (Figure 2b). Thus, prosodic mismatches are processed in real time and the time-course of prosodic modality processing seems to take somewhat longer than the processing of semantic incongruities (but see below).

Finally, results revealed an asymmetric interaction between semantic and modality processing: while no effect of modality congruency was found when participants focused their attention on the semantics of the sentence (Figure 3a), semantic congruency influenced the processing of modality when participants' attention was focused on the intonation contour of the sentence (Figure 3b). Specifically, in this later condition, when the sentence final words were semantically congruous in the context, no P800 component was elicited. We interpreted this finding as an indication that the prosodic mismatch was not detected because the semantics of the sentence was congruous and semantic processing overrode modality processing. By contrast, when the words were semantically incongruous, a P800 was generated, which was taken as evidence that the detection of modality mismatches is facilitated by the presence of semantic incongruities.

Two aspects of these results deserve further comments. First, the late latency of the P800 component to modality incongruity compared to the earlier latency of the N400 component to semantic incongruity could be taken as evidence for a late onset of modality compared to semantic processing. While it may be the case that the semantic aspects of language are processed faster than the modality aspects because the main function of language is to convey meaning, this pattern of results may also reflect the greater difficulty in detecting prosodic in relation to semantic incongruities, as shown by the analysis of the correct responses. Thus, rather than reflecting intrinsic differences in the time course of prosodic and semantic processing, these results may reflect differences in task difficulty. While equating task difficulty is a general problem in cognitive science, it needs to be seriously considered as it may strongly influence the interpretation of the results. A further argument to consider in connection with the question as to whether the long latency of the P800 in the present experiment is related to the difficulty of modality processing comes from a study on pitch processing conducted in our laboratory (Schön et al., 2004). The pitch of sentence final words (F0) was increased by 30% in one condition and by 120% in the other condition. Compared to control sentence endings (no pitch manipulation), pitch variations were associated with increased amplitude of positive components and their latency was shorter for the stronger variations. Second, in order to examine the processing of modality we used the cross-splicing technique. We were very careful in building short (3 words) sentences that always ended with a trisyllabic word starting with unvoiced consonants so that

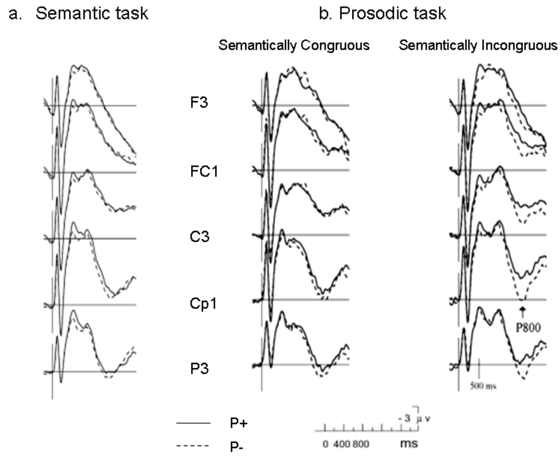


Figure 3. ERPs recorded from lateral electrodes at frontal (F3), fronto-central (FC1), central (C3), centro-parietal (Cp1) and parietal (P3) electrodes in the semantic and prosodic tasks when participants were listening to prosodically congruous and incongruous sentence final words (adapted from Astésano et al., 2004).

a silent phase of 50 ms preceded the stop release. The speech signal was then spliced at the silent phase of the consonant so that no acoustic cues were available to detect the signal manipulation at the splicing point. These sentences were consequently very well-controlled from an acoustic perspective. From a prosodic perspective, they were "prosodic monsters" (no such sentences exist in spoken language), but nevertheless possibly acceptable – prosody is characterized by considerable variability and what is prosodically acceptable or not largely depends upon the speaker, the listener and the context of enunciation. These seemingly contradictory aspects may explain why P800 components were elicited by prosodic incongruities (prosodic monsters) but only when words were semantically incongruous and the task focused participants' attention on the intonation contour.

In conclusion, the results of this experiment allowed us to partially answer the three main questions that were raised above. Modality is processed in real time since sentences with an incongruous prosodic modality elicited a different pattern of brain waves than congruous sentences. However, it remains unclear as to whether the long latency of the observed effect (P800) compared to the semantic congruity effect (N400) reflected delayed modality processing compared to semantic processing or differences in task difficulty. Finally, modality and semantic processing were found to be interactive in such a way that semantic processing was not influenced by modality processing but modality processing was influenced by the semantic content of the sentences. Thus, in the specific conditions of this experiment, semantic processing predominated over modality processing.

4. The pragmatic function of prosody

From a pragmatic point of view, speakers use prosody to organize their message and listeners use it to identify relevant information in an utterance. Communication is optimal when within the context of communication (i.e. the background of shared knowledge or the common mental space shared by both speakers, Fauconnier, 1984), the speaker successfully brings to the foreground the relevant new information (i.e. the figure) s/he wants to transmit to the listener. In other words, the pragmatic structure of discourse is organized around two main concepts, the topic or known information (what we are speaking about) and the focus or new, relevant information (what we are saying about it). For instance, if I ask you *Did your paper get accepted or rejected?* it is likely that you will answer by stressing the word *accepted* in the sentence *Of course, my paper got accepted* so that it would pop out from your answer. The process by which new information becomes cognitively relevant is known as *focus*. While focus can be realized either through the combination of syntactic constructions and prosodic cues or through prosodic cues only, we were specifically interested in the later, in the prosodic marking of focus, and we manipulated contrastive focus that induces an exclusive selection in a paradigmatic class (di Cristo, 1998). Thus, in the example above, the paradigmatic opposition between *accepted* and *rejected* is resolved by a prosodic prominence on the (highly!) relevant word *accepted*. Phonetically, prosodic prominences (focal accents) in French are realized by increases in pitch, duration, intensity, and F0 contour (Astésano et al., 2004; Astésano, 2001; Lacheret-Dujour & Beaugendre, 1999).

Based on this linguistic knowledge about the pragmatic function of focus, the main question we addressed in this experiment was whether listeners make an on-line use of prosodic prominences to process relevant information in discourse contexts (Magne et al., 2005). To this aim, we manipulated the pragmatic relevance of prosodic prominences so that they were either pragmatically congruous or incongruous relative to the discourse context. Short dialogues, composed of a question (Q) and an answer (A) were created that induced specific expectancy regarding the position of focal contrastive accents in the answer. In the examples below, the position of the focal accent (underlined) in [A1] is congruous with [Q1] but is incongruous with [Q2]. The reverse is true for [A2] that is congruous with [Q2] but incongruous with [Q1].

[Q1] Did your paper get accepted or rejected from Brain Talk?

[Q2] Did your paper get accepted in Brain Talk or in Cognitive Linguistics?

[A1] In fact, my paper got <u>accepted</u> in Brain Talk

[A2] In fact, my paper got accepted in <u>Brain Talk</u>

The originality of this design is that the same word in the answer (e.g. *accepted* or *Brain Talk*) can be pragmatically congruous or incongruous according to the preceding question and, consequently, it can be used as its own control. Moreover, this design also allowed examining the influence of the position in the sentence (middle or terminal) of words with focal accents. Participants were asked whether the intonation of the answer was coherent or not in relation to the question. The percent of correct responses and ERPs to words with focal accent were analyzed.

Results revealed that incongruous focal accents on mid sentence words (e.g. *accepted* in [Q2]–[A1]) were associated with increased positivity compared to the same words with congruous focal accents (e.g. *accepted* in [Q1]–[A1]; see Figure 4a). By contrast, incongruous focal accents on sentence final words (e.g. *Brain Talk* in [Q1]–[A2]; see Figure 4b) were associated with increased negativity compared to the same words with congruous focal accents (e.g. *Brain Talk* in [Q2]–[A2]). Interestingly, very similar effects were found when no focal accents were present on sentence middle and final words (see lower parts of Figure 4a and 4b) although they were expected based on the preceding context.

Two main conclusions can be drawn from these data. First, focal accents are processed in real time and their processing depends upon the position of words with focal accents within the sentence. When incongruous focal accents are located on words in mid sentence position, they elicited increased positivity in the 300–800 ms latency band following word onset. We interpreted this finding as reflecting the fact that the listener was surprised by the occurrence of the incongruous focal accent in the middle of the sentence. Indeed, it is known from previous literature that incongruous and unexpected events are associated with increased positivity belonging to the P300 family of components (Sutton et al., 1965; Donchin & Coles, 1988). By contrast, when incongruous focal accents are located on words in sentence final position, they elicited increased negativity peaking around 400 ms following word onset. We interpreted this increased negativity as an N400 component, reflecting the fact that the meaning of final words with incongruous focal accents was more difficult to integrate in the preceding sentence context than the meaning of final words with congruous focal accents. Indeed, the presence of an incongruous focal accent may force the listener to revise the interpretation of the sentence that was built based on the preceding question and beginning of the answer.

Second, the finding of very similar effects for words with incongruous focal accents and for words with no focal accents (whether in mid or final sentence positions) when focal accents are expected based upon the question clearly showed that it was not the processing of the incongruous focal accents that mattered but rather the pragmatic relevance of the focal accents in the discourse context. In other words, the effects were not linked with acoustic differences between words with and without focal accents. This is not to say, however, that the acoustic differences between words with and without focal accents were not processed. Indeed, precise analysis of the time course of the effects for words in middle and final sentence positions showed an earlier onset when incongruous focal accents were present than when they were absent but expected. Thus, differences in acoustic features influenced the onset of the effects but the observed differences most likely reflected the cognitive processing of the coherence between the word being presented (with or without focal accents) and the words that were expected based upon the pragmatic representation of the discourse context.

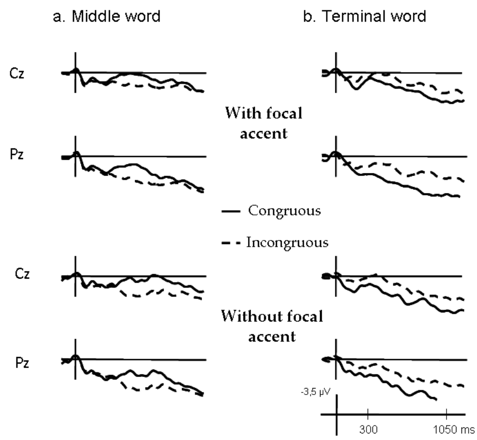


Figure 4. ERPs time-locked to a) mid sentence words and to b) terminal words with or without focal accents that were pragmatically congruous or incongruous in the discourse context. Recordings are from central (Cz) and parietal (Pz) electrodes (adapted from Magne et al., 2005).

5. The metric function of prosody

Previous results have shown that both adults and infants are sensitive to the rhythmic structure of their native language (e.g. Cutler et al., 2001; Nazzi et al., 1998; Ramus et al., 2003). Rhythmic patterns are clearly important for speech segmentation (recognizing meaningful units – words – within a continuous flow of auditory information) and consequently for the acquisition and comprehension of speech. Final syllabic lengthening, reflected by an increase in duration of the final syllable of the final word of prosodic groups, contributes to the rhythmic organization of languages and facilitates speech segmentation. For instance, French listeners use final syllable lengthening to speed up detection of a target syllable located at a rhythmic group boundary in comparison to the same syllable at another location (Dahan, 1996; Christophe et al., 2004).

To further examine the role of final syllable lengthening in speech comprehension, we manipulated orthogonally the metric and semantic aspects of sentence final trisyllabic words (Magne et al., 2007). We created a specific time-stretching algorithm to increase the duration of the penultimate syllable of final trisyllabic words without modifying their timbre or pitch. Moreover, we also created semantic mismatches by replacing the sentence final congruous words by semantically unexpected incongruous words. Thus, we used a 2x2 design with four experimental conditions in which final words of spoken sentences were: 1) both metrically and semantically congruous: *Le concours a regroupé mille candidats* 'The competition brought together one thousand candidates'; 2) metrically incongruous and semantically congruous *Le concours a regroupé mille can<u>di</u>dats* 'The competition brought together one thousand candidates'; 3) metrically congruous and semantically incongruous *Le concours a regroupé mille bigoudis* 'The competition brought together one thousand curlers'; and 4) both metrically and semantically incongruous *Le concours a regroupé mille bi<u>gou</u>dis* 'The competition brought together one thousand curlers'. Finally, participants were asked to focus their attention either on the metric aspects to decide whether the final word was well-pronounced or not (metric task) or on the semantic aspects to decide whether the final word was semantically congruous or incongruous in the sentence context (semantic task).

As in the Astésano et al. (2004) study we addressed three related questions. First, it was of interest to determine whether final syllable lengthening was processed in real time so that final words with increased duration of the penultimate syllable would elicit different patterns of brain waves than final words with congruous metric patterns. Second, we aimed at directly comparing the time course of metric and semantic processing. Finally, we tested the hypothesis that incongruous metric patterns may impede lexical access and word comprehension so that the error rate and amplitude of the N400 component would be larger than for final words with congruous metric patterns. Thus, we expected an interaction between metric and semantic processing.

At the behavioral level, results showed that the error rate was highest in the semantic task when final words were metrically incongruous and semantically congruous. Thus, increased duration of the penultimate syllable clearly seemed to impede lexical access even when words were semantically congruous. Interestingly, results at the electrophysiological level were in line with this interpretation: the amplitude of the N400 component was larger for metrically incongruous than congruous words when participants were focusing their attention on the semantics of the sentences. Thus, the occurrence of metric incongruities hindered lexical access and word comprehension as reflected by increased N400 amplitude. By contrast, when participants focused their attention on the metric aspects, the N400 was followed by larger P700 components to metrically incongruous than congruous words (Figure 5a) thereby reflecting the on-line processing of the metric structure. Finally, semantically incongruous words were associated with larger amplitude N400 components than semantically congruous words in both the semantic and metric tasks (Figure 5b). While such a result was expected when attention was focused at the semantic level, the finding of a similar N400 effect when attention was focused at the metric level was taken to reflect the automatic processing of the meaning of words, independently of the direction of attention.

6. Conclusion

Within these series of experiments we examined several linguistic functions of prosody, modality, focus, and meter segmentation (through final syllable lengthening). In all experiments, we found that prosodic cues were processed on-line and influenced ongoing sentence comprehension. Thus, results revealed interactive effects of modality and semantics, of prosodic focus and pragmatic constraints, and of meter and semantics. Taken together these results therefore favor an interactive view of language processing within which all available linguistic cues (phonetic, prosodic, semantic, syntactic, pragmatic, …) are processed as soon as they are available in order to build up a coherent representation of the utterances (Robert, 1999). These results are in line with recent linguistic theories in which the different aspects of language processing are not considered as structurally and functionally dissociable but rather as closely intertwined and tied together (Altman & Steedman, 1988). For instance, the frontiers between semantics

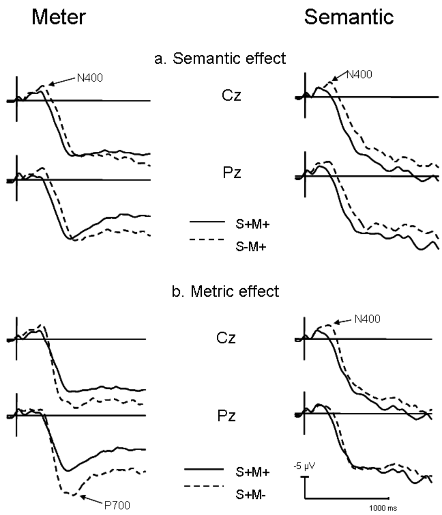


Figure 5. ERPs time-locked to sentence terminal words in the meter and semantic tasks. The overlapped ERPs are elicited by semantic effect (a): semantically congruous (solid line) and incongruous words (dotted line); and by metric effect (b): metrically congruous (solid line) and incongruous words (dotted line). Recordings are from central (Cz) and parietal (Pz) electrodes (adapted from Magne et al., 2007).

and syntax or between semantics and pragmatics are no longer considered as being as clear-cut as was believed during the last 50 years. While this may increase the difficulty in studying on-line language processing, because several types of processes may be going on simultaneously, this may also help develop new views of language processing.

7. Acknowledgements

The different experiments presented here were supported by grants to M. Besson from the International Foundation for Music Research (IFMR #RPA194), from the French Ministry of Research (ACI "Approche plurisdiciplinaire de la complexité linguistique", Dir: Dr Stéphane Robert) and from the Human Frontier Science Program (HFSP #RGP0053). The implementation of the experiments as well as the publication of their results benefitted from comments and suggestions from many colleagues: Monique Chiambretto, Thomas Günter, Reyna Leigh Gordon, Martin Meyer, Andras Semjen, Mario Rossi, Stéphane Robert, Bernard Victorri, and Yves-Marie Visetti.

References

Altmann, G. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 198–238.
Astésano, C. (2001). *Rythme et accentuation en français. Invariance et variabilité stylistique.* Paris: L'Harmattan.
Astésano, C., Besson, M. & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Brain Research. Cognitive Brain Research, 18*, 172–184.
Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten, 87*, 527–570.
Besson, M., Magne, C. & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Science, 6*, 405–407.
Böcker, K.B.E., Bastiaansen, M.C.M., Vroomen J., Brunia, C.H.M. & De Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology, 36*, 706–720.
Christophe, A., Peperkamp, S., Pallier, C., Block, E. & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51*, 523–547.
Cutler, A., Dahan, D. & van Donselaae, W. (1997). Prosody in the comprehension of spoken language: A literature review, *Language and Speech*, *40,* 141–201.
Cutler, A. & Van Donselaar, W.A. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*, 171–195.
Dahan, D. (1996). The role of rhythmic groups in the segmentation of continuous French speech. *Proceedings of the Fourth International Conference of Speech and Language Processing*, 1185–1188. Oct 3–6. Philadelphia.
Darwin, C.J. (1975). On the dynamic use of prosody in speech perception, *Status Report on Speech Research*, *42/43*, Haskins Laboratories, 103–115.
Dawson, G.D. (1951). A summation technique for detecting small signals in a large irregular background. *The Journal of Physiology, 115*, 2–3.
Di Cristo, A. (1998). French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge: Cambridge University Press.
Donchin, E. & Coles, M.G.H. (1988). Is the P300 component a manifestation of context-updating? *Behavioural and Brain Sciences, 11*, 355–372.
Delattre, P. (1966). Les dix intonations de base du Français, *French Review*, *40*, 1–14.
Fauconnier, G. (1984). *Espaces mentaux. Aspects de la construction du sens dans les langues naturelles*. Paris: Les Editions de Minuit.
Gårding, E. (1998). Swedish. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 112–130). Cambridge: Cambridge Univ. Press.
Hirst, D. (1998). British English. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 56–77). Cambridge: Cambridge Univ. Press.
Holcomb, P.J. & Neville, H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281–312.
Jusczyk, P.W. & Hohne, E.A. (1997). Infants' memory for spoken words. *Science, 277*, 1984–1986.
Kutas, M. & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–204.
Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307* (5947), 161–3.

Kutas, M., Neville, H.J. & Holcomb, P.J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing. *Electroencephalography and Clinical Neurophysiology, Supplement 39*, 325–330.

Lacheret-Dujour, A. & Beaugendre, F. (1999). *La prosodie du français*, Paris: CNRS Editions.

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K. & Besson, M. (2005). On-line processing of "pop-out" words in spoken French dialogues. *Journal of Cognitive Neuroscience, 17*, 740–756.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R. & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioural and electrophysiological evidence. *Cerebral Cortex, 17,* 659–668.

McCallum, W.C., Farmer, S.F. & Pocock, P.V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 59*, 477–488.

Molino, J. (2000). Toward an evolutionary theory of music. In N. Wallin, B. Merker & S. Brown, (Eds.), *The origins of music* (pp. 165-176). Cambridge, MA: MIT Press.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination in newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 756–777.

Piotrovski, D. (2008). Phénoménalité et objectivité linguistiques. Paris: Champion, Coll, *Bibliothèque de grammaire et de linguistique*, to appear.

Ramus, F., Dupoux, E. & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. Paper presented at the 15th International Congress of Phonetic Sciences (pp. 337–342). Barcelona, August 3–9, 2003.

Robert, S. (1999). Cognitive invariants and linguistic variability: From units to utterance. In C. Fuchs & S. Robert (Eds.), *Language diversity and cognitive representations (Human Cognitive Processing, Vol. 3)* (pp. 21–35). Amsterdam/Philadelphia: John Benjamins (English revised edition).

Schirmer, A., Kotz, S.A. & Friederici, A.D. (2002). Sex differentiates the role of emotional prosody during word processing. *Brain Research. Cognitive Brain Research, 14*, 228–233.

Schön, D., Magne, C. & Besson, M. (2004). The music of speech: Electrophysiological study of pitch perception in language and music. *Psychophysiology, 41*, 341–349.

Steinhauer, K., Alter, K. & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*, 191–196.

Sutton, S., Braren, M., Zubin, J. & John, E.R. (1965). Evoked potential correlates of stimulus uncertainty. *Science, 150*, 1187–1188.

Vaissière, J. (2005). Perception of intonation. In D.B. Pisoni & R.E. Remez (Eds.), *The handbook of speech perception* (pp. 236–263). Oxford: Blackwell.

The processing of emotional utterances: Roles of prosodic and lexical information

Carly Metcalfe, Manon Grube
Newcastle Auditory Group, Newcastle University

Damien Gabriel
Cognitive Neuro-Rehabilitation Laboratory, University Hospital, Geneva

Susanne Dietrich, Hermann Ackermann
Hertie Institute for Clinical Brain Research, University of Tübingen

Vivian Cook
Education, Communication & Language Sciences, Newcastle University

Andy Hanson
Medical School, Newcastle University

Kai Alter
Newcastle Auditory Group, Newcastle University

Abstract

The present study investigates the interaction between prosodic and lexical information in expressing and recognising emotional states in the encoding and decoding of speech.

The study deals with the processing of interjections in English and is based on previous studies using German data (Dietrich et al., 2006), where it was found that listeners classify interjections into two categories with respect to lexical ambiguity. In the current investigation, behavioural data indicate that interjections can be classified into the two major categories: interjections with low or high lexical information. High lexical information has been demonstrated to facilitate recognition of emotion due to the redundancy in the prosodic and lexical information, e.g. *yuk* can only be classified as disgust. Items with low lexical content in contrast are ambiguous and may be used to express different emotions, e.g. a single–vowel like /a/ may express happiness, surprise, or anger. The intended meaning and its correct interpretation depend solely on the prosodic realisation. The ambiguity in such low lexical interjections leads to greater uncertainty and increased reaction times.

1. Introduction

The expression of emotions in humans forms an integral aspect of communication and is based on a variety of mediums, such as facial expression, vocal expression and body language. Emotional vocal expression is dependent upon physiological factors which can alter the geometry of the vocal tract and the sounds being produced (Schirmer & Kotz, 2006).

Emotional expression using verbal and non-verbal sounds is a crucial part of our survival instincts, in humans as in other species: expressing emotions such as disgust and anger has evolved as a survival strategy. For example, if a mother looks at rotting food and displays negative emotions towards it, such as the exclamation *Yuck!*, her offspring will witness this and learn to associate negative emotions with rotting food as something not to be eaten, thus helping to avoid eating poisonous or unfit food.

In emotional communication, small vocalizations called interjections are used frequently in everyday life (Abelin, 1999): they play a key role and are the focus of the present study. Interjections encompass two main components, lexical and prosodic information, either or both of which can be present. Lexicality refers to the emotion associated with the word-like item used; this relies upon long-term memory acting somewhat like a dictionary, linking the sound with its emotional meaning, such as *wow* with something amazing or impressive. Prosody in the present study refers to the accompanying sound and melodic structure of interjections. The definition of an interjection has remained elusive in previous approaches, namely traditional and modern ones, taking place over the course of the 20th century. Jesperson (1922) classified interjections as "instinctive ejaculations called forth by pain or other intense sensations or feelings", which are non-words, and later categorised two different types of interjections, primary and secondary ones. Primary interjections are true interjections which are only used to express emotion (*ouch*, *yuck*, and *wow*), whereas secondary interjections encompass words with an independent or iconic meaning in addition to an emotional meaning (*all right*, *no way*). More recently, interjections have been classified as a type of word, universally associated with emotions and combining thought and feeling (Wierzbecka, 1992). Kryk (1992) and Abelin (1999) categorised interjections amongst onomatopoeic and discourse particles, i.e. somewhere in-between the two on a sliding scale.

A number of approaches have attempted to rectify this issue (Ameka, 1992; Wharton, 2003; Meng & Schrabback, 1999). All agree that interjections are short, verbal utterances that are at times used to convey emotion, such as *yuck*, *yippee*, and *ouch*. They are part of early language development (prior to twenty-four months of age) and are learnt, i.e. language-specific and not innate (Meng & Schrabback, 1999).

For the purpose of the present study, emotional interjections were categorised into two groups depending upon their meaning: ambiguous and non-ambiguous (Dietrich et al., 2008). Ambiguous utterances do not have a lexical meaning and depend solely on prosody and context for their interpretation. They can be produced in different ways depending upon the context. For instance, *oh* can be either positive ("suddenly the answer came to you") or negative ("the shopkeeper told you that they were out of stock"). For ambiguous interjections, it is impossible to identify the underlying emotion without additional contextual or prosodic information. In the non-ambiguous category in contrast, there is a clear, stable emotional and lexical meaning; e.g. *yum* or *ugh* having a positive and negative meaning, respectively. Combined with prosody, non-ambiguous interjections convey a highly salient emotional meaning – with the obvious exception of sarcasm.

If an interjection is uttered with a list intonation however, the prosodic information is missing, and only lexical information is conveyed. The use of list intonation thus allows for a mismatch condition to be created whereby the interjection is pronounced without the appropriate emotional prosodic information in a more or less neutral manner.

The individual contributions of both lexical and prosodic information to the listener's understanding of an interjection remain to be explored. Very few previous studies have investigated this and only little is known about the use and processing of interjections in the English language. In 1992, Kryk conducted a study on the Polish interjection *no* and contrasted it with its English equivalent *well*. Although Kryk's work allowed for a new classification of an interjection, it did not look into the contributions of the lexical and prosodic components.

Dietrich et al. (2006) investigated German interjections and found that both prosodic and lexical elements were required for emotional understanding, wherein prosody appeared to play a more dominant role. This hypothesis was developed in the present study on English interjections: the aim was firstly to obtain recordings of a variety of interjections, including non-ambiguous and ambiguous ones, and secondly, to investigate the perception of their emotional content. The results will allow comparison of the contributions of prosodic and lexical information as separate components in the perception of the emotional content of English interjections, how the presence of both compares with that of only one component and whether either of the components plays a more dominant role in emotional expression and understanding. The results will in addition aid in the classification of English interjections.

In the first part of the study, recordings of English emotional interjections were obtained for a number of interjections containing either one or both of the components of interest, i.e. lexical or prosodic information. Recordings were obtained from six female speakers, with interjections produced with a normal meaningful intonation, as well as with list intonation, i.e. without prosodic information. In the second part, the emotional content of the interjections was rated as a measure for investigating and disentangling the listeners' processing of prosodic and lexical information in the interpretation of the emotional content of interjections. In the present study, we used the term *valence* to denote a widely accepted measure for the categorisation of emotions into a positive or negative percept.

The study contributes to rectifying the current lack of research into English interjections by providing insights into one of the many aspects of emotional processing. The use of different categories of interjections will allow for comparisons to be made between the processing of lexical and prosodic information and furthermore allow the selection of the most appropriate recordings for further studies, including those involving neurological patients.

2. Speech recordings

2.1. Participants

Six female volunteers (mean age: 21 years; age range: 20–22 years) participated in the recordings, providing written consent. All were monolingual native English speakers, with no prior knowledge of the purpose of the study. All were in good health reporting no oral or neurological disorders. Monetary reimbursement was provided for participation.

2.2. Recordings

Recordings were undertaken in a sound-proof room, in which a microphone (Samson PS01) and buzzer (Radio Shack two-way intercom) were available. Cool Edit 2.1 was utilised to record at 44.1 kHz with a 16 bit resolution rate. The sound card used was Edirol Audio Capture. All recordings were saved on a PC and edited with Cool Edit 2.1, where the usable ones were normalised for amplitude to 75% and cut to appropriate length.

2.3. Procedure

Participants were seated in a soundproof room at a desk on which a microphone (20 cm from the speaker) and buzzer were placed. Written instructions were provided, informing the participant of the procedure and the task to be performed. For the production of each interjection with normal prosodic intonation, participants were provided with stories[1] and asked to read these silently and then produce the given interjection in an appropriate manner a number of times, i.e. to produce at least three usable recordings. This was done for each of the interjections used; for a list of all interjections used, see Table 1. A practice block was provided after which four blocks of recordings took place, involving positive, negative, and ambiguous interjections, respectively, and lastly, all of the aforementioned interjections spoken with list intonation. Between each of the blocks, participants took breaks as desired. The total time for each participant to complete all blocks was approximately thirty minutes.

A total of three recordings per interjection were selected per participant for use in the behavioural ratings experiment. The selection of only three was made by the experimenter based on the clarity of the production, avoiding those that contained interference like coughs, clicks etc. Duplicates were made of the usable recordings in a few exceptions where less than three individual recordings were suitable. The total number of selected recordings was 720, including 198 positive, 108 negative, 108 negative ambiguous, 108 positive ambiguous and 198 list intonation recordings produced by 6 female speakers.

---

1 An example of one of the stories utilised to elicit the interjection *ouch*: "Now that the wax had been applied, it was too late to back out now. As the cloth strip was applied doubt began to grow, and anticipation as seconds later, in one swift motion, the strip was pulled off."

3. Perceptual ratings experiment

The purpose of this experiment was to determine whether the performance and reaction rates of participants in addition to the identification of optimal stimuli obtained in the speech recordings. The data obtained will allow for comparisons of the responses and reaction times between the different valences.

3.1. Participants

Sixteen volunteers (8 female, 8 male; mean age: 21 years; age range: 20–22 years) participated in the behavioural study, providing written consent. All were native English speakers, right-handed and with no prior knowledge of the purpose of the study. All were in good health reporting no auditory or neurological disorders. Monetary reimbursement was provided for participation.

3.2. Procedure

From the obtained speech recordings, 918 usable interjections were used with each interjection occurring three times per speaker for each item in the positive/negative ambiguous, positive/negative non-ambiguous, and list intonation valence classes.

Participants were seated in a soundproof room in front of a computer screen, provided with headphones and a keyboard. Written instructions were given at the beginning of the session. Participants were required to listen to the interjection recordings via the headphones and to categorise each interjection as positive or negative, by pressing either one of two keys labelled with "-" and "+", followed by the enter key. The time allocated per recording was 3 seconds: 1 second for the recording to be played, 1 second for the participant to categorise the interjection and enter the response, and a 1 second break interval prior to the start of the next recording.

A practice session block was provided, using a variety of interjections that had been obtained during the practice blocks of the speech recording sessions, followed by a short break. In the main session, all 720 interjections were played in a randomised order taking about 60 minutes in total.

MATLAB 6.5 was used to play the recordings, collect the participants' responses and measure the reaction times. Positive and negative responses were evaluated as "+1" and "–1", respectively.

3.3. Data Analysis

Mean responses and reaction times of each participant for each interjection were calculated.

The independent variables of interest were class of emotional valence (positive, negative, positive, and negative ambiguous, or list-intonation), interjection, speaker.

One-sample *t*-tests were performed to determine whether mean responses for a given class, interjection, or recording were significantly different from zero, i.e. consistently positive or negative. An ANOVA was performed in addition to compare effects

Table 1. Interjections used in the study sorted by the categories that each interjection falls into. Numbering is used to denote the interjections. Two different types of emotional valence were employed, positive and negative.

Valence class	Interjections
Positive (1)	woohoo, hey presto, yum, wow, yippee, hurray, aha, phwoar, cheers, aww, ooh (1–11)
Negative (2)	oh dear, whoops, crikey, ugh, yuck, ouch (12–17)
Ambiguous: Positive intonation (3) Negative intonation (4)	okay, oh, ah, all right, no way, hmmm (18–23)
List Intonation (5)	woohoo, hey presto, yum, wow, yippee, hurray, aha, phwoar, cheers, aww, ooh, oh dear, whoops, crikey, ugh, yuck, ouch, okay, oh, ah, all right, no way (1–22)

of speaker on responses and reaction times between the different classes of interjection (positive/negative/ambiguous/list intonation). The significance level was $p < .05$.

3.4. Results

Interjections of the non-ambiguous positive and negative valence classes were overall consistently rated as positive and negative, respectively, as is reflected by the largely positive and negative ratings displayed in Figure 1. The group mean ratings for all interjections of valence class 1 (non-ambiguous positive) were all positive, with values ranging from 0.35 to 0.97 and all reaching significance ($p < .05$). The greatest variation in responses occurred for interjections 7, 8 and 10 (*aha*, *phwoar*, and *aww*) as indicated by the error bars in Figure 1. For valence class 2 (non-ambiguous negative), there was an increased variation of responses across the interjections, with group mean values ranging from –0.83 to 0.08. Ratings for four out of the six interjections used reached significance, i.e. interjections 12, 15, 16 and 17 (*oh dear*, *ugh*, *yuck*, and *ouch*) and further showed a similar variance amongst individual participants (0.1125–0.1418).

The ambiguous valence classes 3 (ambiguous positive) and 4 (ambiguous negative) yielded mainly positive and negative ratings, respectively. The same 5 out of 6 interjections (18–23) in either class reached significance. For valence 3 (ambiguous positive), the significantly positive mean ratings ranged from 0.38 to 0.92 which was a greater variation than that seen for valence 4 (ambiguous negative, –0.43 to –0.74). The only ambiguous interjection that did not yield significantly positive or negative ratings in either class was interjection 22 (*no way*), with mean rating values of 0.02 and –0.10 in its positive and negative valence, respectively.

For the list-intonation valence class 5 containing all of the above interjections produced with neutral prosody, mean ratings for 13 out of 22 in total were significantly

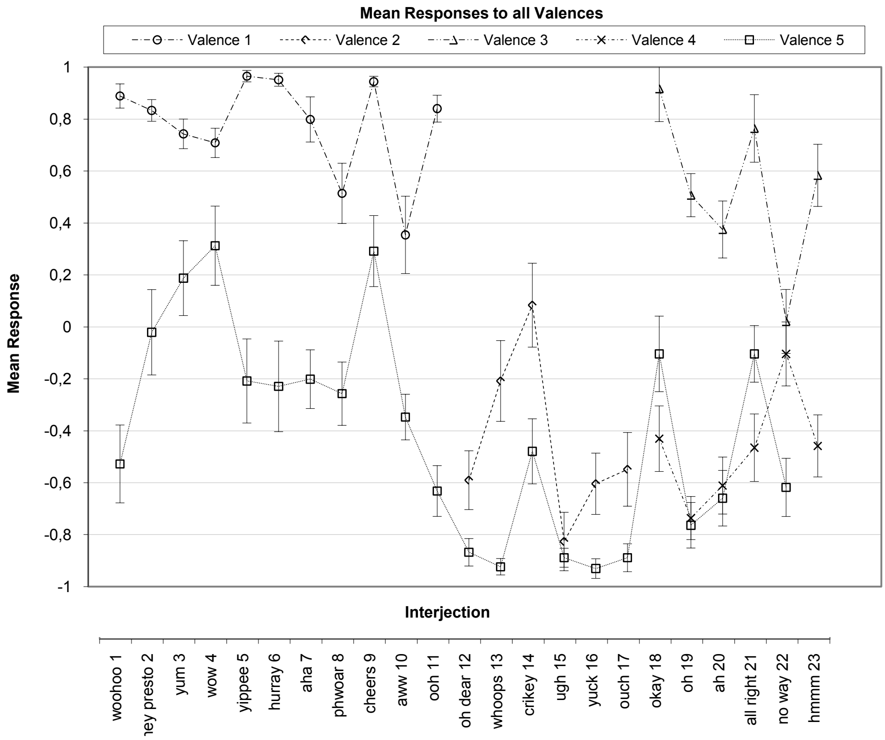


Figure 1. Mean ratings for all interjections of valences 1 (non-ambiguous positive, 1–11); 2 (non-ambiguous negative, 12–17); 3 (ambiguous positive, 18–23); 4 (ambiguous negative, 18–23); and 5 (list-intonation, 1–22). The graph shows the mean ratings by participants identifying interjections as either positive (+1) or negative (–1). Error bars represent the standard error of the mean. The interjections are numbered as in Table 1 and the corresponding words shown on the bottom x-axis.

different from 0. Twelve of those were significantly negative ratings. The only significantly positive ratings for list-intonation items occurred at interjection 9 (*cheers*). Seven out of the eleven corresponding interjections for the original valence 1 (non-ambiguous positive) were insignificant (2–8, inclusive), and the remaining two insignificant results occurred in the interjections 18 and 21 (*okay* and *all right*).

Trends in responses can be seen in Figure 1, showing that the majority (82%) of responses to valence 5 were negative. Although there was a significant difference between valence 5 and valence 2, this was not the case with valence 5 and 4.

In order to specifically address the processing of interjections that are consistently perceived as emotionally valid, the reaction times were analysed for the interjections with significantly positive or negative ratings. Mean reaction times for the three main

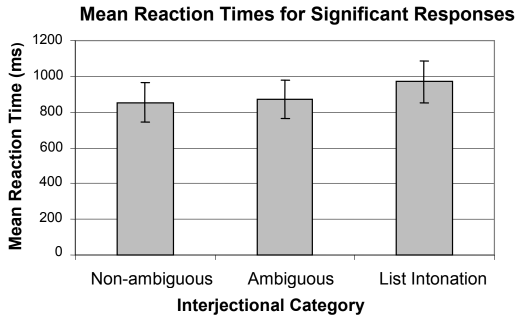


Figure 2. Mean reaction times for each interjection category for data that yielded significantly positive or negative ratings. Error bars represent the standard error of the mean.

categories of stimuli: non-ambiguous interjections, ambiguous interjections and those produced with list intonation are displayed in Figure 2.

The ANOVA performed on reaction times yielded a highly significant effect of category of interjection ($p < .001$). Reaction times were significantly faster for both non-ambiguous and ambiguous interjections than for those produced with list intonation ($p < .001$ and $p < .01$, respectively).

4. Discussion

Both of the studies conducted were original, since there were no previously recorded materials available for this kind of investigation. The choice of both interjections and speakers was made without any priming or preconceived predications. Variations in the results were therefore expected to occur as a reflection of the perception of individual speakers and interjections chosen.

One method for the selection of optimal stimuli would be based on discarding any speakers for whom rating results differed significantly between the groups of valences employed here. Using this method, the stimuli obtained from speakers 2 and 3 would be unsuitable for further studies.

With regard to the non-ambiguous and ambiguous interjections utilised in the studies, those that did not produce consistently significant responses need to be excluded. Interjection 14 (*crikey*) was the only interjection in failing to be properly categorised. This could be due to lack of usage among participants in the behavioural study leading to lower word familiarity.

Clear and significant differences in both perceptual ratings and the reaction times taken by individuals to categorise the emotional content for different categories of interjections based on their content of the key components of lexical and prosodic information were found. The categories of interest were non-ambiguous interjections containing lexical information and produced with matching prosodic (positive or negative) information, ambiguous interjections with no lexical information and produced with positive or negative prosody, and interjections produced with list intonation containing no prosodic information.

Reactions to the list intonation category which took significantly longer than those for any category containing prosodic other group, suggest that purely lexical contributions take longer to recognise and respond to, compared to both lexical and prosodic, and solely prosodic information. Reaction times were shortest for non-ambiguous interjections, followed closely by ambiguous interjections, where both are produced with prosodic information, and longest for list intonation implying that both factors promote the emotional categorisation of utterances.

The large difference between list intonation and ambiguous interjections further indicates that listeners are more reliant upon the prosodic information in order to categorise emotions. This is particularly relevant in the British culture where sarcasm is a prominent feature; the ability to distinguish the speaker's prosody and intonation is of critical importance. Previous research into how prosodic and intonational information contribute to determine whether or not a speaker is utilising sarcasm has produced contradictory findings (Rockwell, 2007). Another hypothesis may be that listeners are simply surprised by the lack of emotional prosody in items spoken with a neutral tone in list intonation. This again necessitates further studies by manipulating emotional prosody, e.g. by means of appropriate speech filtering procedures.

Valences 1 to 4 followed the expected trends as either positive or negative. Valence 5 produced mainly negative responses. This is probably due to the lack of lexical information encoded mimicking a more negative tone. This was particularly evident in valence 1 whereby only interjection 9 (*cheers*) elicited a significant positive response. Three of the remaining interjections (1, 10 and 11 – *woohoo*, *aww*, and *ooh*, respectively) in this valence achieved a significantly negative response with the rest failing to reach significance.

Another possible explanation for increased reaction times and performance rate for valence 5 could be that interjections spoken with a negative/neutral tone are low-arousal stimuli. However, this needs to be investigated further since in this study, the combination and/or difference of valence and arousal was not in focus.

5. Summary

The present study has provided novel stimuli and significant insights into the previously un-researched field involving the processing of emotional utterances in English. Optimal stimuli have been identified for future investigations into the processing of interjections in healthy subjects as well as neurological patients. Concerning the disentanglement of lexical and prosodic information and their roles in emotional expression and processing, the question has yet to be fully answered. Based upon the evidence collected in the present study, it seems likely that lexical information plays a less central role than prosodic information, but that the two combined are most efficient in communicating emotions, consistent with previous work by Dietrich et al. (2006). This, in turn, supports the hypothesis that interjections interpreted using prosodic information alone take longer or are harder to elicit a response to as compared to interjections containing unambiguous lexical information accompanied by the appropriate prosody.

The present study provides a useful and powerful methodological basis for further behavioural studies involving patients with neurological disorders, brain damage or deficits in emotional identification or processing. In cases where the ability to communicate using interjections in these patients remains on a par with healthy individuals, patient-care can be greatly enhanced by increased usage of interjections for effective communication of affective information.

Functional imaging studies and studies investigating EEG responses based on the present type of stimuli will enhance our understanding of the processing of emotional expressions in the human brain, knowledge of which is only in its infancy (Dietrich et al., 2006; Scott & Wise, 2003; Dietrich et al., 2006, 2008).

6. Acknowledgements

This work was generously supported by The Lena Teague Bequest for research into schizophrenia.

References

Abelin, Å. (1999). *Studies in sound symbolism.* Ph.D. dissertation. Gothenburg Monographs in Linguistics 17. Gothenburg: Gothenburg University, Dept. of Linguistics.

Ameka, F. (1992). Interjections: The universal yet neglected part of speech. *Journal of Pragmatics*, *18*, 101–118.

O'Connell, D. C. & Kowal, S. (2005). Where do interjections come from? A psycholinguistic analysis of Shaw's *Pygmalion*. *Journal of Psycholinguistic Research*, *34*, 497–514.

Dietrich, S. A., H., Szameitat, D. P. & Alter, K. (2006). Psychoacoustic studies on the processing of vocal interjections: How to disentangle lexical and prosodic information? *Progress in Brain Research*, *156*, 295–301.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A. & Ackermann, H. (2007). Semiotic aspects of human nonverbal vocalisations: A functional magnetic resonance imaging study. *Neuroreport, 18*, 1891–1894.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A. & Ackermann, H. (2008). Understanding the emotional expression of verbal interjections: A functional MRI study. *Neuroreport*, *19*, 1751–1755.

Jesperson, O. (1922). *Language: Its nature, development and origin*, London: Allen & Unwin.

Kryk, B. (1992). The pragmatics of interjections:The case of Polish *no*. *Journal of Pragmatics*, *18*, 193–207.

Meng, K. & Schrabback, S. (1999). Interjections in adult-child discourse: The cases of German HM and NA. *Journal of Pragmatics*, *31*, 1263–1287.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C. R., Bullmore, E. T., Brammer, M. & Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society*, *London, 265*, 1809–1817.

Rockwell, P. (2007). Vocal features of conversational sarcasm: A comparison of methods. *Journal of Psycholinguistic Research*, *36*, 361–369.

Schirmer, A. & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Review: Trends In Cognitive Science*, *10*, 24–30.

Scott, S. K. & Wise, R. J. S. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication*, *41*, 23–34.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E. & Wise, R. J. S. (2006). Converging language streams in the human temporal lobe. *The Journal of Neuroscience*, *26*, 7328–7336.

Wierzbecka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. New York: Oxford University Press.

Wharton, T. (2003). Interjections, language, and the 'showing/saying' continuum. *Pragmatics and Cognition*, *11*, 39–91.

Prosody and the brain

Inger Moen
Department of Linguistics and Scandinavian Studies, University of Oslo

Abstract

A key issue in the investigation of prosody in the brain damaged population is the search for the neuroanatomical underpinnings of the various prosodic functions. If the physical correlates of prosody, fundamental frequency, intensity, and duration, are of primary importance when it comes to determining hemispheric specialization, we would expect unilateral brain lesions to result in the same type of prosodic disturbance, regardless of the linguistic function of the feature. For instance, if unilateral brain damage leads to problems with the modulation of fundamental frequency, we would expect the problem to be evident regardless of the linguistic function of this feature. If, on the other hand, the linguistic functions are crucial when determining hemispheric specialization, we would expect the behavioural effect of unilateral lesions to vary depending on the linguistic functions of the particular physical dimension. A third possibility would be that both hemispheres are engaged in the processing of prosody, but that they process different acoustic features, for instance that one hemisphere processes fundamental frequency and the other hemisphere processes time. The paper discusses clinical studies, with particular emphasis on Norwegian data, which throw light on the question of hemispheric specialization of prosody.

1. Introduction

Most studies of prosody and the brain have used clinical material, brain damaged patients, as a basis for the investigations. The underlying assumption is that if damage to a particular brain area results in deviant language, then that area is central in the processing of the damaged language functions. In recent years, neuroimaging techniques in the study of brain activity in normal speakers have also been used (for a review, see Ross & Monnot, 2008). I shall primarily discuss the clinical studies, and as far as possible use illustrations from the language I am most familiar with, namely Norwegian.

The interest of clinical linguists in the prosodic features of speech goes back to Monrad-Krohn's famous study of a Norwegian patient with the so-called "foreign accent syndrome". She sounded German while speaking Norwegian. What particularly struck Monrad-Krohn as odd in connection with her deviant speech was her altered prosody (Monrad-Krohn, 1947, p. 411):

> The variations of pitch had by no means disappeared. On the contrary they were rather greater than usual in Norwegian, but they were neither adequate nor quite constant in their inadequacy.

Monrad-Krohn's patient had suffered left hemisphere damage, but the clinical-perceptual impression is that damage to either hemisphere may lead to deviations of sentence prosody. Acoustic investigations have revealed both normal and abnormal characteristics in the intonation of Broca's and Wernicke's aphasics and right-hemisphere-damaged patients. It is, however, unclear to what extent these abnormalities are linguistic in nature and to what extent they are caused by poor control of the physiological mechanisms associated with phonation and fundamental frequency variation, or whether they are the result of a deficiency in the long-range planning of linguistic units (Danly & Shapiro, 1982; Ryalls, 1982; Cooper et al., 1984; Shapiro & Danly, 1985; Cooper & Klouda, 1987).

1.1. The prosodic features of language

The linguistic prosodic features, stress, rhythm, tone, duration, and intonation, correlate with auditory variations in pitch, loudness, tempo, and rhythm. The physical dimensions of prosody are fundamental frequency, intensity, and time. These combine to form the linguistic prosodic units. For instance, stress is in many languages expressed by means of a combination of fundamental frequency, intensity, and duration. Emotional prosody, often referred to as tone of voice, constitutes the background for the perception of the linguistic elements of speech. The vocal parameters of the linguistic elements and of the tone of voice differ mainly in the time scale involved. A particular tone of voice in general covers several utterances, whereas the linguistic vocal parameters are more instantaneous. Increased loudness over a stretch of speech, for instance, is sometimes used to signal anger. At the same time, variations in loudness may also be used to distinguish stressed from unstressed syllables.

1.2. Parts of the brain active in prosodic processing

During the last decades data have accumulated on prosodic disorders as a result of brain damage, data which have produced several major hypotheses concerning the neuroanatomical regions active in prosodic processing. The most straightforward of these hypotheses assumes that all aspects of prosody are processed in the right hemisphere and integrated with linguistic information from the left hemisphere via the corpus callosum (Klouda et al., 1988). Van Lancker (1980) proposes a hypothesis which posits that affective or emotional prosody is controlled by the right hemisphere, whereas the left hemisphere is specialized for linguistic prosody. This is referred to as the functional lateralisation hypothesis. This hypothesis assumes that there is a scale of prosodic features from the most linguistically structured (e.g. Chinese and Norwegian tones) associated with left-hemisphere specialization to the, presumably, least linguistically structured (e.g. emotional tone, voice quality) associated with right-hemisphere specialization. A third major hypothesis assumes that the perception and production of prosody are mainly subserved by subcortical regions and are not lateralized to either of the hemispheres (e.g. Cancelliere & Kertesz, 1990). Several recent studies have supported the hypothesis that individual acoustic cues to prosody may be independently lateralized (e.g. Van Lancker & Sidtis, 1992).

2. Linguistic prosody in patients with brain damage

2.1. Word tones

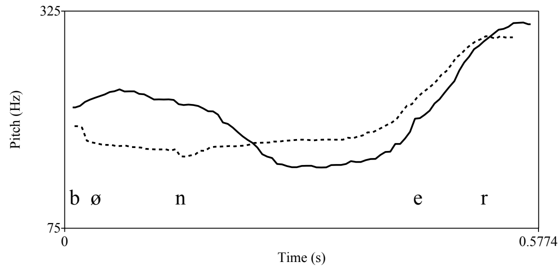
In tone languages differences in pitch at syllable or word level are used to distinguish the lexical meaning of words. In these languages, pitch variations are contrastive in the lexicon. There is general agreement that the primary acoustic correlates of tone are states and movements of the fundamental frequency of the voice (F0), although there may be concomitant changes in time, intensity, and phonation. In Norwegian, a so-called pitch accent language, the tonal contrast is restricted to one syllable in a word and pitch and stress are closely linked. The accented syllable carries one of two phonemically contrastive tones, referred to as Accent 1 and Accent 2 (Figure 1). There are a number of minimal pairs differing only in tone (see e.g. Moen & Sundet, 1996).

Tone languages are quite numerous. However, when it comes to the investigation of language as a result of brain damage, the data from tone languages come from only a limited number of languages, to be presise, from different dialects of Chinese, from Thai, from Swedish, and from Norwegian.

The perception of tone after unilateral brain damage

Studies of the perception of tonal contrasts in brain-damaged speakers of tone languages point to a left-hemisphere dominance in tonal perception (Hughes et al., 1983; Naeser & Chan, 1980; Yiu & Fok, 1995; Gandour & Dardarananda, 1983). Moen and Sundet (1996) investigated the perception of the two tonal contrasts in East Norwegian in eight brain-damaged speakers, four left-hemisphere-damaged aphasic speakers and

Figure 1. The fundamental frequency contours of the Norwegian word tones. Accent 1 (dotted line): *bønder* 'peasants'; Accent 2, (solid line): *bønner* 'prayers', 'beans'.



four right-hemisphere-damaged nonaphasic speakers. The aphasic speakers included two Broca's and two anomic aphasic patients. There was a control group of ten normal speakers. The control group identified all the words correctly. The right-hemisphere-damaged group's performance was comparable to that of the normal controls: only one incorrect identification. In the left-hemisphere-damaged group, on the other hand, only one of the four patients identified all the target words correctly. The other three patients identified from 50% to 92% of the words correctly. The confusion between the two accents was bidirectional: 17% af Accent 1 stimuli and 20% of Accent 2 stimuli were reported as the other accent.

The production of tone after unilateral brain damage

Studies of tone production in the speech of aphasic speakers of Mandarin, Cantonese, Thai, and Norwegian indicate that left-hemisphere-damage may lead to deviant tone production, whereas the tone production of right-hemisphere-damaged patients is relatively spared. Deficient tone production also seems to depend on type and severity of aphasia and time since onset of brain damage. A production deficit is most evident in the severly aphasic patients and more evident shortly after onset of brain damage than later. Non-fluent Broca's aphasics have more problems producing correct tonal contrasts than fluent Wernicke's or conduction aphasic patients (Naeser & Chan, 1980; Packard, 1986; Gandour, 1987; Gandour et al., 1988; Gandour et al., 1992; Ryalls & Reinvang, 1986). An investigation of tonal production in East Norwegian aphasic patients (one Broca and two anomic) tested the ability to differentiate between minimal pairs of words differing only in tone. The test material was a set of drawings illustrating minimal pairs of this type. The target words were written below each drawing. The three aphasic patients all failed to produce correct distinctions between the two accents in some of the minimal pairs (Moen & Sundet, 1999).

To conclude: investigations of the tonal perception and production in brain damaged speakers of tone languages have demonstrated reduced ability in both tasks in left-hemisphere-damaged patients, whereas tonal production and perception is spared in right-hemisphere-damaged nonaphasic patients, indicating that the processing of lexical tones is the property of the left hemisphere. Whether the reduced, but not disrupted, performance of the left-hemisphere-damaged patients is due to residual mechanisms in the left hemisphere, to mechanisms in the right hemisphere, or to subcortical mechanisms is unclear.

2.2. *Lateralisation of stress, rhythm, and intonation*

According to the functional lateralisation hypothesis one would expect linguistic sentence prosody to be lateralised to the left hemisphere. Investigations of the rate of speech in nonfluent aphasics have demonstrated a general slowing down of sentence production with longer syllable and word durations, as well as longer pauses between words, than found in normal speech (Gandour et al., 1994; Danley & Shapiro, 1982; Kent & Rosenbek, 1982, 1983). Investigations of timing relations at the levels of the syllable and the word have reported conflicting results. An investigation of the duration of the stem vowel in words of increasing length produced by a group of nonfluent aphasics (Collins et al., 1983) found that, despite longer absolute durations, the patients shortened the duration of the stem vowel on a relative scale to a degree comparable to that of normals. A similar study (Baum, 1990, 1992), however, found deviant intersyllabic timing relations in the speech of nonfluent aphasics. Baum measured syllable durations in triplets of words increasing in length from one to three syllables, all containing the same root syllable. The aphasic patients exhibited the expected shortening of the root syllable in two-syllable words. But unlike the speech of normals, there was no further decrease in root syllable duration in three-syllable words.

Kent and Rosenbek (1983) looked at variations in intensity in a group of patients with damage in Broca's area and found a significant reduction in the normal syllable to syllable intensity variation for these subjects compared with normal speakers of the same stimuli. Similar results were found in a case study of a Thai speaker (Gandour et al., 1989). This patient also had reduced intensity variation from syllable to syllable compared to a normal speaker in a reading test. The observed flattening of the patient's intensity variation is seen to reflect her inability to produce appropriate variations in stress patterns in larger-sized temporal domains.

Acoustic investigations of fundamental frequency variations (F0) at the level of the sentence have revealed both normal and abnormal characteristics. Danly and Shapiro (1982), investigating English speaking aphasic patients, found that Broca's aphasics retained the normal sentence-final fall in fundamental frequency in short sentences, but that the F0 declination was less pronounced in longer utterances, supporting the clinical impression of monotonous speech. Several studies have reported abnormal durational control but normal lowering of F0 values from the beginning to the end of sentences. Other findings include exaggerated (Danly & Shapiro, 1982; Cooper et al., 1984; Ryalls, 1984) and restricted (Ryalls, 1982) F0 ranges, as well as greater than normal F0 variability (Ryalls, 1984). However, an investigation by Seddoh (2000) found F0 contours comparable to the performance of normal subjects. And a recent study by Baum and Pell (1997) also indicated that the ability to produce normal intonation patterns may be relatively spared for Broca's aphasic patients.

In Norwegian, the prosodic features at sentence level interact with the tones which are features at the syllable and word level. In Norwegian there are three distinct levels of prominence at the sentence level: stress, accent, and focus. Stressed syllables contrast with the unstressed ones. A stressed syllable may be accented, it may have the pitch pattern of either Accent 1 or Accent 2. An accented syllable may be focal or non-focal. Words with a focal accent end in a sharp rise. The largest unit in the prosodic system is the prosodic utterance which may end in a high or a low boundary tone (see e.g. Moen, 2004).

In order to test to what extent left-hemisphere-damaged patients had problems producing sentence prosody, a reading test involving three aphasic patients (Broca's aphasia) and two control subjects was conducted. The subjects were asked to read target words with either Accent 1 or Accent 2 in a fixed syntactic frame (Moen, 2004). The readings were recorded and analysed acoustically (Figure 2). The analysis showed that the patients' timing and rhythm were abnormal compared to the controls: syllables, words, and utterances were abnormally long. The increased duration was partly due to interspersed pauses between individual syllables and words. There was also a tendency to give all syllables equal prominence, in other words to fail to distinguish between stressed and unstressed syllables. The fundamental frequency contours, on the other hand, were relatively normal. There was a clear distinction between Accent 1 and Accent 2. Focal accent was indicated with a rise and the intonational utterance ended in an appropriate low boundary tone. This study does not support the hypothesis that all features of sentence prosody are the property of the left hemisphere. It supports the hypothesis that prosodic features may be independently lateralised.

3. Lateralisation of emotional prosody

Emotional, or affective prosody, has been seen as primarily a property of the right hemisphere (Ross, 1981; Ross & Mesulam, 1979). However, a number of recent studies of aphasic patients have found both production and perception impairment of affective prosody (for a review see Ross & Monnot, 2008). A study of a Norwegian patient with a Broca type aphasia showed that he could not recognise different types of emotional prosody although his brain damage was limited to the left hemisphere. The patient was mildly agrammatic with word finding problems and what sounded like a foreign accent. Deviant prosody was an important feature of his foreign-sounding speech. Acoustic analysis of his speech revealed limited F0 variation at word and utterance level. The patient was first investigated in 2006 (Moen et al., 2007). He was tested in a follow up study with a Norwegian version of *Profiling Elements of Prosodic Systems* (Peppé & McCann, 2003). On the test items of the perception of emotional prosody he performed at chance level. He was unable to complete the test items of the production of emotional prosody. These studies bring into question the assumption of a right hemisphere superiority in the processing of affective prosody. Ross and Monot (2008), however, have offered an alternative explanation for why left hemisphere lesions might disrupt the processing of affective prosody. They argue that reduced affective-prosodic comprehension and production in patients with left hemisphere damage is a secondary result of a reduced ability to integrate the linguistic message, associated with the left hemisphere, with the prosodic-affective message, associated with the right hemisphere.

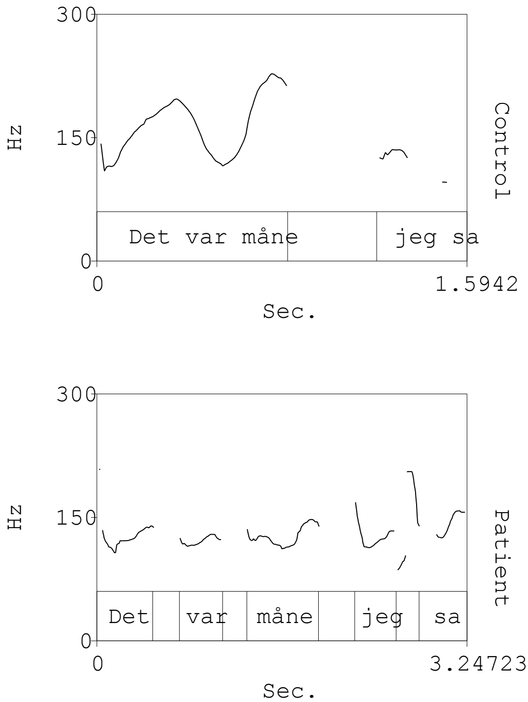


Figure 2. Fundamental frequency contours of the sentence *Det var "måne" jeg sa* 'It was "moon" I said', spoken by a control subject (upper contour) and an aphasic patient (lower contour).

4. Neuroimaging studies

In a series of functional magnetic resonance imaging studies Gandour and collaborators investigated the perception of tones, contrastive stress on words, and fundamental frequency at the sentence level in Mandarin Chinese, a tone language with four lexical tones (Gandour et al., 2003; Gandour et al., 2004; Tong et al., 2005).

In the investigation of the perception of tonal contrasts, the subjects were presented with pairs of words and asked to determine whether they were the same or different. The listeners, native speakers of Mandarin, showed activation in both hemispheres, but most in the left hemisphere.

In the intonation experiment, the stimuli were 36 pairs of three syllable utterances and 44 pairs of one syllable utterances. These were pronounced with two different intonation patterns. There was thus an interaction between the lexical tones and the intonation. The listeners were asked to judge whether pairs of utterances were the same or different. Again, the listeners showed effects in both hemispheres.

In a third experiment the perception of fundamental frequency at the sentence level combined with the perception of contrastive stress – a linguistic feature produced with a combination of fundamental frequency, intensity, and duration – was tested. The test material consisted of sentences with contrastive stress either in initial position or in final position. The sentences were produced with one of two intonation patterns – a declarative versus an interrogative modality. The subjects were asked to concentrate either on the stress pattern or on the intonation pattern when they were presented with pairs of sentences. They were asked to indicate whether the sentences differed or not. The listeners showed activity over widely distributed areas and greater left hemisphere activity in the stress task than in the intonation task. We see that these neuroimaging studies present a more complex picture than the clinical studies. There is clear evidence that both hemispheres are involved in the processing of prosody.

5. Conclusions and directions for future research

Investigations of the tonal production and perception in brain damaged speakers of tone languages have found reduced ability in both tasks in left-hemisphere-damaged aphasic patients as opposed to right-hemisphere-damaged nonaphasic patients where this ability is spared. This indicates that the processing of lexical tones is the property of the left hemisphere. Deviations in the production of stress and timing relations between syllables and words have also been found in the speech of left-hemisphere-damaged aphasic patients. It is, however, unclear to what extent the reduced, but not totally disrupted, performance of the left-hemisphere-damaged patients is due to residual mechanisms in the left hemisphere, to mechanisms in the right hemisphere, or to subcortical mechanisms.

Acoustic investigations of fundamental frequency at the level of the sentence have revealed both normal and abnormal characteristics in the speech of aphasic patients with left-hemisphere damage. On the basis of the current clinical evidence it is unclear to what extent the production of fundamental frequency above the level of the word is lateralized to one of the hemispheres.

A number of studies of aphasic patients have found both production and perception impairment of affective prosody. This brings into question the assumption of a right hemisphere superiority of affective prosody. It does, however, not rule out the possibility that the impairment of affective prosody in left-hemisphere-damaged patients is a secondary result of a reduced ability to integrate the affective information associated with the right hemisphere with the linguistic information associated with the left hemisphere, which leaves open the question of hemispheric specialization of affective prosody.

Recent neuroimaging studies of the perception of prosody in neurologically healthy individuals present a more complex picture than the clinical studies. The neuroimaging studies indicate that both hemispheres are involved in the processing of sentence prosody.

Although the nature of the mechanisms underlying reduced tonal perception and production are at present unclear, the current evidence indicates that tonal processing is impaired following left hemisphere damage in tone languages. There are, however, few investigations of tonal processing in patients with subcortical damage, such as patients with Parkinson's disease. Further investigations of this patient group might expand our knowledge of the neuroanatomical bases for the processing of tonal contrasts. Most studies of tones in the brain damaged population have focused on single words or short utterances in relatively few languages. Future research will, hopefully, include longer stretches of speech and data from a wider range of the numerous tone languages of the world. We also need more data on the interaction between morphological and syntactic information and prosodic information. In the study of prosody and the brain, both clinical data and neuroimaging studies are crucial since they provide supplementary evidence. Lesion studies tell us what areas are necessary for normal functioning, while imaging studies tell us what areas participate in that function.

References

Baum, S. (1990). Acoustic analysis of intra-word syllabic timing relations in anterior aphasia. *Journal of Neurolinguistics*, *5*, 321–332.

Baum, S. (1992). The influence of word length on syllable duration in aphasia. Acoustic analyses. *Aphasiology*, *6*, 501–513.

Baum, S. & Pell, M. M. (1997). Production of afftective and linguistic prosody by brain-damaged patients. *Aphasiology*, *11*, 177–198.

Cancelliere, A. & Kertesz, A. (1990). Lesion localisation in acquired deficits of emotional expression and comprehension. *Brain and Cognition*, *13*, 133–147.

Collins, M., Rosenbek, J. & Wertz, R. (1983). Spectrographic analysis of vowel and word duration in apraxia of speech. *Journal of Speech and Hearing Research*, *26*, 224–230.

Cooper, W. E. & Klouda, G. V. (1987). Intonation in aphasic and right-hemisphere-damaged patients. In J. H. Ryalls (Ed.), *Phonetic approaches to speech production in aphasia and related disorders* (pp. 59–77). Boston/Toronto/San Diego: Little, Brown and Company.

Cooper, W. E., Soares, C., Nicol, J., Michelow, D. & Goloskie, S. (1984). Clausal intonation after unilateral brain damage. *Language and Speech*, *27*, 17–24.

Danly, M. & Shapiro, B. E. (1982). Speech prosody in Broca's aphasia. *Brain and Language*, *16,* 171–190.

Gandour, J. (1987). Tone production in aphasia. In J. H. Ryalls (Ed.), *Phonetic approaches to speech production in aphasia and related disorders* (pp. 45–57). Boston/Toronto/San Diego: Little, Brown and Company.

Gandour, J. & Dardarananda, R. (1983). Identification of tonal contrasts in Thai aphasic patients. *Brain and Language*, *18*(1), 98–114.

Gandour, J., Dechongkit, S., Ponglorpisit, S. & Khunadorn, F. (1994). Speech timing at the sentence level in Thai after unilateral brain damage. *Brain and Language, 46*, 419–438.

Gandour, J., Petty, S. H. & Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language*, *35*, 201–240.

Gandour, J., Petty, S. H. & Dardarananda, R. (1989). Dysprosody in Broca's aphasia: A case study. *Brain and Language, 37*, 232–257.

Gandour, J., Ponglorpisit, S., Khunadorn, F., Dechongkit, S., Boongird, P., Boonklam, R. & Potisuk, S. (1992). Lexical tones in Thai after unilateral brain damage. *Brain and Language*, *43*, 275–307.

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X. & Tong, Y. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*, 185–200.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X. & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, *23*, 344–357.

Hughes, C. P., Chan, J. L. & Su, M. S. (1983). Aprosodia in Chinese patients with right cerebral hemisphere lesions. *Archives of Neurology*, *40*, 732–736.

Kent, R. D. & Rosenbek, J. (1982). Prosodic disturbance and neurologic lesion. *Brain and Language, 15*, 259–291.

Kent, R. D. & Rosenbek, J. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research, 26*, 231–249.

Klouda, G. V., Robin, D. A., Graff-Radford, N. R. & Cooper, W. E. (1988). The role of callosal connections in speech prosody. *Brain and Language*, *35*, 154–171.

Moen, I. (2004). Dysprosody in Broca's Aphasia: A study of Norwegian brain damaged patients. In *Teoreticeskie problemy jazykoznanija: Sbornik statej k 140-letiju kafedry obscego jazykoznanija Filologiceskogo fakul'teta Sankt-Peterburgskogo gosudarstvennogo universiteta*. St. Petersburg: Filologiceskij fakul'tet Sankt-Peterburgskogo gosudarstvennogo universiteta (pp. 341–356).

Moen, I., Becker, F., Günter, L. & Berntsen, M. (2007). An acoustic investigation of pitch accent contrasts in the speech of a Norwegian patient with the foreign accent syndrome. *Proceedings of 16th International Conference of Phonetic Sciences*, Saarbrücken, Germany.

Moen, I. & Sundet, K. (1996). Production and perception of word tones (pitch accents) in patients with left and right hemisphere damage. *Brain and Language*, *53*, 267–281.

Moen, I. & Sundet, K. (1999). An acoustic investigation of pitch accent contrasts in the speech of a Norwegian patient with a left-hemisphere lesion (Broca's aphasia). In B. Maassen & P. Groenen (Eds.), *Pathologies of speech and language. Advances in clinical phonetics and linguistics* (pp. 221–228). London: Whurr Publishers.

Monrad-Krohn, G. H. (1947). Dysprosody or altered melody of language. *Brain*, *70*, 405–423.

Naeser, M. A. & Chan, S. W.-C. (1980). Case study of a Chinese aphasic with the Boston Diagnostic Aphasia Exam. *Neuropsychologia*, *18*, 389–410.

Packard, J. L. (1986). Tone production deficits in nonfluent aphasic Chinese speech. *Brain and Language*, *29*, 212–223.

Peppé, S. & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, *17*, 345–354.

Ross, E. D. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology, 38*, 561–569.

Ross, E. D. & Mesulam, M.-M. (1979). Dominant language functions of the right hemisphere?: Prosody and emotional gesturing. *Archives of Neurology*, *36*, 144–148.

Ross, E. D. & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in the right hemisphere. *Brain and Language, 104*, 51–74.

Ryalls, J. H. (1982). Intonation in Broca's aphasia. *Neuropsychologia*, *20*, 355–360.

Ryalls, J. (1984). Some acoustic aspects of fundamental frequency of CVC utterances in aphasia. *Phonetica, 41*, 103–111.

Ryalls, J. & Reinvang, I. (1986). Functional lateralisation of linguistic tones: Acoustic evidence from Norwegian. *Language and Speech*, *29*, 389–398.

Seddoh, S. A. (2000). Basis of intonation disturbance in aphasia: Production. *Aphasiology, 14*, 1105–1126.

Shapiro, B. E. & Danly, M. (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain and Language*, *25,* 19–36.
Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X. & Lowe, M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *Neuroimage*, *28*, 417–428.
Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *International Journal of Human Communication*. *13*, 227–277.
Van Lancker, D. & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, *35*, 963–970.
Yiu, E. M.-L. & Fok, A. Y.-Y. (1995). Lexical tone disruption in Cantonese aphasic speakers. *Clinical Linguistics and Phonetics*, *9*(1), 79–92.

The role of a left-edge tone in speech processing

Mikael Roll
Department of Linguistics, Lund University

Abstract

Prosodic events associated with the right edge of intonation phrases influence their interpretation as well as the parsing of prosodic and syntactic phrase boundaries. Tonal events associated with the left edge, on the other hand, would appear to be more subtle and their functions less obvious. Nevertheless, the left edge of main clauses is often associated with prosodic prominence. The present study reviews work on Swedish speech production and processing arguing that there is a left-edge boundary tone associated with the beginning of main clauses in Standard Swedish. The left-edge boundary tone is used in speech processing to distinguish main clauses from subordinate clauses.

1. Introduction

1.1. Intonation and syntactic processing

Prosodic phrasing guides the syntactic parsing in language processing (e.g. Marslen-Wilson et al., 1992; Speer et al., 1996). Depending on the phrasing, the word string *When it rains cats and dogs stay home* may be interpreted in two different ways. For example, *rains* might be associated with a boundary tone, as in *When it rains* || *cats and dogs stay home*. Due to the preceding phrase boundary, indicated by (||), *cats and dogs* is then interpreted as the subject of the main clause verb *stay*. A boundary tone might also be associated with *dogs*, i.e. *When it rains cats and dogs* || *stay home*. *Cats and dogs* is then understood as a complement of *rains*, leaving *stay* without a subject, yielding an imperative interpretation.

If *stay* has another subject, e.g. *people*, the syntactic structure is only compatible with a boundary tone associated with *dogs*, as in *When it rains cats and dogs* || *people stay home*. The boundary tone cannot felicitously be associated with *rains*, i.e. **When it rains* || *cats and dogs people stay home*. If it were, *cats and dogs* would initially be interpreted as the subject of *stay*. For *people* to be integrated in the syntactic structure, *cats and dogs* would need to be reinterpreted as the complement of *rains*. Thus, the disambiguating word *people* would yield an intonation-induced "garden-path" effect, since due to the prosodic phrasing, it mismatches the syntactic structure that the listener has activated.

1.2. Right edge and reprocessing

Using Event-Related Potentials (ERPs), Steinhauer et al. (1999) found a "P600" effect at the syntactically disambiguating word in prosody-induced garden-path sentences in German. The P600 is a positive deflection peaking around 600 ms after stimuli indicating an anomalous syntactic or semantic structure. The increased positivity is assumed to reflect structural reprocessing (Osterhout & Holcomb, 1992; Hagoort et al., 1993; Bornkessel et al., 2002; Bornkessel & Schlesewsky, 2006). After the syntactically disambiguating word of easily resolved garden-path sentences, such as German object relatives, a positive peak termed "P345" sometimes precedes the P600 at 300–400 ms, forming a biphasic pattern (Friederici et al., 2001; Frisch et al., 2004). The first P345 phase of the ERP effect has been analyzed as indexing the detection of the syntactic anomaly, whereas the second P600 phase would show the actual reprocessing (Friederici et al., 2001).

1.3. Right edge and left edge

The difference in phrasing in the example sentence of Section 1.1 was mainly indicated by boundaries that show where the right edge of the first phrase is located. Generally, prosodic events associated with the right edge have a clearer role in the processing of phrases than left-edge prosody. Boundary tones, final lengthening, and pauses associated with the right edge of phrases influence the hearer's interpretation in terms of speech act and phrase boundary location. Left-edge prosody usually does not in fact change speech act interpretation. Furthermore, when a right-edge boundary tone indicates that a phrase ends, any following speech material is highly likely to belong to a subsequent phrase, making a left-edge intonational marking of the next phrase more superfluous for phrasing purposes. This distinction is reflected in the ERP effects following intonation changes at the right and left edge in German. Right-edge boundary tones are followed by a "Closure Positive Shift" (CPS) in the ERPs, coinciding with the perception of a phrase boundary (Steinhauer et al., 1999). Sentence-initial pitch accents, on the other hand, give rise to a "P200" effect, showing perception of the physical pitch change, but possibly no higher-order cognitive processing, such as semantic or syntactic integration (Heim & Alter, 2006).

Although prosodic marking of the left edge might seem superfluous, the left edge of intonation phrases is often associated with prosodic prominence. In English, the first word of utterances tends to be accented regardless of whether it expresses new or given information (Gussenhoven, 1985), due to metrical phrase structure constraints (Horne, 1991). Shattuck-Hufnagel (2000) analyzes English stress shift as an instance of early accent placement. Furthermore, speakers have been observed to associate the left edge of intonation phrases with boundary tones in a number of languages. In European Portuguese, for instance, Vigário (2003) and Frota (2003) report on an optional boundary tone with a primary association to the left edge of intonation phrases and a secondary association to the first prosodic word of phrases. Results such as these leave the question open as to whether left-edge intonation does not have more impact on the on-line syntactic processing than it might seem.

2. Left-edge tone in production

2.1. Previous studies

Investigating discourse structure in Swedish read speech, Horne (1994) observed a high (H) tone associated with the last syllable of the first prosodic word in an utterance. The H could not readily be interpreted as a focal accent, since it was associated with words expressing given information. The conclusion in Horne (1994) was that the H was a boundary tone associated with the right edge of the first prosodic word in an utterance. This rising tone is often very prominent and has been interpreted as related to speaker engagement in task-related dialogues (Horne et al., 2001).

2.2. Parsing function of the left-edge tone

In Horne (1994) and Horne et al. (2001), the early H tone was examined in relation to prosodic structure and speaker engagement. It was not assumed that the H tone was related to marking the left edge of utterances, since it did not occur at the very beginning of the utterance, but rather *followed* the first word accent. Therefore, the H tone was basically interpreted as a word boundary tone. No syntactic parsing function was assumed for the tone. Roll (2004, 2006) however, searching for a prosodic cues for disambiguating embedded main clauses and canonical subordinate clauses, found that speakers used the left-edge H tone to distinguish between the two syntactically different structures.

2.3. Word order and syntactic structure

In Swedish, main clause structure is characterized by the word order S-V-SAdv (Subject-Verb-Sentence Adverb), as in *Ölen rinner inte* '(lit.) the beer flows not', where the verb *rinner* 'flows' precedes the sentence adverb *inte* 'not'. Canonical subordinate clause structure, on the other hand, is indicated by the opposite order between the sentence adverb and the verb, i.e. S-SAdv-V, as in *Jag sa att ölen inte rinner* '(lit.) I said that the beer not flows'. In spoken language, embedded main clauses are often used instead of canonical subordinate clauses, as a means of expressing assertion. Thus, in the embedded main clause *Jag sa att ölen rinner inte* '(lit.) I said that the beer flows not', the verb *rinner* 'flows' precedes the sentence adverb *inte* 'not' as in an ordinary main clause, instead of following it, as it would in a canonical subordinate clause.

2.4. Main verb and embedded clause structure

Since embedded main clauses express assertion, they co-occur with assertive main clause verbs such as *säga* 'say', but not with nonassertive verbs such as *vilja* 'want'. The sentence **Jag vill att ölen rinner inte* '(lit.) I hope that the beer flows not' is not interpretable, since *vill* 'want' expresses uncertainty regarding the truth value of the embedded clause, whereas the embedded main clause *ölen rinner inte* '(lit.) the beer flows not' asserts its own truth value.

The resistance of nonassertive main clause verbs to be followed by embedded main clauses was used by Roll (2004, 2006) to construct unambiguous canonical subordinate clauses. Thus, the embedded clause in *Jag sa att* [*ölen rinner inte*] has unambiguous main clause structure due to the linear order between the verb *rinner* 'flows' and the sentence adverb *inte* 'not'. The corresponding *Jag vill att* [*ölen rinner lite*] '(lit.) I want that the beer flows a little' contains the VP adverb *lite* 'a little' instead of the sentence adverb *inte* 'not'. VP adverbs follow the verb both in main clauses and canonical subordinate clauses. Thus, the word order of the embedded clauses involving VP adverbs is structurally neutral. However, after a nonassertive verb, embedded clauses have unambiguous *subordinate clause structure* since nonassertive main clause verbs cannot be followed by embedded main clauses.

2.5. Left-edge tone and embedded clause structure

Roll's (2004, 2006) results showed that embedded main clauses, but not canonical subordinate clauses, are associated with a H tone at the left edge similar to what Horne (1994) and Horne et al. (2001) found in non-embedded main clauses. In Figure 1, there is a H tone in the last syllable of the first word *ölen* 'the beer' of the embedded main clause in *Jag sa att ölen rinner inte* '(lit.) I said that the beer flows not' (top). The tone is absent in the corresponding syllable in the canonical subordinate clause of *Jag vill att ölen rinner lite* '(lit.) I want that the beer flows a little' (bottom). Thus, the conclusion can be drawn that speakers used the left-edge H tone as a phrase-initial boundary tone to mark the beginning of a main clause.[1]

3. Left-edge tone and syntax in the brain

In order to find out whether listeners use the left-edge tone as an initial boundary tone, Roll et al. (2009) tested the online processing of embedded main clauses with and without left-edge boundary tones. The stimulus sentences were of the kind shown in [1–4]. All test stimuli consisted of a main clause, in the examples *Besökaren menar alltså* '(lit.) The visitor thinks thus' and an embedded main clause, *att familjen känner ju det* '(lit.) that the family feels of course that'. The sentence adverb *ju* 'of course' following the verb *känner* 'feels' shows the main clause structure of the embedded clause. Example [1] represents the condition AssH (assertive, H). It has the assertive main clause verb *menar* 'thinks' in the main clause, and a H left-edge boundary tone associated with the last syllable of the subject *familjen* 'the family' in the embedded main clause. In [2], the condition Ass∅ (assertive, no H), with no left-edge boundary tone in the embedded main clause, is shown. The intonation contour of [1] and [2] is seen in Figure 2. The conditions NassH (nonassertive, H) and Nass∅ (nonassertive, no H) are illustrated in [3] and [4]. These conditions had nonassertive main clause verbs instead of assertive, in this case *hoppas* 'hopes'. The subjects were asked to judge the word order as acceptable or unacceptable.

[1] *AssH*

Besökaren	menar	alltså	att	familjen H	känner	ju	det	på	kvällen
The.visitor	thinks	thus	that	the.family	feels	of.course	that	in	the.evening

[2] *Ass∅*

Besökaren	menar	alltså	att	familjen ∅	känner	ju	det	på	kvällen
The.visitor	thinks	thus	that	the.family	feels	of.course	that	in	the.evening

[1] The distribution of the left-edge H between the first two content words was also explored. The tone coincided with the first peak of the word accent associated with the second content word. Standard Swedish accent I is expressed as (H)L*, where the H is optional, and the L* is associated with the accented syllable. Accent II has the pattern H*L, i.e. the H is associated with the accented syllable. The difference between the presence and the absence of a left-edge H is therefore the presence or the absence of a first peak in the second word if it has accent I. If the second word has accent II, the left-edge H is also associated with its first peak, increasing the height of the obligatory H*.

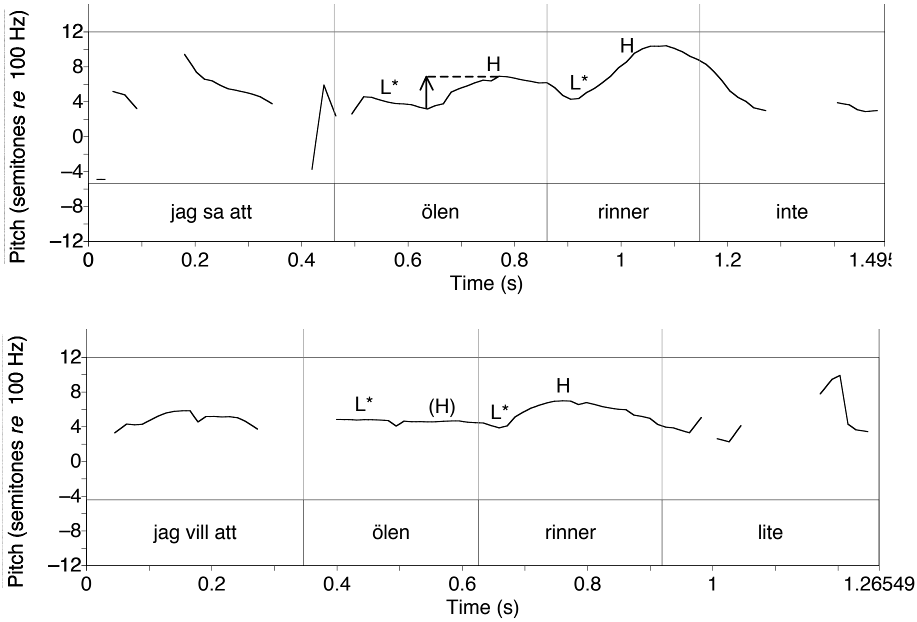


Figure 1. Intonation contour of the embedded main clause sentence *Jag sa att ölen rinner inte* '(lit.) I said that the beer flows not' (top), and the canonical subordinate clause sentence *Jag vill att ölen rinner lite* '(lit.) I want that the beer flows a little' (bottom). There is a H tone in the last syllable of the subject *ölen* 'the beer' in the embedded main clause, which is absent in the canonical subordinate clause.

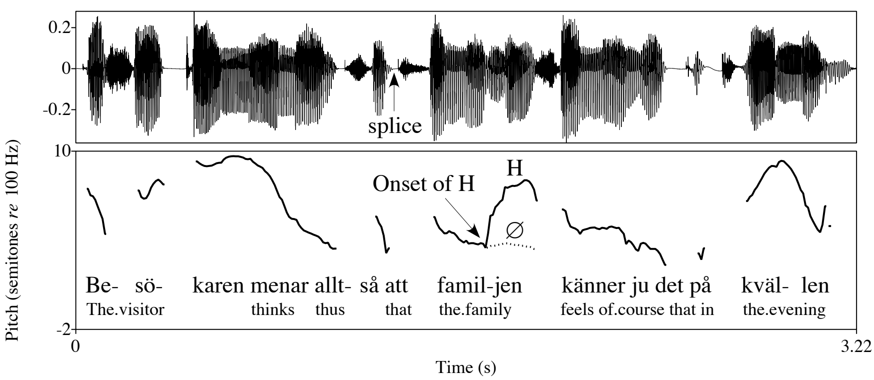


Figure 2. Waveform and intonation contour of the test sentences in [1] (AssH) and [2] (Ass∅). In [1], there is a left-edge boundary tone in the last syllable of the embedded clause subject (H), which is absent in [2] (∅). Originally published in Roll et al. (2009).

[3] *NassH*

*	Besökaren	hoppas	alltså	att	familjen H	känner	ju	det	på	kvällen
	The.visitor	hopes	thus	that	the.family	feels	of.course	that	in	the.evening

[4] *Nass∅*

*Besökaren	hoppas	alltså	att	familjen ∅	känner	ju	det	på	kvällen
The.visitor	hopes	thus	that	the.family	feels	of.course	that	in	the.evening

For each condition, 40 stimulus sentences were created by cross-splicing the main clauses with the subordinate clauses. Half of the sentences were produced with a H boundary tone in the last syllable of the embedded clause subject, and half without a H. The other half of the sentences were produced by lowering or raising the fundamental frequency of the original sentences using PSOLA resynthesis. The test sentences were balanced with 160 control sentences in order not to make the sentence adverb predictable, giving a total of 320 sentences.

3.1. ERP effects of the left-edge tone

The ERPs for AssH ([1]) and Ass∅ ([2]) are displayed in Figure 3. When the ERPs were timed to the left-edge boundary tone, the first effect was an anterior positive peak between 200 and 250 ms following the H onset. It was interpreted as a P200 effect, reflecting the perception of the physical pitch change, but no higher-order processing, similar to what Heim and Alter (2006) found for sentence-initial pitch accents in German.

The presence or absence of a left-edge boundary tone did have an effect, however, on the processing of the sentence adverb. In condition Ass∅ ([2]), where the sentence adverb had not been preceded by a H left-edge boundary tone, it was followed by an increased positivity, seen already in Figure 3. When the ERPs were timed to the sentence adverb (Figure 4), an increased biphasic positive effect between 300–400 ms and 600–800 ms emerged. This effect was probably a combined P345-P600 effect, encountered at the syntactically disambiguating point in easily resolved garden-path sentences. In other words, since there was no H boundary tone at the beginning of the subordinate clause in the Ass∅ condition, the listeners activated a canonical subordinate clause structure for the embedded clause. When the sentence adverb indicated embedded main clause structure, they had to reprocess the syntactic structure. The process was confirmed by the behavioral results. The word order was judged as correct in 68% of the sentences with a left-edge boundary tone in the embedded main clause, but only in 52% of the sentences lacking a left-edge tone.

3.2. ERP effects of nonassertive verbs

The nonassertive main clause verbs in condition NassH ([3]) led to a strong activation of subordinate clause structure for the embedded clause. Since the sentence adverb again contradicted the activated structure, the subjects had to reprocess the sentence structure, a process that was reflected in an increased P600 effect between 350 and 700 ms after the sentence adverb in the ERPs (Figure 5).

Figure 3. ERPs at the frontal electrode FZ timed to the onset of the H left-edge boundary tone (AssH) or the absence of a tone (Ass∅) after assertive verbs (*Besökaren menar alltså att familjen H/∅…* 'The visitor thus thinks that the family H/∅…'). A P200 is visible after the H. The sentence adverb *ju* 'of course' occurs at around 500 ms, and a following P345 is seen in the Ass∅ condition.

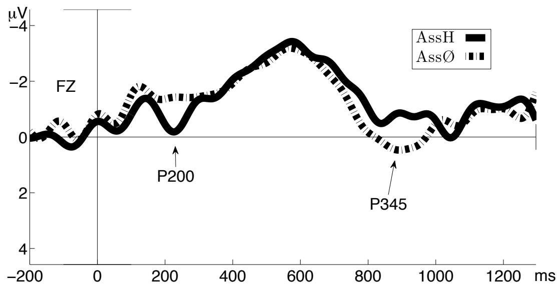


Figure 4. ERPs at the central electrode CZ timed to the sentence adverb of the embedded main clause (*Besökaren menar alltså att familjen H/∅ känner ju…* 'The visitor thus thinks that the family H/∅ feels of course…') a biphasic P345-P600 effect at 300–400 and 600–800 ms after the sentence adverb in the Ass∅ (no tone) condition was revealed.

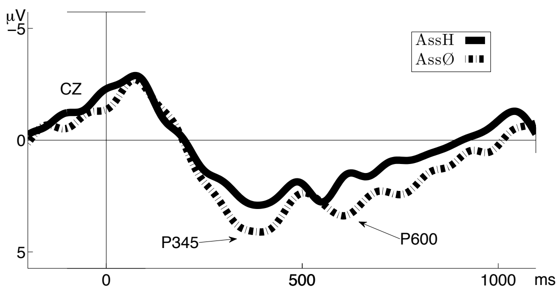
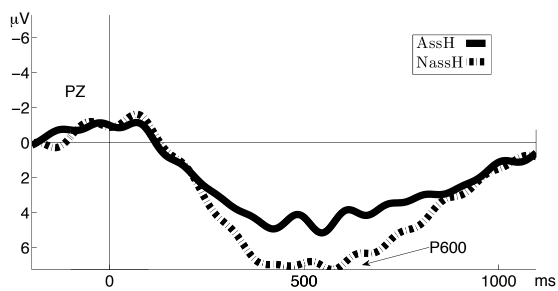


Figure 5. P600 effect after the sentence adverb *ju* 'of course' when following nonassertive verbs (NassH, dotted line), as in *Besökaren* *<u>hoppas</u>* *alltså att famljen känner ju…* 'The visitor thus <u>hopes</u> that the family feels of course…' as compared to assertive verbs (AssH, solid line), as in *Besökaren* *<u>menar</u>* *alltså att familjen känner ju* 'The visitor thus <u>thinks</u> that the family feels of course…'.



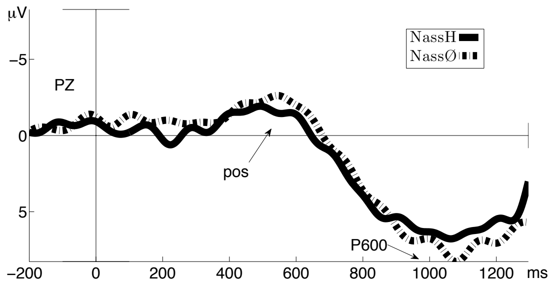


Figure 6. ERPs following a H boundary tone (NassH) or the absence of a boundary tone (Nass∅) after nonassertive verbs, *Besökaren hoppas alltså att familjen H/∅…* 'The visitor thus hopes that the family H/∅…'. There is an increased positivity between 450 and 700 ms after the H onset.

3.3. Assertiveness and left-edge tone

The difference between the biphasic P345-P600 effect of the absence of the boundary tone and the concentrated P600 effect of the nonassertive verbs most likely indicates that the verb pragmatics had a stronger effect on the syntactic structure preference than the left-edge boundary tone. This result was confirmed by the behavioral effect of the main clause verb pragmatics. With nonassertive main clause verbs, the acceptability of the word order of the embedded main clause dropped from 68% to 24%.

Although the verb pragmatics had a stronger effect, it did not cancel out the effect of the left-edge boundary tone. Even in the adverse conditions of nonassertive verbs, the left-edge boundary tone had a significant effect on the word order judgments. Without a left-edge H, the acceptability of the word order decreased from 24% to 19%. If listeners hear a nonassertive main clause verb, and thus strongly activate a canonical subordinate clause structure for the embedded clause, one might expect them to start reprocessing the embedded clause already when hearing the H left-edge boundary tone signaling an upcoming embedded main clause. As is seen in Figure 6, there was a hint of reprocessing activity in the ERPs after left-edge boundary tones following nonassertive verbs in condition NassH ([3]), as compared to the absence of a tone in condition Nass∅ ([4]). A weak increased posterior positivity was observed between 450 and 700 ms following the H onset.

4. Summary

Prosodic features associated with the right edge of intonation phrases are known to influence the parsing of syntactic structures. Reports on prosodic prominence associated with the left edge of phrases in English and the use of left-edge boundary tones in European Portuguese have opened the question as to whether left-edge prosody might also influence syntactic processing.

In previous studies on read speech, a H was observed in the last syllable of prosodic words at the left edge of utterances in Standard Swedish. These earlier studies examined the prosodic structure of main clauses. Roll (2004, 2006) however, by investigating embedded clauses, came to the conclusion that speakers used the left-edge tone to signal the beginning of main clauses as opposed to canonical subordinate clauses. In a later study, Roll et al. (2009) used ERPs to monitor the online processing of the left-edge tone in embedded main clauses. Listeners were found to enhance a main clause interpretation when hearing a H tone in the last syllable of the first prosodic word of a clause. In other words, there is evidence that a left-edge boundary tone is a cue to the syntactic parsing in spoken Swedish.

5. Acknowledgements

The study has been supported by grant 2004-2569 from the Swedish Research Council. The present version of this article has been substantially improved from the initial submission, thanks to constructive criticism from Merle Horne.

References

Bornkessel, I. & Schlesewsky, M. (2006). The extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113*(4), 787–821.

Bornkessel, I., Schlesewsky, M. & Friederici, A. D. (2002). Beyond syntax: Language-related positivities reflect the revision of hierarchies. *Neuroreport, 13*(3), 361–364.

Friederici, A. D., Steinhauer, K., Mecklinger, A. & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology, 47*, 193–221.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K. & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*, 305–323.

Frisch, S., Beim Graben, P. & Schlesewsky, M. (2004). Parallelizing grammatical functions: P600 and P345 reflect different cost of reanalysis. *International Journal of Bifurcation and Chaos, 14*(2), 531–549.

Frota, S. (2003). The phonological status of initial peaks in European Portuguese. *Catalan Journal of Linguistics, 2*, 133–152.

Gussenhoven, C. (1985). Two views of accent: A reply. *Journal of Linguistics, 21*, 125–138.

Hagoort, P., Brown, C. & Groothusen, J. (1993). The syntactic positive shift as an ERP-measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Heim, S. & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis, 66*, 55–68.

Horne, M. (1991). Why do speakers accent 'given' information? *Proceedings of Eurospeech 1991* (Vol. 3), 1279–1282.

Horne, M. (1994). Generating prosodic structure for synthesis of Swedish intonation. *Working Papers*, Dept. of Linguistics, Lund University 43 (pp. 72–75).

Horne, M., Hansson, P., Bruce, G., Frid, J. Accent-patterning on domain-related information in Swedish travel dialogues. *International Journal of Speech Technology, 4*(2), 93–102.

Marslen-Wilson, A. W. D., Tyler, L. K., Warren, P., Grenier, P. & Lee, C. S. Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology Section A, 45*(1), 73–87.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Roll, M. (2004). Prosodic cues to the syntactic structure of subordinate clauses in Swedish. M. A. Thesis, Lund University.

Roll, M. (2006). Prosodic cues to the syntactic structure of subordinate clauses in Swedish. In G. Bruce & M. Horne (Eds.), *Nordic prosody: Proceedings of the IXth conference, Lund 2004* (pp. 195–204). Frankfurt am Main: Peter Lang.

Roll, M., Horne, M. & Lindgren, M. (2009). Left-edge boundary tone and main clause verb effects on syntactic processing in embedded clauses – An ERP study. *Journal of Neurolinguistic, 22*(1), 55–73.

Speer, S. R., Kjelgaard, M. M. & Dobroth, K. M. The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research, 25*(2), 249–271.

Steinhauer, K., Alter, K. & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature, 2*(2), 191–196.

Vigário, M. (2003). Prosody and sentence disambiguation in European Portuguese. *Catalan Journal of Linguistics, 2*, 249–278.

Fishing for information: The interpretation of focus in dialogs

Ulrike Toepel

Neuropsychology and Neurorehabilitation Service, University Hospital and University of Lausanne

Ann Pannekamp, Elke van der Meer

Department of Cognitive Psychology, Humboldt University of Berlin, Germany

Abstract

Recognizing whether something heard during a conversation adds information to someone's state of knowledge happens quite automatically. We are not really aware of the linguistic cues we use to focus listeners' attention to pieces of relevant information we want to highlight and, in addition, how we are guided to recognize them. Using event-related potentials (ERPs), the current study investigates the online perception of spoken dialogs in the presence versus absence of adequate prosodic cues for focus identification. We found that the processing of information foci is accompanied by focus-related positive-going brain markers irrespective of their prosodic adequacy. In addition, missing prosodic markings on information foci elicited negative-going brain potentials.

1. Introduction

Conversations and other information exchanges between humans usually do not consist of single sentences but often take the form of dialogs. The structure of these verbal exchanges is not arbitrary but follows specific organization principles projecting beyond the borders of single sentences. Halliday (1967) introduced the term "information structure" to subsume the organizing principles of connected utterances. In a simplistic way, information units can be divided into parts comprising of novel or contextually non-derivable information (e.g. contrastive statements) as opposed to previously encountered assertions. Novelties or contrasts within a verbal exchange can also be called "information focus". On the other hand, information that has previously been mentioned in the discourse or can be inferred from it is referred to as "non-focus" or "given" information.

Within a dialog between conversation partners, the proportion and content of focused and non-focused information are subject to constant dynamics. When new information is introduced into the discourse, it is focused. Subsequently, it turns into a common ground, non-focused part of the message when it is spoken about again (Grosz & Sidner, 1986).

German, the language investigated here, has several linguistic strategies for rendering focused information particularly salient (see Toepel (2006) for an overview). First, these can be related to syntactic parameters like word order changes: *It was the CLOWN who amused the children (and not the pony)*. Second, semantic-pragmatic contexts such as questions can induce the focussing of a sentence element: *Who amused the children? The CLOWN did*. Third, in spoken conversation, elements to be focused can be overtly highlighted by sentence prosody and focus accentuation *The CLOWN amused the children (and not the pony)*.

The study at hand investigates the interplay of semantic-pragmatic and prosodic information in discourse interpretation in adults. It is part of a long-term project that is concerned with the developmental course of discourse interpretation abilities in children at various ages. In particular, the project aims at identifying local and global brain markers for the perception of focused dialog information. Moreover, the project is designed to explore the influence of prosody on focus identification in younger and older children as well as in the adult control group presented here.

Event-related potentials (ERPs) constitute a valuable tool to investigate online discourse processing in adults and children effectively and without health risks. With respect to the goals of our project, ERPs can be used to determine the importance of semantic-pragmatic versus prosodic cues for dialog interpretation. That is, brain markers related to the perception of matching and non-matching associations between semantic-pragmatic (contextually induced) focus and the (prosodic) focus accentuation can be directly compared. In addition, we complement local ERP analyses by topographical global measures of the electric field on the scalp surface over time (Murray et al., 2008). By doing so, comparisons with previous ERP studies on focus perception are possible but also refinements in terms of topographic modulations between process-

Table 1. Examples of the dialog materials with literal translations.

	Novelty focus condition	Correction focus condition
Focus-inducing context question	Wen hat Thomas gefragt? Who did Thomas ask?	Hat Thomas Anne gefragt? Did Thomas ask Anne?
Prosodically congruent answer	Thomas hat LISA gefragt. Thomas did LISA ask.	Thomas hat LISA gefragt. Thomas did LISA ask.
Prosodically incongruent answer	*Thomas hat Lisa gefragt. *Thomas did Lisa ask.	*Thomas hat Lisa gefragt. *Thomas did Lisa ask.

ing conditions. These latter would in turn be indicative of a differential engagement of neural generators for either focus/accentuation type (Lehmann, 1987).

Previous ERP studies in adults evince several effects induced by the perception of focus structures in the visual and auditory domain. For visual sentence perception, Bornkessel and colleagues (2003) reported a positive-going waveform between 280 and 480 ms at centro-parietal electrodes when a novel sentence element was focused by syntactic means (word order change). A similar ERP component elicited by contrastive focus elements was reported by Cowles et al. (2007). Moreover, Cowles and colleagues described an additional negative-going centro-parietal ERP effect peaking around 400 ms when the focus position in a sentence included an inappropriate focus referent. The pattern was interpreted as an N400-like effect likely reflecting a hampered integration of the sentence elements into the discourse structure.

In studies on the auditory perception of dialogs, a focus-related positive shift has been shown by Hruska and colleagues (2004) when listeners encountered a semantic-pragmatically induced novelty focus which was adequately marked by accentuation. The rather sustained ERP effect started ~400 ms after the focus onset and was present at centro-parietal electrodes. A similar finding for the perception of novelty focus was reported by Toepel et al. (2004) who, in addition, evinced the independence of the focus-related positive shift from the actual prosodic realization of a semantic-pragmatically focused element. The focus-related positive shift could also be replicated for the perception of contrastive information, and was again found to be independent from the actual prosodic marking of a contextually induced focus (Toepel et al., 2007). However, when a pragmatic focus was not marked by an adequate accentuation pattern, an additional N400 effect preceded the focus-related positive shift.

With respect to the low-ranging age groups of our long-term project, the study design is directed towards the *implicit* processing of focus information. That is, the participants were not instructed about the varying focus and accentuation types, and were furthermore only asked simple comprehension questions after random numbers of dialogs. Within this child-friendly design, the ERPs for the processing of novelty as well as contrastive focus (i.e. corrections) were investigated. Moreover, the perception of both focus types was explored under conditions where the contextually induced pragmatic focus is adequately realized in terms of prosody or not (see Table 1).

2. Methods and procedures

2.1. Participants

Thirty-one university students took part in the study (15 male, 16 female; mean age: 23.8 years; age range: 19–32 years) and were paid for participation. All volunteers were right-handed (Oldfield, 1971) and without any known neurological or hearing disorders.

2.2. Dialog materials

The four dialog conditions were formed by a context question and an answer (see Table 1). The wh-pronoun in the novelty focus type question prompted the introduction of a new discourse referent in the consecutive answer sentence. On the other hand, the contrastive question introduced a discourse referent which is corrected in the successive answer. The two question types were combined with their prosodically appropriate answers carrying a focus accent on the critical noun. Moreover, both question types were also combined with a prosodically inappropriate answer bearing no focus accent. The inappropriate answers were identical for the two question types they were combined with. Thus, during the experiment two prosodically correct and two prosodically incorrect dialog conditions were presented.

The recording of the dialogs was done by two trained female speakers of Standard German. One speaker was provided with a list of all questions and one speaker was in charge of producing the answers in a collaborative way. For the realization of appropriate question-answer pairs, the questioner asked for novel information (wh-pronoun) or uttered a potentially contrastive query. The respondent in turn produced the respective prosodically appropriate answers (novelty or correction focus accentuation). For the realization of the inappropriate answer type, the questioner uttered the potentially contrastive query that was this time responded to by an answer repeating all information already given by the question. Thus, the answers only conveyed given information realized without focus accentuation. Forty dialogs per type were produced by the speakers, that is, 160 dialogs all in all.

Speech recordings were made in a sound-attenuated booth at a sampling rate of 44.1 kHz (16bit, mono). Each sentence was saved in an individual file; loudness was consecutively adapted. Analyses of the durational and fundamental frequency (F0) were carried out with the PRAAT software (`http://www.praat.org`) for the different types of answer sentence prosody.

The overall sentence duration for the novelty accentuation was significantly shorter than for the inadequate prosody (1551 ms [*SD* = 96.7] vs. 1665 ms [*SD* = 114]; $t(78) = -4.16$, $p < .01$). The average sentence duration for corrections was also significantly shorter than the duration of answers bearing an inadequate prosody (1572 ms [*SD* = 96.9] vs. 1665 ms [*SD* = 114]; $t(78) = -3.07$, $p < .01$). In the position of the focus element, the noun was significantly longer when bearing a novelty accentuation as opposed to an inadequate prosody (408 ms [*SD* = 50.0] vs. 362 ms [*SD* = 49.2];

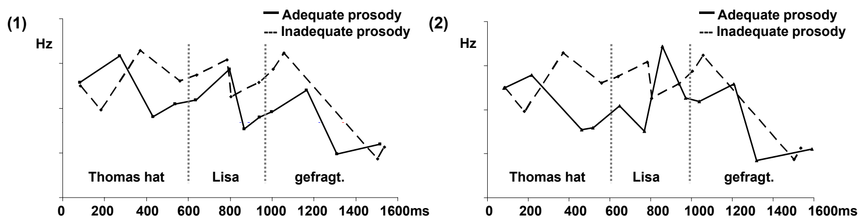


Figure 1. F0 contours over answer sentences with (left) novelty accentuation (solid line) versus inadequate prosody (dashed line); and answer sentences with (right) correction accentuation (solid line) versus inadequate prosody (dashed line).

$t(78) = 4.21$, $p < .01$). The noun carrying the correction focus accent was also significantly longer than when produced with the inadequate prosody (426 ms [$SD = 58.8$] vs. 362 ms [$SD = 49.2$]; $t(78) = 5.33$, $p < .01$). Thus, both focus accents have longer durations than the non-focused nouns.

Figure 1 shows the F0 contours after averaging the onset, minimal, maximal, and offset F0 values over sentence constituents in the 40 sentences per condition (left: novelty vs. inadequate prosody, right: correction vs. inadequate prosody). The accent on the novelty focus (solid line in left panel) is realized with a rising F0 contour while the correction accentuation (solid line in right panel) is associated with a falling-rising pattern with a pronounced rise on the focused noun. The prosody of inadequate information on the noun (dashed line) is also realized with a fall-rise in the F0 contour. The F0 excursions over the noun with novelty focus accentuation are significantly higher than with inadequate prosody (66.63Hz [$SD = 29.37$] vs. 41.3Hz [29.97]; $t(39) = 3.75$, $p < .05$). Also, the tonal movement over the noun with correction accentuation is more pronounced (96.88Hz [$SD = 47.99$] vs. 41.3Hz [29.97]; $t(39) = 6.01$, $p < .05$).

2.3. Experimental procedure

The dialogs were presented via loudspeakers in a sound- and electromagnetically shielded chamber. Participants were seated in front of a PC monitor. Each dialog trial started with the presentation of a question. After an inter-stimulus interval of 2000 ms, the answer to the question followed. During the auditory presentation of each dialog, a crosshair was present on the screen. Volunteers were instructed to fixate the crosshair to avoid ocular movements. After the presentation of each dialog, a blink phase of 3000 ms followed that was marked by the presentation of a smiley on the screen. After a random number of trials, a question mark appeared indicating that the participant would now have to answer a comprehension question.

An experimental session typically lasted approximately 30min., no longer than 90min. including electrode preparation.

2.4. EEG recordings and analyses

The EEG was recorded from 59 Ag/AgCl cap-mounted electrodes whose positions were in accordance with the international 10–20 system (Jasper, 1958). The ground electrode was positioned above the sternum. The bipolar vertical electrooculogram (VEOG) was recorded from electrodes placed above and below the right eye. The bipolar horizontal electrooculogram (HEOG) was recorded from positions at the outer canthus of each eye. Electrode impedances were kept below 5 kΩ. The EEG was acquired with XREFA amplifiers at a sampling frequency of 500 Hz. Recordings were online referenced to the left mastoid and offline re-referenced to the average reference of all cap electrodes.

General analysis strategy

The main objective of this study was to examine the online brain mechanisms during the perception of focused information in the presence and absence of adequate prosodic markings. Event-related brain responses were analyzed by means of local and global measures of the electric field at the scalp. These analyses can serve to differentiate effects following from modulations in the strength of electric responses of statistically indistinguishable brain generators from topographic modulations in the electric fields. Changes in the topography of the electric field on the scalp are, in turn, indicative of alterations in the configuration of intracranial generators. Details of the analysis formulae have been given elsewhere (Murray et al., 2008; Michel et al., 2004). All EEG analyses were carried out with the Cartool Software (`http://brainmapping.unige.ch/Cartool.htm`).

Initial data treatment

After data recording, EEG epochs were semi-automatically scanned for eye and muscle artifacts and other noise transients which were rejected from further data analysis. A bandpass filter from 0.1–40 Hz was applied to each participant's data set. Only participants with more than 75% of artifact-free trials per condition were included in further analyses. The EEG data were averaged per participant and condition between –100 to 1000 ms relative to the onset of the focused noun. A baseline correction was applied to the period of 100 ms before noun onset. In a second step, grand averages were computed across subjects.

EEG waveform modulations

As a first level of analysis, we analyzed the ERP waveforms of each electrode time-point-wise in a series of pair-wise comparisons (*t*-tests). Two conditional comparisons were performed, i.e. the perception of novelty focus with versus without focus accentuation, and the processing of correction focus in the absence versus presence of prosodic focus marking. Temporal auto-correlation at individual electrodes was controlled for by a temporal threshold of 20ms. Further, a statistical threshold of 95% confidence was applied. Thus, conditional differences are only considered significant when lasting longer than 20 ms and bearing p-values < .05. The results of the initial time-point-wise *t*-tests across all electrodes served to define time-windows for the ANOVAs on ERP components in an observer-independent data-driven manner.

In order to allow comparisons with previous ERP studies on focus perception by means of ANOVAs, several regions of interest (ROI) on the scalp were formed by averaging EEG data from spatially adjacent electrodes. At the midline electrodes, three ROIs were defined (midline frontal: FPZ, AFZ, FZ; midline central: FCZ, CZ, CPZ; midline parietal: PZ, POZ, OZ). For the lateral electrodes, six ROIs were formed (frontal left: FP1, AF7, F9, AF3, F7, F5, F3, FT7; frontal right: FP2, AF8, F10, AF4, F8, F6, F4, FT10; central left: FC3, FC5, FT7, C3, C5, T7, CP3, CP5, T; central right: FC4, FC6, FT8, C4, C6, T8, CP4, CP6, TP8; parietal left: P3, P5, P7, P9, TP9, PO3, PO7, O1; parietal right: P4, P6, P8, P10, PO4, PO8, O2).

Global electric field analyses

Modulations in the strength of the electric field at the scalp were assessed by means of Global Field Power (GFP; Lehmann & Skrandies, 1980). For this purpose, the data of each participant were compared with a time-point-wise paired *t*-test separately for answer sentences conveying adequately accented novelty and correction foci versus the inadequate prosodic realization. GFP is calculated as the square root of the mean of the squared value recorded at each electrode (versus the average reference). It represents the spatial standard deviation of the electric field at the scalp and yields larger values for stronger electric fields. Modulations in the GFP or response strength of the electric field, respectively, are orthogonal to changes in the response topography described below. Thus, power and topography changes can co-occur but can also be present in varying time periods. The observation of a GFP modulation in the absence of a topographic modulation would most parsimoniously be interpreted as the modulation of statistically indistinguishable neural generators across experimental conditions. In contrast, topographical modulations point to differences in the neural generators between conditions, and provide a global means to define ERP components.

The aspect of topographical changes is addressed by the analysis of Global Dissimilarity (Lehmann & Skrandies, 1980), a direct index of condition-wise configuration differences between electric fields independent of their strength (i.e., power-normalized data are compared). Statistical differences in the Global Dissimilarity between two conditions are identified by a Monte Carlo nonparametric bootstrapping procedure (Murray et al., 2008). Neurophysiologically, topographic changes between conditions in certain time periods indicate differences in the brain's underlying active generators (Lehmann, 1987). Yet, periods of Global Dissimilarity differences do not allow inferring the number and type of different electric field configurations. For this purpose, the ERP data for the time intervals revealing topographic changes between conditions were submitted to a topographic pattern (i.e. electric field map) analysis based on a hierarchical clustering algorithm (Murray et al., 2008). This method is independent of the reference electrode (like the GFP analysis) and is insensitive to pure amplitude modulations across conditions (topographies of power-normalized maps are compared). The output of the clustering algorithm is an optimal number of maps (i.e., the minimal number of maps that accounts for the greatest variance of the dataset) as determined by a modified Krzanowski-Lai criterion (Murray et al., 2008). The pattern of map occurrence observed in the group-averaged data is statistically tested by comparing each of these maps with the moment-by-moment scalp topography of an individual participant's ERPs for each condition. This procedure is referred to as "map fitting". For the map fitting, each time point of each ERP from each participant is labeled according to the electric field map with which it best correlates spatially (Murray et al., 2008). The result of the fitting is a table of relative map presence in milliseconds. It provides the amount of time over a given interval that each map as identified by the clustering of the group-averaged data best accounts for the response from an individual participant and condition. These values are then submitted to a repeated measure ANOVA using condition and map as within-subject factors. When the ANOVA reveals an interaction condition x map, and map A is present significantly longer in one condition and map B in another condition, it can be inferred that brain responses over the given time period are better explained by different map topographies.

All global reference-free analyses were computed with a temporal threshold criterion of 20 ms effect length and $p < .05$ as statistical threshold for effect size.

3. Results

In what follows, we first present the ERP deflections and results of the global analyses for the processing of novelty focus with adequate versus inadequate prosodic marking. In succession, the electrophysiological data for the perception of correction focus with appropriate versus inadequate prosodic markings are reported.

3.1. Brain markers for the perception of novelty focus

In Figure 2a, the processing of novelties with adequate accentuation on the critical noun and with inadequate prosody is illustrated by ERP waveform displays and a statistical scatter plot displaying the time-point-wise ERP differences between conditions.

The condition with adequate accentuation (blue line) shows a long-lasting frontal negativity. Furthermore, a positive shift starts at approximately 400 ms after the focus onset at posterior electrodes. The condition with inadequate prosody (red line) reveals a central-posterior negativity peaking around 400 ms. The negativity is followed by a positive shift most pronounced at posterior electrodes.

Initial time-point-wise paired *t*-tests across all electrodes revealed ERP modulations over the 180–320 ms, the 350–420 ms and the 600–800 ms interval after noun onset. Based on these initial indications on condition-wise ERP waveform differences, we conducted ANOVAs over the defined ROIs. There is an interaction between condition and region visible in all time periods at lateral (180–320 ms: $F(2, 60) = 21.78$, $p < .01$; 350–420 ms: $F(2, 60) = 9.84$, $p < .01$; 600–800 ms: $F(2, 60) = 6.26$, $p < .01$) and at midline electrodes (180–320 ms: $F(2, 60) = 14.27$, $p < .01$; 350–420 ms: $F(2, 60) = 5.98$, $p < .05$; 600–800 ms: $F(2, 60) = 4.48$, $p < .05$). Additionally, a condition x hemisphere interaction was found at lateral electrodes in the time-period between 350 and 420 ms ($F(1, 30) = 5.67$, $p < .05$). To rule out interaction effects between condition and region, new ROIs were conducted for frontal (frontal left, midline frontal and frontal right), central (central left, midline central and central right), and parietal (parietal left, midline parietal and parietal right) electrodes. A main effect condition was present at frontal electrodes in all time-periods (180–320 ms: $F(1, 30) = 24.90$, $p < .01$; 350–420 ms: $F(1, 30) = 8.48$, $p < .01$; 600–800 ms: $F(1, 30) = 8.45$, $p < .01$), and also at parietal electrodes (180–320 ms: $F(1, 30) = 19.83$, $p < .01$; 350–420 ms: $F(1, 30) = 9.31$, $p < .01$; 600–800 ms: $F(1, 30) = 5.24$, $p < .05$). These frontal and parietal effects of condition that were not existent at central electrodes were already visible in the paired *t*-tests (see Figure 2a).

Changes in the response strengths of the electric field between the conditions as measured by GFP were apparent between 392 and 420 ms (Figure 2b). Modulations of the response topography as indicated by Global Dissimilarity were yielded over the 180–304 ms, the 344–460 ms, and the 685–720 ms intervals (Figure 2c).

The topographic pattern analysis (over time periods yielding topographic dissimilarities between the condition with adequate and inadequate prosodies) and the results of the map fitting are displayed in Figure 2d. Below the bar graphs of each fitting period, the maps that best explain the group-wise brain responses during the respective interval are shown. That is, the maps displayed below the bar graphs were present significantly longer in one processing condition than in the other. Significant condition x map interactions were revealed over the 180–304 ms interval (3 maps: $F(2, 60) = 14.81$, $p < .01$), the 344–460 ms interval (4 maps: $F(3, 90) = 9.74$, $p < .01$), and the 685–720 ms period (2 maps: $F(1, 30) = 19.48$, $p < .01$). Thus, over all three time intervals that revealed topographic modulations, differential electric field maps accounted for the collective ERP responses to the perception of novelty focus.

3.2. Brain markers for the processing of correction focus

In Figure 3a, the processing of corrections with adequate accentuation on the critical noun and with inadequate prosody is illustrated by ERP waveform displays and a statistical scatter plot displaying the time-point-wise ERP differences between conditions.

The condition with adequate prosody (green line) elicits a centro-posterior positive ERP starting around 400 ms after the onset of the focused and accented noun. In contrast, the condition realized with an inadequate prosody on the noun (red line) shows a centro-posterior negativity peaking around 400 ms. The negative ERP peak is followed by a positive shift.

Initial time-point-wise paired *t*-tests across all electrodes revealed ERP modulations over the 370–500 ms, the 550–620 ms, and the 700–820 ms intervals after noun onset. Based on these initial indications on condition-wise ERP waveform differences, we conducted the ANOVA for the ROIs. A main effect of condition was existent in both time periods at lateral (370–500 ms: $F(1, 30) = 34.60$, $p < .01$; 550–620 ms: $F(1, 30) = 10.08$, $p < .01$; 700–820 ms: $F(1, 30) = 12.85$, $p < .01$) and midline electrodes (370–500 ms: $F(1, 30) = 36.20$, $p < .01$; 550–620 ms: $F(1, 30) = 10.74$, $p < .01$; 700–820 ms: $F(1, 30) = 15.24$, $p < .01$). Additionally, a condition x region interaction was present in the time period between 550 and 620 ms at lateral electrodes ($F(2, 60) = 5.47$, $p < .05$). The dissolution of the interaction revealed a main effect

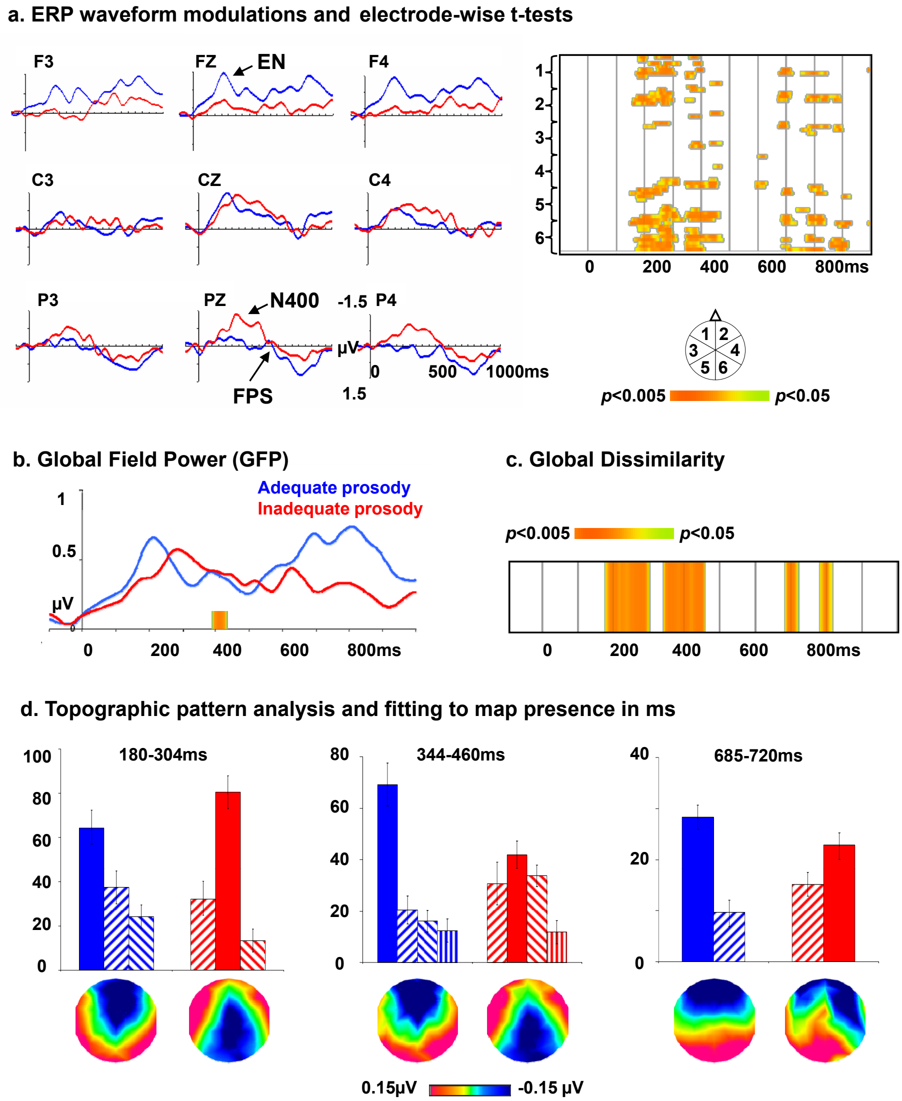


Figure 2. Perception of novelty focus on the noun, with adequate accentuation (blue line) and inadequate prosody (red line). (a) ERP waveforms from nine representative electrodes (low-pass filtered with 7Hz). (b) GFP modulations and (c) Global Dissimilarity index over the post-stimulus period. (d) Results of the topographic pattern analysis and relative map presence (in ms) over the three stable periods of brain responses as revealed by the fitting procedure. Maps below the bar graphs show the electric field topographies that best explain the group-wise brain responses to the prosodically adequate condition (below the blue bars) and to the prosodically inadequate condition (below the red bars).

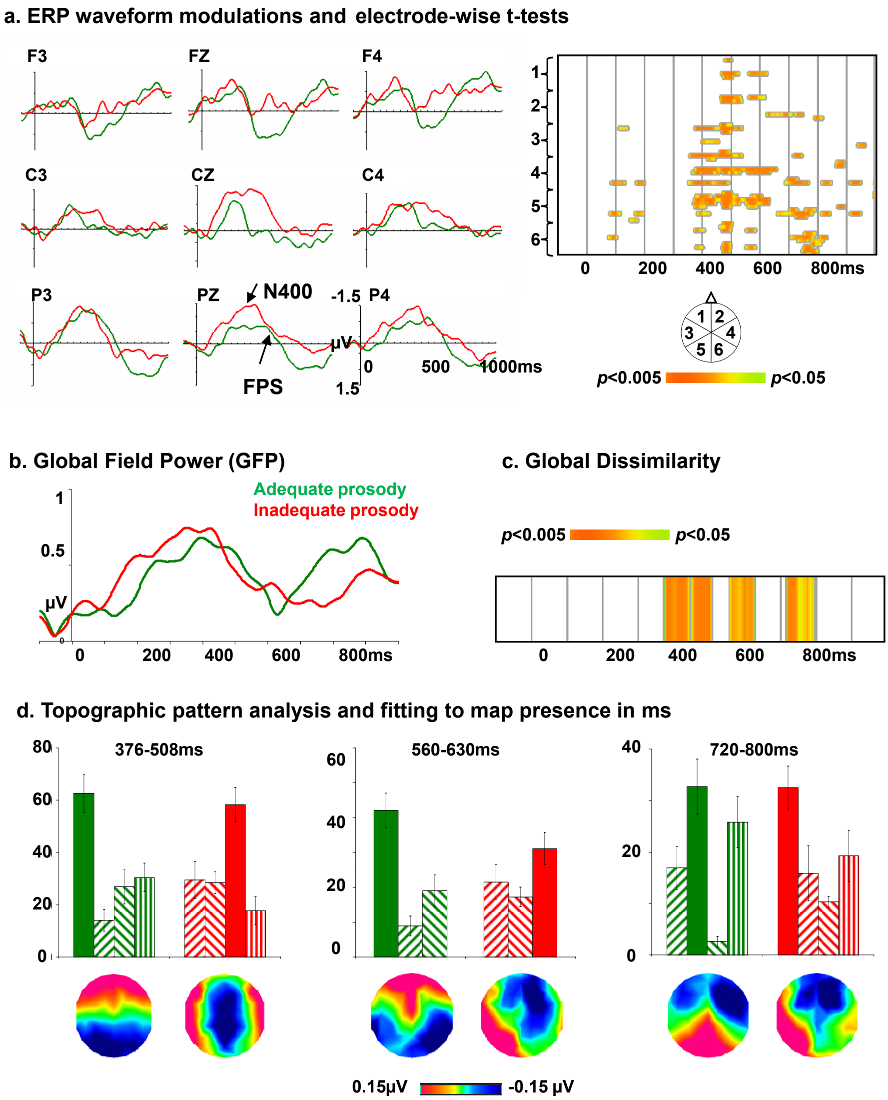


Figure 3. Perception of correction focus on the noun, with adequate accentuation (green line) and inadequate prosody (red line). (a) ERP waveforms from nine representative electrodes (low-pass filtered with 7Hz). (b) GFP modulations and (c) Global Dissimilarity index over the post-stimulus period. (d) Results of the topographic pattern analysis and relative map presence (in ms) over the three stable periods of brain responses as revealed by the fitting procedure. Maps below the bar graphs show the electric field topographies that best explain the group-wise brain responses to the prosodically adequate condition (below the green bars) and to the prosodically inadequate condition (below the red bars).

condition at frontal ($F(1, 30) = 4.96$, $p < .05$) and parietal electrodes ($F(1, 30) = 5.04$, $p < .05$). Also, a condition x region effect was present between 700 and 820 ms (lateral: $F(2, 60) = 6.01$, $p < .05$; midline: $F(2, 60) = 5.17$, $p < .01$). This interaction manifested itself in a main effect condition at parietal electrodes ($F(1, 30) = 11.78$, $p < .01$).

No modulations in the response strength between the conditions as measured by GFP were observed (Figure 3b). Modulations of the response topography as indicated by Global Dissimilarity were yielded over the 376–508 ms, the 560–630 ms, and the 720–800 ms intervals (Figure 3c).

The topographic pattern analysis (over time-periods yielding topographic dissimilarities between the condition with adequate and inadequate prosodies) and results of the map fitting are displayed in Figure 3d. Below the bar graphs of each fitting period, the maps that explain best the group-wise brain responses during the respective interval are shown. Thus, the maps displayed below the bar graphs were present significantly longer in one processing condition than in the other. Significant condition x map interactions were revealed over the 376–508 ms interval (4 maps: $F(3, 90) = 12.16$, $p < .01$), the 560–630 ms period (3 maps: $F(2, 60) = 8.36$, $p < .01$), and the 720–800 ms interval (4 maps: $F(3, 90) = 5.55$, $p < .01$). Thus, over all three time intervals that revealed topographic modulations, differential electric field maps accounted for the collective ERP responses to the perception of correction focus.

4. Discussion

The work presented in this contribution aimed at identifying the implicit perception of focus information in the absence versus presence of prosodic highlighting. Local and global measures of the electric field on the scalp surface were utilized to investigate its online consequences.

When processing target sentences containing adequately realized novelties, ERP waveforms showed a long-lasting negativity over frontal electrodes that started directly after the onset of the focus position (Figure 2a). Similar findings were reported by Heim et al. (2006) and Hruska et al. (2004) whenever a particular information focus was expected due to contextual requirements and also prosodically realized. In our study, we could also replicate the previous finding with global topographic analyses. As apparent from the left-most electric field topographies in Figure 2d, the map with a frontal negative maximum best explained listener's responses to the expected and adequately realized novelty focus in the earliest time period. The particular effect has been termed Expectancy Negativity (EN). We propose that the EN in our study is due to the presence of the wh-word in the context question that in turn leads to the listeners' expectation of a prosodically marked novelty focus in the target sentence. Yet, it is only when this expectation is also fulfilled in terms of accentuation that the EN is induced.

While the novelty focus question obligatory needs a sentence element bearing novelty information to be felicitously answered, the correction context is ambiguous. That is, the correction question could be followed by the repetition of the questioned information as well as by corrective information. As apparent from the ERP data, there was no EN effect present at the earliest time period when correction foci were perceived. Hence, the ambiguous correction context seemed to decrease listeners' expectations on the (prosodic) form of the target sentence.

When processing the inadequate accentuation of novelty focus, the ERP analyses evinced an additional posterior effect in the earliest time window. This centro-posterior negative maximum can also be seen in the left-most map topographies in Figure 2d. We interpret the response pattern as the early onset of an auditory N400, in particular since a similar topographic map was still predominant in the successively analyzed time-period, that is, during the "classical" N400 window (Diaz & Swaab, 2007; Van den Brink & Hagoort, 2004; Van Petten et al., 1999). The respective topographic N400 map related to perceiving the inadequate novelty accentuation is visualized in the middle part of Figure 2d.

When perceiving inadequate correction intonation, participants' brain responses did not differ topographically earlier than in the N400 window (see left-most map topographies in Figure 3d). Moreover, the context-to-prosody mismatch effect started slightly later than for inadequate novelty accentuation. In the ERP waveforms, the mismatch was reflected in a centro-posterior negativity (Figure 3a), seen also for the perception of the inadequate novelty prosody. We propose that the later onset of the mismatch effect for correction prosody is due to the ambiguous correction context resulting in less strong expectations on the prosodic realization of the correction than on the novelty focus as elaborated above.

Also, when comparing the topographic maps for both prosodically inadequate conditions in approximate time windows (core maps in Figure 2d versus left-most maps in Figure 3d) differences in the N400 topographies become apparent. Whether these varying scalp topographies were indeed due to differences in the underlying neural generators between conditions is something that will be explored later during the course of the project.

To summarize the above findings, electrophysiological consequences of perceiving prosodically inadequate focus markings (novelties or corrections) are reflected in centro-posterior negativities relative to the onset of the focus element. We interpret these negativities as N400 patterns induced by the expected but not realized focus prosody, thus as online brain markers related to information structural violations. Similar N400 effects related to the processing of information structural inconsistencies have been reported, albeit mostly in the visual domain (Cowles et al., 2007; Hagoort & Van Berkum, 2007; Nieuwland & Van Berkum, 2006; Bornkessel et al., 2003). In these studies, the N400 effects were mainly caused by violations of general world knowledge and discourse reference. However, in the auditory domain, the N400 has been consistently reported for information structural violations induced by inadequate focus accentuations (Toepel et al., 2007; Magne et al., 2005; Hruska & Alter, 2004). Thus, the findings of our study converge well with notions of prosodically induced difficulties in processing information structures beyond the single sentence level.

In addition to the brain responses with negative amplitudes for the perception of focus accentuation, we obtained further ERP effects with positive amplitudes for the processing of sentence elements in focus positions irrespective of their prosodic realization. When processing adequately realized novelty focus elements, a positive-going waveform was elicited that is most pronounced over posterior electrodes and started around 500 ms (Figure 2a). According to the global topographic analyses, a stable brain state for this focus-related positivity was present around 700 ms. During this time, a bilateral posterior positive maximum was apparent in the right-most electric field maps as shown in Figure 2d.

In contrast, when correction foci with adequate accentuation were processed, the positive-going waveform started around 350 ms after the focus onset and was more widely distributed over the scalp (Figure 3a). Specifically, the focus-related positivity was present over bilateral frontal sites within the first two periods of the topographic analyses (left-most and middle maps in Figure 3d). During the last time period of stable brain patterns, a bilateral but centro-posterior positive maximum was apparent (right-most topographic maps in Figure 3d).

When focus positions without adequate prosodic markings were encountered, the above discussed N400 responses are followed by ERPs with positive-going amplitudes. These positivities resembled the positive-going ERPs already observed when focus positions with prosodically adequate realizations were perceived by listeners. For processing novelty focus, the topographic analyses indicated a stable period around 700 ms with a slightly left-lateralized positive maximum (right-most map topographies in Figure 2d). When correction foci with inadequate prosody were encountered, these left-lateralized positive maxima were apparent during both late periods of stable responses (middle and right-most map topographies in Figure 3d).

Taken together, the perception of semantic-pragmatic information foci induced ERP effects with centro-posterior positive maxima. We term the effect Focus-related Positive Shift (FPS) as it resembled commonalities with several previous reports on focus-induced processing effects in the visual (Cowles et al., 2007; Bornkessel et al., 2003) and in the auditory domain (Toepel et al., 2007; Hruska et al., 2004). Yet, it should be pointed out that most of these studies used varying terms for the phenomenon. Our data evinced that the information focus of utterances in discourse was interpreted by listeners irrespective of its prosodic surface marking but was based on the semantic-pragmatic interplay of context and target (current) utterance. Whenever semantic-pragmatic foci were perceived, ERP effects with centro-posterior maxima were induced. However, our findings also refined prior reports on prosody-independent focus-related positivities (Toepel et al., 2007). By extending our analysis methods to topographic pattern analyses, we could identify discrete periods of stable map presences for either condition. In particular, we showed that the electric field topographies differed as a function of the processed focus type, and as a function of presence versus absence of a prosodic focus marking. As these topographic differences are indicative of modulations in the intracranial generators (Lehmann, 1987) further research will also explore the neural networks active during the varying processing conditions.

Our work revealed further evidence for the importance of contextual information in the course of processing connected speech utterances (i.e. dialogs). We showed that unambiguous language contexts induced the strongest expectancies on the information focus in the subsequent utterance, and that discourse requirements restricted the interpretation of focus information. When information that is supposed to be in the focus of the listener did not bear an adequate prosodic surface form, N400 responses indicated the recognition of the surface inadequacy. Yet, the contextual prerequisites were still sufficient to enable the interpretation of the information focus as indicated by the FPS responses.

Our present findings (and the developmental work in progress) will henceforth be relevant for discussions related to neurocognitive modelling of language perception at the discourse level (Hagoort & Van Berkum, 2007; Friederici, 2002).

5. Further work

Extended topographic analyses on our data set indicated the utilization of varying brain networks when listeners perceive focus information in the presence versus absence of adequate prosodic realizations. Whether these differences in the surface activity on the scalp were due to the activation of differential neural networks in the brain will be the subject of further investigations. Moreover, since the current report only provides a subset of data from a longer-termed developmental project, ongoing work will be concerned with developmental aspects of focus perception.

6. Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, grant-number ME 1362/12-1). We would also like to thank Sylvia Stasch in Leipzig for data recording.

References

Bornkessel, I., Schlesewsky M. & Friederici A. D. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter- vs. intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 871–882.

Cowles, H. W., Kluender, R., Kutas M. & Polinsky M. (2007). Violations of information structure: An electrophysiological study of answers to wh-questions. *Brain and Language, 102*, 228–242.

Diaz, M. T. & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research, 1146*, 85–100.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science, 6*, 78–84.

Grosz, B. J. & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*(3), 175–204.

Hagoort, P. & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 801–811.
Halliday, M. A. K. (1967). Notes on transitivity and theme in English (Part 2). *Journal of Linguistics*, 3, 199–244.
Heim, S. & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, 66, 55–68.
Hruska, C. & Alter, K. (2004). How prosody can influence sentence perception. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp. 211–226). Berlin: Mouton de Gruyter.
Jasper, H. H. (1958). Report of the committee on the methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology, 10*, 370–375.
Lehmann, D. (1987). Principles of spatial analysis. In A. S. Gevins and A. Reymond (Eds.), *Methods of Analysis of Brain Electrical and Magnetic Signals, vol. 1* (pp. 309–354). Amsterdam: Elsevier.
Lehmann, D. & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*, 609–621.
Magne, C., Astésano, C., Lacharet-Dujour. A., Morel, M., Alter, K. & Besson, M. (2005). Online processing of "pop-out" words in spoken French dialogues. *Journal of Cognitive Neuroscience, 17*(5), 740–756.
Michel, C. M., Murray, M. M., Lantz G., Gonzalez, S., Spinelli, L. & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology, 115*, 2195–2222.
Murray, M. M., Brunet, D., Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography*, 20, 249–264.
Nieuwland, M. S., Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098–1111.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.
Pannekamp, A., Toepel, U., Alter, K., Hahne, A. & Friederici, A. D. (2005). Prosody-driven sentence processing. *Journal of Cognitive Neuroscience, 17*(39), 407–421.
Steinhauer, K., Alter, K. & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191–196.
Toepel, U. & Alter, K. (2004). On the independence of information structural processing from prosody. In A. Steube (Ed.), *Information structure: Theoretical and Empirical Aspects* (pp. 227–240). Berlin: Mouton de Gruyter.
Toepel, U. (2006). Contrastive topic and focus information in discourse – Prosodic realisation and electrophysiological brain correlates. *MPI Series in Human Cognitive and Brain Sciences, 66*.
van den Brink, D. & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*(6), 1068–1084.
Van Petten C., Coulson S., Rubin S., Plante, E. & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394–417.

Processing and modelling

Neocerebellar emulation of language processing

Giorgos Argyropoulos
Language Evolution and Computation Research Unit, University of Edinburgh

Abstract

Despite the growing literature of empirical findings in support of the involvement of the cerebellum in language processing, the identity of the underlying linguistic computations per se has attracted little, if any, attention at all. Emphasizing language comprehension here, a neurolinguistic model of the posterolateral cerebellum is proposed, grounded on a recently formulated psycholinguistic processor, capturing aspects of rapid, predictive, noise-resistant, and covertly imitative language comprehension processes.

1. Introduction

1.1. The linguistic question on the cerebellum

The lateral cerebellum (henceforth CB) exhibits a pronounced reciprocal expansion with frontal cortical areas (Dow, 1942), and a striking volume increase in hominoids as opposed to monkeys (e.g. MacLeod et al., 2003; Whiting & Barton, 2003). In particular, its reciprocal connectivity with Broca's and Wernicke's area as well as with the dorso-lateral prefrontal cortex (Leiner et al., 1991; Schmahmann & Pandya, 1997; Middleton & Strick, 1998), along with its cytoarchitectural homogeneity, suggestive of unitary CB computations (e.g. Bloedel, 1992; Schmahmann, 1997; Wolpert et al., 1998), encourage the formulation of a neurolinguistic model that makes reference to psycholinguistic processes pertaining to such CB neurocomputations. However, despite the psychopathological evidence for a "lateralized linguistic cerebellum" (Mariën et al., 2001), little if anything has been said about the involvement of such computations per se in language processing, with research on the CB linguistic psychopathology predominantly constrained in the examination of frontal-like syndromes of language disorders that CB deficits might induce, either because of the functional deafferentization of the cortical language-related loci, or because of the disruption of the CB modulatory role, as in verbal working memory (Silveri, 1994; Molinari et al., 1997). The present paper provides a synthesis of psycholinguistic models of language comprehension processes with work on the computations of the posterolateral CB.

1.2. State estimation in language comprehension

Wilson & Knoblich (2005) had reviewed the involvement of covert imitative production mechanisms in the efficient and noise-resistant perception of conspecifics in contexts of increased predictability ("low Kalman gain"). Based on their observations, Pickering and Garrod (2007) propose that production holds a causal role in the efficient perception of noisy/ambiguous linguistic input and in online prediction, and construct a psycholinguistic processor that uses a production-based language "emulator" (Grush, 2004), which is "controlled by feedback from a Kalman filter" (Pickering & Garrod, 2007, p. 108), weighing predictions against analysis of the input at each step. In cases where the prediction is strong and the input noisy, the internal model the language production system provides exerts strong top-down influence to the "input analysis system". Among others, the processor captures cases of on-line prediction in multi-level language processing (see Pickering & Garrod (2007) for a brief review), noise resistance in speech perception (e.g. phoneme restoration phenomena (Warren, 1970), where listeners are unable to detect a phone that is occluded by physically similar noise, based on expectations generated in word or a fortiori sentential contexts), cases of "shallow processing" (Sanford & Sturt, 2002) in sentence comprehension (as manifested, for instance, in Wason and Reich's (1979) "verbal illusions", e.g. *No head injury is too trivial to be ignored*), as well as cases of involvement of speech production mechanisms in speech perception (e.g. Watkins et al., 2003).

2. The model

2.1. Foundations: The CCNMC

Considerations of the neural instantiation of such psychological processes have primarily involved the mirror neuron literature: Wilson & Knoblich (2005) explicitly identify mirror neurons as the neural foundation of the covert imitative involvement of the production system in the perception of conspecifics in contexts of increased predictability. In the same vein, Iakoboni (2005) and Hurley (2008) use the concepts of internal models and emulation to capture the function of mirror neurons in higher cognitive processes. However, what has been ignored in such considerations is the CB (Miall, 2003): the "Cerebellar Cortico-Nuclear Microcomplex" (henceforth CCNMC), i.e., the fundamental CB functional unit, provides the ground of "internal" (Ito, 1984), or "forward" models (Kawato et al., 1987)[1], or "emulators" (Grush, 2004), or, similarly, the neural analogue of the Kalman Filter (Paulin, 1989, 1997).

A CCNMC consists of a microzone, i.e., a small area of the CB cortex, and a small number of nuclear cells. Simply put, the microzone receives two kinds of input, mossy fibers and climbing fibers, and the output is carried by the deep CB nuclear cells. The set of mossy fiber inputs is transformed by the granule cells whose axons form the parallel fibers. The axons of the Purkinje cells, the only output cells of the microzone, are sent to a small group of vestibular or CB nuclear neurons. Long Term Depression (LTD) occurs at parallel fiber-to-Purkinje cell synapses after conjunctive activation of these synapses together with climbing fiber (transmitting the error signals)-to-Purkinje cell synapses (Ito et al., 1982).

2.2. Introducing the NCBKFLP

The "Neocerebellar Kalman Filter Linguistic Processor" (henceforth NCBKFLP; see Figure 1) is proposed here to provide the neurofunctional grounding of Pickering and Garrod's (2007) psycholinguistic processor. Closely following the functional principles of Ito's CCNMCs, the linguistic internal models of the NCBKFLP would learn on the basis of the discrepancies between the predicted state and the actual state that the cortically instantiated "input analysis system" (ibid.) enters: CB predictions transmitted by the ventrolateral neodentate nucleus via the phylogenetically newer parvocellular part of the red nucleus and the ventrolateral nucleus of the thalamus to the mirror neurons of Broca's area are trained by the error signals transmitted via the climbing fibers, with long-term depression (LTD) occurring at the synapses of Purkinje cells-parallel fibers of the posterolateral CB cortex. The inferior olivary nuclei receive signals both directly from the dentate nucleus and indirectly (for references, see Ito, 2008; Schmahmann, 2001) from prefrontal association areas via the zona incerta. In that way, the predictions of the CB linguistic internal model can be compared with the actual output of

[1] In the motor domain, forward models predict the sensory consequences of movements from an efference copy of issued motor commands. Inverse models compute necessary feedforward motor commands from desired movement information. Both kinds of internal models are assumed to be predominantly located in the cerebellum (Kawato, 1999).

the cortical input analysis system. An efference copy of the state of the cortical input analysis system is transmitted by the cerebro-ponto-cerebellar pathway (see Schmahmann & Pandya, 1997), with mossy-parallel fibers reaching the Purkinje cells of the posterolateral CB. A CCNMC might thus connect to the cerebral loop as a reliable copy of the language model in Wernicke's area, with the language process being alternatively conducted by Broca's area acting on the CCNMC rather than on Wernicke's, adaptively avoiding the conscious effort needed for the exploration of cortical loci (Ito, 2000, 2008).

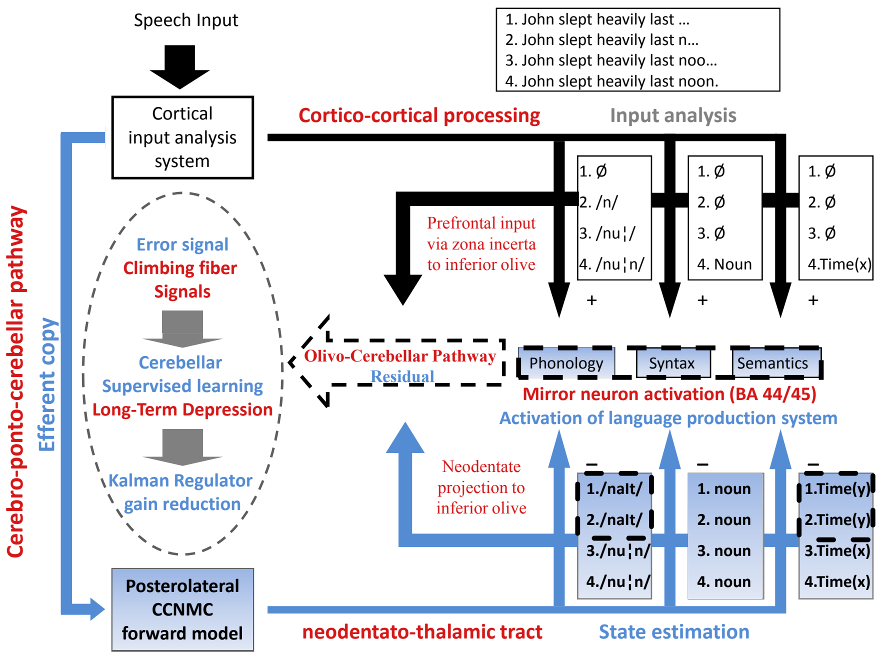


Figure 1: The Neo-Cerebellar Kalman Filter Linguistic Processor provides the neural grounding of Pickering & Garrod's (2007) processor: Mossy fiber inputs, originating at the pontine nuclei, convey, via the parallel fibers, the state of the cortical input analysis system to Purkinje cells at the (right) posterolateral CB cortex. Purkinje cells via their inhibitory connection to the ventrolateral part of the dentate nucleus, via the parvocellular red nucleus and the ventrolateral thalamus, signal the expected state to the mirror neurons of Broca's area, triggering the covert involvement of the language production system. Discrepancies between the CB predictions and the actual sensory input are conveyed as error signals back to the posterolateral CB cortex, via the climbing fibers that originate at the inferior olivary nuclei, which receive indirect connections from the cortical language related areas via the zona incerta and direct projections from the dentate nucleus. Long Term Depression occurs at the parallel fiber to Purkinje cell synapses of the posterolateral cerebellar cortex after conjunctive activation of these synapses together with climbing fiber-to-Purkinje cell synapses.

2.3. Suggestive evidence

Evidence for CB error signalling and error-driven learning in the linguistic domain has already been provided: Fiez et al.'s (1992) CB patient showed deficient supervised learning in a concurrent discrimination word learning task, and likewise failed to detect instances of his erroneous performance in word generation tasks. In a PET study of processing syntactically ambiguous sentences, significant CB activation was found at the point of reanalysis of the sentence (Stowe et al., 2004), suggestive of the CB involvement in error signalling (Jueptner & Weiller, 1998). In the same vein, the gain of a disrupted Kalman regulator cannot be lowered on the basis of the residuals: lack of practice-induced facilitation in linguistic tasks is characteristic in CB patients, who poorly reduce their reaction times and accuracy rates across blocks, as opposed to normal controls in a variety of word generation tasks (Fiez et al., 1992; Gebhart et al., 2002). These findings reflect a more domain-general pattern of lack of performance optimization in CB deficits, most recently demonstrated in Ferrucci et al. (2008), where CB Transcranial Direct Current Stimulation specifically impaired the practice-dependent proficiency increase in verbal working memory, while TDCS over the prefrontal cortex induced an immediate change in the WM task but left the practice-dependent proficiency unchanged.

Some first suggestive evidence for the involvement of the CB in the restoration of noisy linguistic percepts comes from studies in dyslexia accomodatable within the framework of the "Cerebellar Deficit Hypothesis" (Nicolson et al., 2001): the pronounced difficulties of dyslexic subjects in perceiving speech in noisy contexts (Sperling et al., 2005) has been attributed to deficient CB internal models in speech processing (Ito, 2008). The CB predictions matching the input processed by the cortical input analysis system induce "cancellation of reafference" (Nelson & Paulin, 1995; Blakemore et al., 1998), which, at the phonological level, would explain why white noise in an occluded fricative is harder to detect in contexts of highly frequent words (strong top-down influence) as opposed to infrequent ones (Samuel, 1981).

"Not noticing the evidence" (Sanford & Sturt, 2002) in the cases of "verbal illusions" (Wason & Reich, 1979), "pragmatic normalization" (Fillenbaum, 1974), and "shallow processing" phenomena seem to involve the same CB function at the sentential level, where an incorrect mental model is constructed (Garnham & Oakhill, 1996), bypassing the processes of the syntax-semantics interface and providing an interpretation based on the habitual arrangement of the semantic components of the parsed sentence. Internal models of the posterolateral CB have been considered to acquire the dynamics of the manipulation of a particular mental model stored in the temporoparietal cortex as the control object, and thus facilitate processing below the level of awareness (e.g. Ito, 2000). The CB may thus pre-emptively output an arrangement of the semantic components (e.g, `dog´(x), man´(y), bite´(AGENT, PATIENT)`) of the input sentence (e.g. the often pragmatically normalized sentence *The man bit the dog*) based on their predicted, statistically prominent arrangement `(dog´(x), man´(y), bite´(x,y))`, compensating for the often noisy/ambiguous linguistic input, and bypassing the slower cortical computations of the syntax/semantics interface that would guarantee the correct, yet slower and infrequent interpretation `(bite´(y,x))`. Suggestively, the CB has already been shown to be involved in non linguistic illusions of state estimation, as the "size-weight illusion" (Hubbard & Ramachandran, 2004), where CB patients showed a minimized effect of the illusion to different degrees as opposed to normal control subjects.

2.4. Mirror neuron-cerebellar connectivity

In the model thus proposed, the connectivity between mirror neurons in Broca's area and the right posterolateral CB guarantees a functional cooperation of imitation and emulation mechanisms, respectively. While the properties of the connectivity of BA 44/45 mirror neurons with the right posterolateral CB have not yet been researched, it has suggestively been held that CB afferents via the dentatothalamic tract to the F5, i.e., Broca's homologue in monkeys (Petrides & Pandya, 1994) might have input-output relations with the mirror neurons in that area representing the "desired state" in tool-use learning (Imamizu et al., 2003). In the same vein, the estimates of the right posterolateral CB emulator (in cases of a low Kalman regulator gain) would be transmitted to the mirror neurons of Broca's area, triggering the covert employment of production mechanisms in language comprehension.

3. Cortical rules and cerebellar habits

3.1. Syntactic habits as NCBKFLP output

Efferent copies of sentential inputs of the form of [N(oun), V(erb), N(oun)] can be transmitted from the cortical input analysis system to the Purkinje cells of the posterolateral CB cortex via the mossy fiber afferents (the cerebro-ponto-cerebellar pathway). Indeed, the length of the parallel fibers is sufficient to synaptically connect many microzones (e.g. Mugnaini, 1983) and transmit composite contexts to the microcomplexes, while the mediating granule cells ensure that each parallel fiber carries a combination of activity on several mossy fibers, with parallel fibers thus conveying combined contextual representation to Purkinje cells. Combinations of such elementary linguistic categories can be trained in the same fashion that elementary moves are trained into compounds in the CB (Thach et al., 1992). This linguistic internal model, via routinization, can reliably copy the dynamics of the thematic role assignment process, rapidly providing to such input sequences the output of the form [Agent = N1, Action = V, Theme = N2], with the neodentate pre-emptively transmitting this information back to the language-related cortical loci, ensuring the ease and speed of processing active sentences (Forster & Olbrei, 1974).

The cortical neurolinguistic chauvinism take on subcortical structures has often marginalized findings of subcortical linguistic deficits not reflecting frontal-like patterns. The most prominent example is Pickett (1998), the first report in the literature of sentence comprehension deficits attributable to CB pathology, where normal control subjects made errors in the interpretation of 4% of the active sentences and

10% of the passive sentences, whereas CB patients, surprisingly, made errors on 11% of both active and passive sentences, showing no effect of Voice manipulation, with the lateral CB patients of the overall group strikingly performing better than normal control subjects. Such performance patterns, interpretable as "mixed results" (Justus, 2004) from the frontal-like perspective, are even more clearly contrasted by the ones of Broca's agrammatic aphasics in comprehending passive sentences (e.g. Saffran et al., 1998), who exhibit an exaggerated degree of such heuristics-induced (strategic or not) misinterpretations.

3.2. Cerebellar foundations of pseudosyntax

According to Townsend & Bever's (2001) "Late Assignment of Syntax Theory" (LAST), sentence comprehension employs both algorithmic (the actual syntax) and faster, pre-emptive "heuristic, associative" computations (the "pseudosyntax"), the latter utilizing statistically valid information to elicit an initial meaning/form hypothesis, based on which the former will construct a detailed syntactic analysis. The NCBKFLP captures such cases of two-stage serial models of sentence comprehension, with the habits/rules divide being instantiated in the slower cortico-cortical versus the faster (Doya, 1999) corticocerebellar modality of language processing, respectively: the emulational function of the CB advocates in favour of its involvement in such pseudosyntactic operations, as opposed to the actual, algorithmic processes or the very representations of the syntactic templates per se. Characteristically, the notion of "neurofunctional redundancy" has been invoked for the CB (emulated) linguistic representations, as CB aphasia is significantly milder than classical aphasic syndromes, owing to maximal prefrontal compensation (Fabbro et al., 2004). The CB is also fundamental in associative learning in both motor and higher cognitive tasks (Drepper et al., 1999), also forming in many comparative neurocognitive models a fundamental locus of habits (Mishkin et al., 1984), the domain generality of which, supported by its cytoarchitectural homogeneity and its reciprocal connectivity with the language-related cortical loci makes it promisingly inclusive of syntax and semantic processing habits. Thus, with the abolishment of the NVN pseudoparse, i.e., the heuristic of the highly entrenched theta-role assignment of [Agent = N1, Action = V, Theme = N2], the overall "cost for the passive structure" (Ferreira, 2003) now pertains equally to the active voice sentences as well. This provides a concrete neurolinguistic basis for Pickett's (1998, p. 103) own suggestion that "the linguistically impaired cerebellar subjects… can perform the full range of linguistic processing, but their mental shortcuts have been disrupted, so that processing 'easy' sentences requires more conscious effort and is performed less automatically."

The "weak coupling" of the internal model with the copied target system (Paulin, 1997) in language processing-state estimation thus translates into the "good-enough" representations constructed by hearers/readers in sentence comprehension (Ferreira et al., 2002) and the subsequent misinterpretations induced.

Townsend and Bever (2001), assuming a different perspective on the neural instantiation of LAST, propose that pseudosyntactic operations are instantiated in Broca's area, holding that the NVN pseudoparse seemingly found in agrammatic aphasics is but a non-linguistic strategy. However, this immediately makes the proposal that the NVN heuristic is supported by Broca's area run into unfalsifiability issues. Moreover, Broca's area is best construable as a part of the "procedural" cortico-striatal and cortico-cerebellar circuit (Ullman, 2004), so, certain pseudosyntactic operations are conceivably CB-dependent. Finally, Townsend & Bever's proposal (2001) heavily relies on the findings of Linebarger et al. (1983) on the preservation of grammatical knowledge in Broca's aphasics, against which, however, Grodzinsky & Finkel (1998) have shown that agrammatic aphasics exhibit severe deficits in grammaticality judgment tasks involving structures with dependencies between traces and their antecedents, thus showing (regardless of the syntactic framework employed) that the agrammatic deficit is structure-, but not task-dependent. The CB thus appears to be at least one neuroanatomical structure supporting pseudosyntactic operations, and at most the only locus undertaking such linguistic operations.

4. The NCBKFLP in grammaticalization

Grammaticalization is definable as the diachronic process that leads lexical items to grammatical ones inside constructional morphosyntactic contexts. In Argyropoulos (2008), it was proposed that the neocerebellum and the basal ganglia might provide the neural foundation for the domain-general automatization phenomena that constitute the cognitive core of grammaticalization operations (Givón, 1979, and especially 1989; Bybee, 1998; Lehmann, 2004), and that current neurolinguistic research on those structures allows the first steps towards the articulation of a neurolinguistics of grammaticalization.

Given that the products of grammaticalization operations are introduced as instances of optimized linguistic performance at the intra-generational level of language transmission (Haspelmath, 1998), it is important to consider the neurocognitive mechanisms underlying the optimization of repeatedly processed linguistic repertoires, rather than the storage of linguistic representations per se: the significance of the CB in the optimization of thought and language processing, as first proposed by Leiner et al. (1986), promotes this structure as a promising candidate, "a fortiori" given its fundamental role in the practice-induced shift from a controlled (prefrontal and anterior cingulate cortex) to an automated (Sylvian-insular) language processing circuit in linguistic tasks (see Fiez & Raichle, 1997), i.e., the neurocognitive shift suggested to underlie the realization of grammaticalization processes (Givón, 1989).

The very construal of "pseudosyntax" as the process stage of "syntactic habits" (Townsend & Bever, 2001) presupposes the involvement of adaptive demotions from stage 2 to stage 1 operations that routinization of language processing would induce. Characteristically, Kempson & Cann (2007) account for the procliticization (see [1], taken from Bouzouita, 2002) of object pronouns in the transition from Medieval to Renaissance Spanish as "a natural subsequent step of routinization, involving the call-ing up of actions associated with the verb together with those associated with the clitic with a single lexical look-up mechanism".

[1] mas los rompan luego
but CL break.3PL afterwards
'but break them afterwards'

CB supervised learning provides such a "look-up table" or "shortcut-circuit" for mappings originally developed by the "time-consuming cortico-cortical processing" (Doya, 1999, p. 970). Evidence for the involvement of the right posterolateral CB in language processing in a constrained search space has been provided in imaging and behavioural studies: increased CB activation was found in the "FEW" condition in a stem completion task, i.e., where the candidate completions of the stem were few in number (e.g. PSA-), as opposed to increased frontostriatal activation in the "MANY" condition (e.g. STA-), interpreted as involving the selection of a particular response among competing alternatives (Desmond et al., 1998). Gebhart et al. (2002) lend similar emphasis to the size of the search space as a factor affecting the performance of their right posterolateral CB patients, as performance is spared in the subordinate category generation (e.g. stimulus: MONEY, response: *dollar, pound, euro,* etc.) task, where many correct word responses are available, yet is poor in the antonym (e.g. stimulus: KIND, response: *rude*) and verb generation (e.g. stimulus: CHEF, response: *cook*) tasks, where few correct word responses are available.

The "Linear Fusion Hypothesis", predicting that "items that are used together fuse together" (Bybee, 2002), applies widely in grammaticalization operations, with affixation, agglutination, and merger as characteristic cases of universal "syntagmatic coalescence" (Lehmann, 1995) phenomena. Independently of whether chunking in motor/cognitive behaviour is supported by CB state estimation (Paulin, 1989) or by a CB composing function (Thach et al., 1992), the NCBKFLP directly accounts for the automated chunks that hearers process (Bybee, 2002) on the basis of the covert generation of an anticipated co-occurring item B in response to a perceived item A. The impaired anticipatory planning found early in CB patients (Leiner et al., 1987), as well as the severe impairments found in cognitive associative learning (Drepper et al., 1999) advocate in favour of such a kind of CB involvement.

Finally, the core, semantic change in grammaticalization processes (as opposed to the secondary aspect of the formal and further semantic reductions that accompany such shift) has been accounted for by grammaticalization researchers in terms of a demotion from a cognitive "foreground" to a cognitive "background" in routinized discourse processing (Boyer and Harder, forthcoming; Lehmann, 2004). Given the above considerations then, the fact that neuroscientists have used identical terms and notions to explicate the significance of the CB in the automatization of thought processing (Thach, 1998) is suggestive of a perspective of a fruitful synthesis that historical linguistics may reach with recent developments in the field of the neuropsychology of language.

5. Conclusion

The introduction of computations of state estimation in psycholinguistic models provides the cornerstone for the articulation of concrete neurolinguistic models of the posterolateral cerebellum beyond its commonly held strictly modulatory role, along with a set of falsifiable experimental hypotheses to be assessed in forthcoming work by the author.

6. Acknowledgements

My special thanks to Prof. Merle Horne and Mikael Roll for their support, as well as to Dr. Thomas Bak, Dr. Kasper Boye, Dr. Ronnie Cann, Prof. Fernanda Ferreira, Prof. Talmy Givón, Prof. Peter Harder, Prof. Martin Haspelmath, Prof. James Hurford, Prof. Christian Lehmann, Prof. Chris Miall, Prof. Martin Pickering, Dr. Patrick Sturt, and Prof. David Townsend for the encouragement and their helpful comments – I hold exclusive responsibility for any mistakes here. I gratefully acknowledge the Pavlos and Elissavet Papagiannopoulou Foundation for the financial support.

References

Argyropoulos, G. P. (2008). The subcortical foundations of grammaticalization. In A. D. M. Smith, K. Smith & R. F. i Cancho (Eds.), *The evolution of language: Proceedings of the 7th international conference on the evolution of language* (pp. 10–17). Singapore: World Scientific Press.

Barton, S. B. & Sanford, A. J. (1993). A case-study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition*, *21*, 477–487.

Blakemore, S.-J., Wolpert, D. M. & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1*(7), 635–640.

Bloedel, J. R. (1992). Functional heterogeneity with structural homogeneity: How does the cerebellum operate? *Behavioral and Brain Sciences, 15,* 666–678.

Bouzouita, M. (2002). Clitic placement in old and modern Spanish: A dynamic account. MSc. Dissertation. King's College London.

Boye, K. & Harder, P. (forthcoming). Evidentiality: Linguistic categories and grammaticalization. *Functions of Language.*

Bybee, J. L. (1998). A functionalist approach to grammar and its evolution. *Evolution of Communication*, *2*, 249–278.

Bybee, J. L. (2002). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 107–132). Amsterdam: John Benjamins.

Desmond, J. E., Gabrieli, J. D. E. & Glover, G. H. (1998). Dissociation of frontal and cerebellar activity in a cognitive task: Evidence for a distinction between selection and research. *Neuroimage, 7,* 368–376.

Dow, R. S. (1942). The evolution and anatomy of the cerebellum. *Biological Reviews of the Cambridge Philosophical Society, 17,* 179–220.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks, 12,* 961–974.

Drepper, J., Timmann, D., Kolb, F. P. & Diener, H. C. (1999). Non-motor associative learning in patients with isolated degenerative cerebellar disease. *Brain, 122,* 87–97.

Fabbro, F., Tavano, A., Corti, S., Bresolin, N., De Fabritiis, P. & Borgatti, R. (2004). Long-term neuropsychological deficits after cerebellar infarctions in two young adult twins. *Neuropsychologia, 42,* 536–545.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164–203.

Ferreira, F., Bailey, K. G. D. & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Ferrucci, R., Marceglia, S., Vergari, M., Cogiamanian, F., Mrakic-Sposta, S., Mameli, F., Zago, S., Barbieri, S. & Priori, A. (2008). Cerebellar transcranial direct current stimulation impairs the practice-dependent proficiency increase in working memory. *Journal of Cognitive Neuroscience, 20*(9)*,* 1687–1697.

Fiez, J. A., Petersen, S. E., Cheney, M. K. & Raichle, M. E. (1992). Impaired non-motor learning and error detection associated with cerebellar damage. *Brain, 115,* 155–178.

Fiez, J. A. & Raichle, M. (1997). Linguistic processing. In J. D. Schmahmann (Ed.), *The Cerebellum and Cognition, International Review of Neurobiology, 41* (pp. 233–54). San Diego: Academic Press.

Fillenbaum, S. (1974). Pragmatic normalization: Further results for come conjunctive and disjunctive sentences. *Journal of Experimental Psychology, 102*, 574–578.

Forster, K. I. & Olbrei, I. (1973). Semantic heuristics and syntactic analysis. *Cognition*, *2*, 319–347.

Garnham, A. & Oakhill, J. V. (1996). The mental models theory of language comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 313–339). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gebhart, A. L., Petersen, S. E. & Thach, W. T. (2002). Role of the posterolateral cerebellum in language. *Annals of the New York Academy of Science, 978,* 318–333.

Givón, T. (1989). *Mind, code and context: Essays in pragmatics.* New York: Academic Press.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Grodzinsky, Y. & Finkel, L. (1998). The neurology of empty categories. Aphasics' failure to detect ungrammaticality. *Journal of Cognitive Neuroscience, 10*(2)*,* 281–292.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27,* 377–435.

Haspelmath, M. (1998). Does grammaticalization need reanalysis? *Studies in Language, 22,* 315–51.

Hubbard, E. M. & Ramachandran, V. S. (2004). The size-weight illusion, emulation, and the cerebellum. *Behavioral and Brain Sciences, 27,* 407–408.

Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences, 31,* 1–58.

Iacoboni, M. (2005). Understanding others: Imitation, language, empathy. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From cognitive neuroscience to social science. Volume 1: Mechanisms of imitation and imitation in animals* (pp. 77–99). Cambridge, MA: MIT Press.

Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T. & Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proceedings of the National Academy of Sciences of the United States of America, 100*(9), 5361–5466.

Ito, M. (1984). *The cerebellum and neural control.* New York: Raven Press.

Ito, M. (2000). Neural control of cognition and language. In A. Marantz, Y. Miyashita & W. O'Neil (Eds.), *Image, language, brain* (pp. 149–162). Cambridge, MA: MIT Press.

Ito, M. (2002). Historical review of the significance of the cerebellum and the role of Purkinje cells in motor learning. *Annals of the New York Academy of Sciences, 978,* 273–288.

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience, 9,* 304–313.

Ito, M., Sakurai, M. & Tongroach, P. (1982). Climbing fiber induced depression of both mossy fiber responsiveness and glutamate sensitivity of cerebellar Purkinje cells. *Journal of Physiology (Lond.), 324,* 113–134.

Jueptner, M. & Weiller, C. (1998). A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. *Brain, 121,* 1437–1449.

Justus, T. (2004). The cerebellum and English grammatical morphology: Evidence from production, comprehension, and grammaticality judgments. *Journal of cognitive neuroscience, 16*(7), 1115–1130.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering, 82,* 35–45.

Kawato, M. (1999). Internal models for motor control and trajectory planning, *Current Opinion in Neurobiology*, *9,* 718–727.

Kawato, M., Furukawa, K. & Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics, 57,* 169–185.

Kempson, R. & Cann, R. (2007). Dynamic syntax and dialogue modelling: Preliminaries for a dialogue-driven account of syntactic change. *Current Issues in Linguistic Theory, 284*, 73–101.

Lehmann, C. (1995). *Thoughts on grammaticalization.* Munich: Lincom Europa. (First published as *akup* 48, Institut für Sprachwissenschaft, Universität zu Köln, 1982.)

Lehmann, C. (2004). Theory and method in grammaticalization. *Zeitschrift fur Germanistische Linguistik 32*(2), 152–187.

Leiner, H. C., Leiner, A. L. & Dow, R. S. (1986). Does the cerebellum contribute to mental skills? *Behavioral Neuroscience, 100*(4), 443–54.

Leiner, H. C., Leiner, A. L. & Dow, R. S. (1987). Cerebro-cerebellar learning loops in apes and humans. *Italian Journal of Neurological Science, 8,* 425–436.

Leiner, H. C., Leiner, A. L. & Dow, R. S. (1991). The human cerebro-cerebellar system: Its computing, cognitive, and language skills. *Behavioural Brain Research*, *24*, 113–128.

Linebarger, M. C., Schwartz, M. F. & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13,* 361–392.

MacLeod, C. E., Zilles, K., Schleicher, A., Rilling, J. K. & Gibson, K. R. (2003). Expansion of the neocerebellum in Hominoidea. *Journal of human evolution, 44,* 401–429.

Mariën, P., Engelborghs, S., Fabbro, F. & De Deyn, P. P. (2001). The lateralized linguistic cerebellum: A review and a new hypothesis. *Brain and Language, 79,* 580–600.

Miall, R. C. (2003). Connecting mirror neurons and forward models. *NeuroReport, 14*(16), 1–3.

Middleton, F. A. & Strick, P. L. (1998). Cerebellar output: Motor and cognitive channels. *Trends in Cognitive Sciences, 2*(9), 348–354.

Mishkin, M., Malamut, B. & Bachevalier, J. (1984). Memories and habits: Two neural systems. In G. Lynch, J.L. McGaugh & N.M. Weinberger (Eds.), *Neurobiology of learning and memory* (pp. 65–77). New York: Guildford Press.

Molinari, M., Leggio, M.G. & Silveri, M. (1997). Verbal fluency and agrammatism. In J.D. Schmahmann (Ed.), *The cerebellum and cognition, International Review of Neurobiology, 41* (pp. 325–339). San Diego: Academic Press.

Mugnaini, E. (1983). The length of cerebellar parallel fibers in chicken and rhesus monkey. *Journal of Computational Neurology, 220,* 7–15.

Nelson, M.E. & Paulin, M.G. (1995). Neural simulations of adaptive reafference suppression in the elasmobranch electrosensory system. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 177*(6), 723–736.

Nicolson, R.I., Fawcett, A.J. & Dean, P. (2001). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences, 24,* 508–511.

Paulin, M.G. (1989). A Kalman filter theory of the cerebellum. In M.A. Arbib & S.-I. Amari (Eds.), *Dynamic interactions in neural networks: Models and data* (pp. 239–259). New York: Springer.

Paulin, M. (1997). Neural representations of moving systems. In J.D. Schmahmann (Ed.), *The Cerebellum and Cognition, International Review of Neurobiology, 41,* (pp. 515–533). San Diego: Academic Press.

Petrides, M. & Pandya, D. (1994). Comparative architectonic analysis of the human and the macaque frontal cortex. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. IX) (pp. 17–58). New York: Elsevier.

Pickering, M.J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105–110.

Pickett, E.R. (1998). *Language and the Cerebellum.* Doctoral dissertation, Brown University.

Saffran, E.M., Schwartz, M.F. & Linebarger, M.C. (1998). Semantic influences on thematic role assignment: Evidence from normals and aphasics. *Brain and Language, 62,* 255–297.

Samuel, A.G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110,* 474–494.

Sanford, A.J. & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*(9), 382–386.

Schmahmann, J.D. & Pandya D.N. (1997). The cerebrocerebellar system. *International Review of Neurobiology, 41,* 31–60.

Silveri, M.C., Leggio, M.G. & Molinari, M. (1994). The cerebellum contributes to linguistic production: A case of agrammatic speech following a right cerebellar lesion. *Neurology, 44,* 2047–2050.

Sperling, A.J., Lu, Z.-L., Manis, F.R. & Seidenberg, M.S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience, 8*(7), 862–863.

Stowe, L.A., Paans, A.M.J., Wijers, A.A. & Zwarts, F. (2004). Activations of "motor" and other non-language structures during sentence comprehension. *Brain and Language, 89,* 290–299.

Thach, W.T. (1998). What is the role of the cerebellum in motor learning and cognition? *Trends in Cognitive Sciences, 2*(9), 331–337.

Thach, W.T., Goodkin, H.P. & Keating, J.G. (1992). The cerebellum and the adaptive coordination of movement. *Annual Review of Neuroscience, 15,* 403–442.

Townsend, D. & Bever, T.G. (2001). *Sentence comprehension: The integration of habits and rules.* Cambridge, MA. MIT Press.

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92,* 231–270.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392–393.

Wason, P. & Reich, S. S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology, 31,* 591–597.

Watkins, K. E., Strafella, A. P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41,* 989–994.

Whiting, B. A. & Barton, R. A. (2003). The evolution of the cortico-cerebellar complex in primates: Anatomical connections predict patterns of correlated evolution. *Journal of Human Evolution, 44,* 3–10.

Wilson, M. & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131*(3)*,* 460–473.

Wolpert, D. M., Miall, R. C. & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2*(9)*,* 338–347.

On the emergence of early linguistic functions: A biological and interactional perspective

Francisco Lacerda
Department of Linguistics, Stockholm University

Abstract

Human infants typically utter their first words within the second half of their first year of life and about one year later they tend to take their first steps towards the full exploration of the linguistic combinatorial power of their ambient languages. From the overwhelming exposure to continuous speech utterances produced by other speakers, the infant derives the underlying linguistic structure of the utterances occurring in the ambient language.

This chapter proposes a model of how general perception, production, and interaction mechanisms may account for the typical process of early language development leading the infant to the discovery of words embedded in the continuous speech of the ambient language.

1. Introduction

The human infant's language acquisition is often pictured as "effortless". Although it is not possible to produce an independent and reliable assessment of how "effortless" first language acquisition really is, the typical second-language learner's experience suggests that learning a new language after childhood appears to call for much more overt work to master a new language than what seems to be the case for the young infant acquiring the ambient language. This contrast between the level of explicit commitment necessary to learn a second language and the "easy way" in which the infant appears to acquire its native language presents somewhat of a linguistic paradox. After all, a second-language learner has, by definition, acquired a first, native, language and therefore it could be expected that the first language might be advantageous when learning a new language. Yet, mastering a second language at an acceptable performance level typically demands active and explicit study of the new language's phonetic, phonological, and syntactic aspects. Why does the naive young infant acquire its native language without requiring the structured language tuition that seems to be necessary to teach a new language to an already full-blown speaker of a native language?

A classical answer to this question builds on innate language-oriented processes (Pinker, 1994) in combination with the notion of critical periods (Lenneberg et al., 1965; Lenneberg, 1967) and proposes essentially that an infant's biological predisposition to language learning is dampened after the first years of life. The underlying notion is that the unfolding biological program for genetically determined language learning is available only early in life. However, while there is no question that the infant's biological development interacts with the language learning process, viewing language acquisition from an interactional perspective where the infant is pictured in the typical ecologic settings of early exposure to the ambient language provides a more flexible and rather plausible view of the process (MacWhinney, 1999). As it will be shown below, when the ecological aspects of the infant's typical language acquisition settings are taken into account, the "ill-formed" sentences, interruptions and sudden shifts in focus generally revealed by the analysis of infant-directed speech become much more coherent and plausible as the basis for emerging linguistic functions in early infancy. By the same token, the adult learner's ecological settings along with their matured cognitive development seem to create a language-learning scenario with strong biases towards an analytic and comparative approach. In addition to the biological development per se, the adult has a life experience that largely exceeds that of the infant's and explores previous knowledge of the world and how spoken language is used to communicate with others. Besides the dramatic neurological development between infancy and adolescence (Chugani et al., 1987), the changes in processing ability are also a consequence of the adaptive and specialization process associated with language acquisition, which feeds back on itself. In other words, the prerequisites for language learning in immature versus mature speakers are very different not only because of the biologically determined differences (that may or may not have a direct impact on language learning ability per se) but because of the substantial differences in the individual's experiences of spoken communication and previous knowledge of the world and properties of objects in it.

From a general perspective, the native language acquisition process can be seen as grounded already in the pre-natal period, when the foetus' auditory system becomes mature enough to represent the sounds that reach it in the womb. Of course, this is not language acquisition in a strict sense, nor is it an innate propensity to acquire language. It is a simple bias towards one frequent type of sound heard in the womb (Vince et al., 1985; Murphy & Smyth, 1962; Querleu et al., 1988). Such a bias can be seen as a consequence of general memory and sensory representations of acoustic events. It arises because the acoustic characteristics of vocalizations (primarily the prosodic speech characteristics of the pregnant woman's utterances) are rather different from those of other competing sounds in the womb's acoustic environment (Walker et al., 1971). Thus, the differentiated physical properties of the utterance sounds may in general be enough to induce a spontaneous representation of speech sounds in contrast to non-speech sounds but it may even be enhanced by the recurrent co-occurrences of hormonal levels associated with the woman's general state of alertness during vocalizations. These factors suggest that pre-natal exposure to the mother's voice filtered by the bodily structures around the foetus may create biases towards the most frequent prosodic patterns occurring in the mother's speech. After birth, such biases are likely to trigger the newborn's interest for the mother's voice because her speaking style necessarily matches a pattern of acoustic representations known to the newborn, as demonstrated in some classical experiments on neonatal speech recognition (de Casper & Fifer, 1980; de Casper & Spence, 1986; Spence & de Casper, 1987; Lecanuet et al., 1987; Moon & Fifer, 1990; Nazzi et al., 1998). Given the ecological context of the initial mother–infant interactions, the newborn's interest for the prosodic characteristics of the mother's voice is readily picked up by the mother who spontaneously experiments with her voice and rapidly discovers that enhancing the prosodic characteristics of her utterances is a good means to maintain the communication link with her infant. In this spontaneous way the mother's speaking style quickly converges to what is usually named by infant-directed speech (IDS) style (also referred to as "motherese" or "parentese"). The typical characteristic of this speaking style is an enhanced prosodic contour (Fernald & Simon, 1984; Fernald, 1989; Sundberg, 1998) compared to what can be expected in adult-directed speech (ADS) (Fernald & Mazzie, 1991).

The proper process of language acquisition starts as the infant is involved in interaction with actors, objects, or events in their ecological context. Although the newborn's initial bias towards speech is likely to play a positive role in the initial stages of the language acquisition, the bias is obviously not a necessary pre-requisite for successful language acquisition. Adults and older siblings in the infant's neighbourhood tend to explore intuitively other means of directing the infant's attention towards speech utterances. The acquisition of spoken language in hearing infants with deaf, sign-language speaking parents, demonstrates that the initial bias is not a necessary condition for the language acquisition process. Indeed, the present account of the initial stages of the language acquisition process suggests that general memory and sensory representation in combination with interaction processes underlie the emergent process of first language acquisition, given the infant's focus of attention on the speaker or on external objects. The language acquisition scenario sketched here assumes a 1- or 2-month-old infant interacting with caregivers or siblings in a typical ecologic setting, offering the opportunity of complex multisensory exposure or interaction with humans, objects, or events in the infant's neighbourhood. It is also assumed that by the end of the first year of life, the infant shows clear signs of having established some sound-meaning representations (Roug et al., 1989; Vihman & McCune, 1994; Vihman, 1996). Although still lacking the fine articulatory skills necessary to produce controlled and reliable utterances, the one-year-old responds adequately to words and utterances heard in the immediate language environment. Throughout the first year of life, the infant shows progressively more and more specific responses to words or sentences. The infant looks at the relevant objects that the adults' or older siblings' utterances refer to, thereby disclosing its capacity to associate the sound of certain words with objects or events somehow available nearby. The faculty of establishing tentative sound-object or sound-event associations is remarkable by itself but the processes underlying the development of such a faculty are even more astonishing given the complexity of the infant's physical world. Indeed, rather than being exposed to "simple" isolated words that might be easy to relate to available objects or events, the infant is normally exposed to continuous speech where not only potentially relevant words are embedded in more or less complex contexts but where there are also several plausible available objects or events that might be referred to. Therefore, a phonetic and psychological account of the early stages of language acquisition on the basis of the infant's perceptual and motor capabilities and of the caregivers' behaviour in the ecological setting of a typical adult–infant interaction poses an enormous challenge. Studies of early language acquisition, with all their wide implications and multidisciplinary character raise fundamental scientific issues that are triggering the inspiration and commitment of researchers from many different scientific areas.

2. An ecologic approach to language acquisition

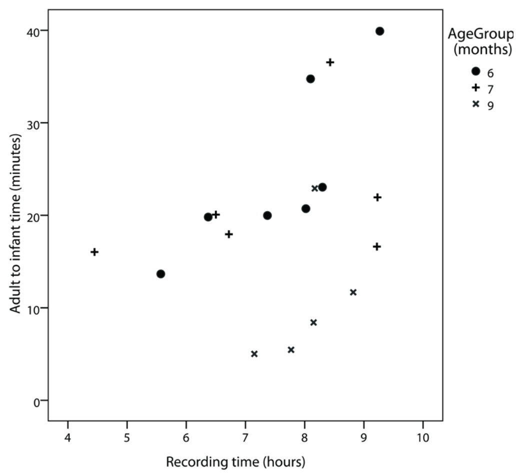
2.1. ETLA

Typical speech does not consist of words uttered in isolation. Instead, words are uttered in sequences where word boundaries are diffuse unless the speaker actually makes a pause. This is also true for the type of speech the infant is exposed to. The "ill-formed" incomplete utterances reaching the infant in a natural communication setting appear not to contain enough information to enable the infant to find out something useful about the linguistic structure of its ambient language. When interacting with infants, adults typically use utterances containing series of words rather than just isolated words. Although the syntactic structure of the utterances produced by adults in such interaction settings may be somewhat simpler in relation to what they would have produced in an adult-to-adult conversion (Thiessen et al., 2005), the acoustic pattern reaching the infant represents nevertheless whole utterances, where individual sound segments are embedded in a speech sound continuum that is affected by co-articulation both within and across the individual word boundaries. Under these circumstances, discovering the underlying linguistic structure implicit in the utterances is bound to be a difficult task in as much as there is no apparent acoustic indication of what the sound elements embedded in the utterances may be. However, considering the infant's sensitivity to regularities in the transition probabilities between consecutive sounds (Saffran & Thiessen, 2003; Saffran et al., 2001; Aslin et al., 1998; Saffran et al., 1996) and the characteristics of typical IDS, the early steps of the language acquisition process may after all be accounted by the interaction between more parsimonious general-purpose processes leading to language rather than by the a priori assumption that the language acquisition process relies on initial linguistic knowledge (Pinker, 1994; Pinker, 2002). Indeed, if the interaction between the infant and the adult is seen in its ecologically relevant context, the ability to detect statistic regularities at both the segmental and supra-segmental levels becomes an important resource in accounting for the infant's early language acquisition process. Of course this does not imply that the infant actually uses statistical computations to pick likely "word-candidates" from the continuous string of sounds to which it is exposed. The processes underlying early language acquisition can in fact have simpler roots than that. If basic aspects of the infant's interaction with its linguistic environment are taken into account along with the characteristics of the typical ecological settings and the adult behaviour when interacting with the infant, a relatively simple model of early language acquisition can be achieved, even if the model only calls for general-purpose, non-linguistically motivated, sensory representation and general interaction processes. Attempting to account for the early language acquisition in terms of such general and non-linguistically motivated principles is the goal of the proposed Ecological Theory of Language Acquisition, ETLA (Lacerda & Sundberg, 2006). This theoretical framework suggests that independently motivated memory and sensory representations along with interaction processes integrated in the infant's ecological context are key factors accounting for the infant's initial discovery of the ambient language's linguistic structure.

2.2. Scenarios of adult–infant interaction

A critical aspect of the present account of the language acquisition process is the long-term interaction between the infant and the adult. In its daily home environment, the infant is typically immersed in an acoustic environment where IDS is just a small fraction of all the acoustic stimuli the infant experiences (van de Weijer, 1999). It should be pointed out that IDS is used here in a generic sense to refer to speech that is directed to the infant. There is no implication that the "IDS-typical" prosodic characteristics actually should be present during these adult–infant interaction periods. The only assumption is that the adult manages to catch the infant's attention and involvement in the speech communication situation, but that does not necessarily have to be done by using expanded F0-contours. In other words, the relevant aspect is the adult's sensibility and ability to maintain the infant's focus on the speech communication situation while the means to catch the infant's attention, per se, are not of primary importance. Data adapted from van de Weijer (1999) was used to compute the actual amount of time the infant was exposed to IDS, either from an adult or sibling, within each of the

Figure 1. Amount of time per recording period that an infant listened to IDS from an adult, according to data in van de Weijer (1999, Table 2.3). Three age ranges: 6 months of age (circles), 7 months of age (diamonds) and 9 months of age (squares).



recording sessions. Figure 1 shows the amount of time that adult speakers spent talking to the infant, in relation to the total duration of the recording session, and Figure 2 shows the amount of time an older sibling of the infant spent speaking to the infant during the same sessions.

The different symbols in the figures indicate the infant's age in months (symmetrically rounded). Forcing the data from each of the age ranges into linear regression models with "Adult to infant time" as the dependent variable against "Recording time", revealed that the slopes of the regression curves within each age group were rather low. The non-standardized coefficients obtained were 0.101, 0.026 and 0.095 for the 6, 7 and 9 months age ranges, respectively. Essentially this means that the infant is exposed to IDS from the adults only 3–10% of the time. The pattern of rather low IDS incidence is again illustrated in Figure 3, where the overall IDS time relative to the non-IDS time and the remaining recording time is displayed in pie-charts.

To the extent that van de Weijer's (1999) data is representative of how an infant's exposure to its ambient language looks like in a Western culture, it is apparent that most of the infant's language input is speech heard in the background or between speakers who interact with each other but not directly with the infant. Under these circumstances, the argument of the "poverty of stimulus" would indeed seem justified[1] but the underlying assumption, that the IDS impact would be proportional to relative frequency of occurrence of the IDS, is probably wrong. More than the infant's general exposure to the ambient language, the short periods of adult–infant interaction during which the infant is focused on the adult's speech are very likely to be the most

1 In the specific case of language acquisition, this argument states that the variability in the speech input-reaching the infant is so vast that it is virtually impossible to extract underlying linguistic meaning from the immediate co-occurrences of particular utterance-object (or utterance-action) instances, unless the infant has some form of innate knowledge about linguistic principles.

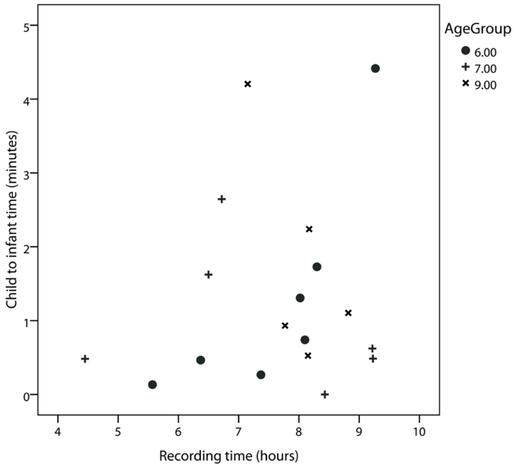


Figure 2. Amount of time per recording period that an infant received speech input from a child, according to data in van de Weijer (1999, Table 2.3). Data sorted into three age ranges: 6 months of age (circles), 7 months of age (diamonds) and 9 months of age (squares).

important time windows in terms of the early language acquisition process. Another aspect to consider is the type of utterances that the infant hears in IDS. Again, according to van de Weijer's (1999) data, most (45%) of the adult to infant utterances in his analyzed material fall within the category "others", i.e. fillers, interjections, vocatives, or social expressions. This category of utterances is likely to be a limited source of linguistic information, as far as picking the underlying syntactic and combinatorial properties of language is concerned. Yet, the expressions involved in this category are still likely to contribute to the formation of socially adequate sound-meaning associations as the utterances tend to occur in rather well-defined interaction contexts. Thus, while the category "other" may not contribute to lexical, syntactic, or phonological development as such, the occurrence of such utterances in certain typical situations is likely to provide the infant with implicit information on how the expressions may be linked to certain aspects of the social interplay. The next most frequent types found in van de Weijer's material were declarative utterances (23.5%), followed by interrogatives (9.7%) and imperatives (8.4%). Van de Weijer's data for this 6-to-9-months old infant also corroborates the findings of a previous study of Dutch-speaking mothers speaking to their 2-year-old children (Snow et al., 1976). Although two years of age is far from the developmental stage envisaged by the present account on the emergence of early language acquisition, the consistency between these two sets of data is reassuring.

From the perspective of early language acquisition, the time windows during which the adult engages in active interaction with the infant can therefore be seen as particularly relevant from an experimental and theoretical point of view because they provide optimal observation periods during which the infant is intensively exposed to IDS at the same time that the adult speaker also tends to be focused on the infant's responses to the messages conveyed by that IDS.

There are several potential adult–infant interaction scenarios. One scenario pictures

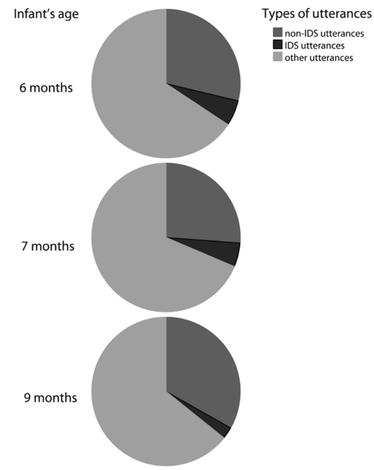


Figure 3. Distribution of IDS, non-IDS and other events during recording times when the infant was 6, 7 and 9 months old (based on data from van de Weijer, 1999, Table 2.3).

the infant as a passive listener of the ambient language. According to van de Weijer's data and the intuition from everyday-life experience, this should be the most common language exposure early in life. The infant may not explicitly pay attention to the content of what is said in the background acoustic ambience, but is nevertheless exposed to the sounds of the ambient language. Under such conditions language acquisition is probably mainly guided by the principles of statistical learning, provided that the infant attends to the sound of the ambient language. To the extent that van de Weijer's data is representative of this scenario, the average 2.56 hours/day (van de Weijer, 1999, p. 36) that the infant can listen to speakers of its ambient language offers by itself an outstanding opportunity of deriving the spoken language's phonotactic and prosodic regularities. Indeed, if infants are able to pick up statistic regularities in only a few minutes of continuous speech (Saffran et al., 1996; Aslin et al., 1998; Saffran et al., 2001), they can also be assumed to be able to detect the statistical regularities embedded in the naturally produced utterances of their ambient language. Even considering that the complexity of the ambient speech is larger than that of the sound sequences used by Saffran and colleagues, data revealing that 4-month-old infants do recognize the sound structures of their names (Mandel et al., 1995; Mandel & Jusczyk, 1996) strongly suggests that infants manage to detect such statistical regularities even in natural listening conditions.

But while the statistical regularities available from the general phonotactic properties of the ambient language may lay the ground for the identification of word-like sequences, retrieving the linguistic referential function from such word candidates is less likely on the basis of passive exposure to the ambient language alone. Therefore the scenario of early language acquisition involving active adult–infant interaction and IDS becomes a natural complement to such pure statistical learning. Under these circumstances, the adult's monitoring of the infant's reactions during the interaction situation is likely to provide a very strong link between the sounds of words and the objects or events that they refer to. In reality, this interaction scenario should be further specified to reflect the goals of the adult during the interaction. In a situation in which the adult attempts to present novel objects to the infant, it may be expected that the utterances produced tend to present the object's name and features while making sure of keeping the object within the infant's visual field. If no specific target objects are present in the interaction setting, the adult's discourse may instead be more likely to describe the setting, the infant's general state of mind or just follow the infant's momentary focus of attention, as a narrator would describe the world for the infant (Nelson, 1988). In both interaction situations, the co-occurrence of acoustic stimuli (the referential utterances) along with objects or events that they refer to should promote learning of the linguistic referential function.

3. Adult–infant interaction and the emerging lexicon

During adult–infant interaction periods, the infant's focus of attention is monitored by the adult and it is likely that the adult utterances produced under such circumstances have some relation to whatever the adult perceives to be the infant's current focus of attention or needs. Co-occurrence of utterances with objects or situations to which the utterances may relate appears as a potentially determinant factor influencing early language acquisition. To investigate the extent to which the emergence of linguistic structure may be predicted from utterances produced under such focused interaction settings, an explicit mathematical model was created and applied to actual adult–infant interaction data. In line with the hypothesis that focused adult–infant interaction contains optimal information to trigger early language acquisition, the model should be able to generate plausible lexical candidates using data from naturalistic interaction. The model itself assumes no a priori linguistic knowledge but it can store representations of sensory information and it has a similarity metric that can be applied to the representations.

3.1. How large is the infant's acoustic space?

In an attempt to achieve an estimate of the number of "possible" sounds available to the naive infant, the following combinatorial exercise was carried out:

Assuming that the infant can discriminate 17 frequency bands and that within each of these bands the infant can discriminate 17 intensity steps (from the hearing threshold to the maximum threshold within each of these bands), how many different sounds would the infant be able to discriminate, provided that any intensity change in any of the frequency bands could be detected as a "novel sound"?

The combinatorial explosion created by the combination of each of these levels, in each of the bands leads to $17 * 17 * 17 * \ldots * 17 = 17^{17} \approx 10^{20}$ potentially discriminable sounds. This is an astronomic number! If it were to represent human body cells, it

Table 1. Transcribed example of a mother's speech to her 3-month-old son.

Start time	End time	Utterance
1.060	1.873	aa
3.672	4.929	haa just
5.988	6.604	hamm
7.689	8.896	a å vet du va
9.586	13.701	vem va inte med på frukosten också aa
16.560	18.408	heehahaha jaa
20.552	21.291	haa
24.101	24.593	ja
28.413	29.916	jahaa
31.124	31.937	va de så
33.884	37.137	å storasyster fick yoghurt ojj
41.474	42.410	ooj
43.840	49.113	å storasyster fick yoghurt och stora stora mackor a men vet du va
49.754	53.352	ja jaa ja men vet du va vi ska ju prata
54.337	55.003	vet du va
55.988	56.802	tittuut
57.516	59.783	aa men vet du vi kan prata xxx

would be necessary to have on the order of 10 000 000 individuals to correspond to a total of 10^{20} cells! The estimate is, of course, crude but still it gives an indication of the immensely vast acoustic search space in which the infant must navigate to pick up the relevant acoustic signals involved in speech communication and associated with meaning. Indeed, against the background of such a number of possible sound alternatives, the argument of the poverty of stimulus appears as quite plausible. However, because the speech sounds occurring in the infant's ecological environment are not random samples drawn from this acoustic domain but instead sounds used in connection with descriptions of objects or events in the infant's neighbourhood, the probability of repetition of sound sequences within an adult–infant interaction situation is actually enormously higher by comparison to what might be expected if the sounds were just drawn at random from that search space. This is the gist of the proposed ETLA and the model of lexical emergence that will be described below.

3.2. Adult-infant interaction data

The data used to test the model has been collected at Stockholm University's Phonetics laboratory. Several aspects of adult–infant interaction are being studied using high-quality video and audio recordings of sessions where adults and infants engage in interaction under naturalistic but controlled laboratory conditions. The ecologic relevance of these recordings is assumed to be acceptable. The sessions last for about 30 minutes, which is longer than the average daily adult to infant speaking time reported by van de Weijer (1999). The recording situation involves free play in a studio that is arranged to look like a small living room, where cameras and wireless microphones are placed so that they do not interfere with movements nor draw attention. A selection of age-adequate toys is available in the room and some of the toys are given labels that the care-giver is expected to use spontaneously throughout the recording session.

Table 1 shows a transcription excerpt of the first 60 seconds of a 16 minute long recording session where one mother interacts with her 3-month old son. The start and end time for each utterance is given, along with the transcription of the mother's utterance. Although in the transcriptions of the utterances spaces are used as delimiters to separate consecutive words, within-utterance delimiters are displayed only for ease of visualization but are actually not considered in the analysis of the materials because the model simulating the infant's early language acquisition stages is not to have a priori linguistic knowledge. Rather than word-boundaries, the only plausible delimiters at this level are the actual pauses between utterances. To be sure, the background exposure to the ambient language referred to above may contribute to statistical information on transitional probabilities between possible consecutive sounds and could help the infant to derive phonotatic biases towards spotting possible word boundaries along the speech continuum of an utterance but it will not be considered at this stage.

3.3. Main features of the model

The current emergent lexicon model is intended to demonstrate how general and non-linguistic processes may nevertheless lead to the emergence of linguistic structure under the ecological settings of typical adult–infant interaction early in life. The model has no knowledge of what a word or any other grammatical component is, but it has the ability to store representations of acoustic signals. In fact the model does not even know what a "speech signal" is, but this aspect is bypassed in the current version as the model is fed with segmented speech signals that effectively select the relevant information right from the beginning. Thus, at this point the model input consists of character strings simulating the actual acoustic signal rather than the signal itself (a prototype version of the model that accepts audio inputs is being evaluated) but the model includes all the relevant functionality, even though it is much simplified because of the more treatable string input. To represent the fact that input signals are treated holistically as sequences of sounds in which words or other grammatical categories do not stand out as elements, the string representations are stripped from all space delimiters before being processed by the model. Indeed, only the pauses between utterances are marked by delimiters.

Table 2. Lexical candidates derived by the model under assumption of 20 sec. memory span and memory primacy.

Infant's age: 2.97 months
Number of lexical candidates found in "mamma" utterances: 11
Assumed memory span: 20 s and memory primacy
Total recorded duration: 974.449 sec.
Effective total utterance time: 651.407 sec.
Lexical items' rate: 0.01129 candidates/sec. (rel. total recorded duration)
Effective lexical items' rate: 0.01688 candidates/sec. (rel. total utterance time)

Items (memory primacy):	Repetitions within span
va	6
mmm	6
jaa	4
mhmhmhmhmhm	3
jaha	2
titta	2
ja	2
mhmhmhmhmhmhmhmhm	1
intefärdig	1
kucka	1
haa	1

The basic model components are memory storage of the incoming stimuli representations and a recursive similarity measure that compares the representation of each new input with previously stored representations. In addition to these basic functions, a "memory span" variable and a "memory primacy/recency" switch were included in order to study the impact of different types of temporal memory effects on the emergent lexicon.

3.4. Emerging lexical candidates

Using a memory span of 20 seconds and memory primacy dominance (Beckman & Edwards, 2000), the model generates the possible lexical candidates presented in Table 2 from the full 16 minutes of the data.

The data shows that interjections, expressions, and isolated words occur repeatedly in the adult's utterances, even when considering that the infant's memory span only stores the last 20 s of utterances and that lexical candidates are selected only when they match the initial parts of an utterance. To study how primacy and recency effects would result into different sets of lexical candidates, the same speech material was processed under the assumption of no bias towards primacy or recency effects (Table 3) as well as recency effects only (Table 4).

In this recording session, a doll named "kucka" was given to the mother to use in her interaction with the infant. However, in this case, the mother did not actually refer to the toys in the typical repetitive way that usually is observed in IDS at this age. She instead started speaking about something that had happened at home during breakfast earlier in the morning, involving the infant and an older sibling. Except for the fact that the recorded adult–infant session did not turn out to have the planned focus on the doll, the mother's strategy was perfectly adequate. She skilfully explored the infant's experience of the situation earlier in the morning to maintain the communication link and the infant's interest on the interaction situation. Indeed it can be argued that by capturing the infant's attention, the mother actually increases the probability of successfully introducing the doll's name, although she did not have the opportunity of exposing the infant to the target name as would have been the case if the infant was interested in the doll right from the beginning. From the perspective of early language acquisition, the strategy used by this mother must therefore be seen as productive. Of course it is not possible to guarantee that the mother was actually trying to introduce the target doll to the infant. She might have had another agenda as well. However, the potentially lower rate of direct exposure to the target word is likely to be balanced by the increased attention that the infant displayed during the periods in which the word was introduced. The name of the doll was repeated four times during the session, although not always within the 20 s of memory span selected for this run of the model. The number of repetitions was lower than what has been observed in other situations (Lacerda & Sundberg, 2006) but the model still indicates that the target word would have been picked as a lexical item. Another word that also pops up is *titta* 'look' that the mother uses to catch the infant's attention.

Similar results were obtained for recency memory bias (Table 4) but the target word "kucka" was actually lost under these conditions because the mother had uttered the word emphatically, with a rather long final vowel that was transcribed as a double vowel and therefore did not match the other forms.

To appreciate the differences in discourse style between IDS and ADS, the model was applied to a segment of spontaneous adult-direct speech produced by a Swedish politician during a radio interview (Table 5). With a memory span of 20 s and no memory biases, the model picked up only one possible lexical candidate based on the recurrence principle: *eh*, used by the speaker to fill in pauses in her speech.

3.5. A tentative comparison of pause durations in IDS and ADS

Finally, to investigate the general rhythmic characteristics of the speech signal reaching the infant, the length of the pauses between the utterances was also measured. The study of pauses is a frequent theme in discourse analysis and the present results are no more than just an example of what can be observed in infant-directed speech. A proper analysis of the significance and distribution of pauses requires the consideration of a number of factors that fall outside the scope of the present chapter. Thus, the data presented here must be seen as an indication of how pauses in IDS towards a 3-month-

Table 3. Lexical candidates derived by the model under assumption of 20 sec. memory span and with no primacy or recency biases.

Infant's age: 2.97 months
Number of lexical candidates found in "mamma" utterances: 18
Assumed memory span: 20 s and no memory bias.
Total recorded duration: 974.449 sec.
Effective total utterance time: 651.407 sec.
Lexical items' rate: 0.01847 candidates/sec. (rel. total recorded duration)
Effective lexical items' rate: 0.02763 candidates/sec. (rel. effective utterance time)

Items (no primacy or recency effects):	Repetitions within span
ja	10
jaa	8
mmm	7
va	6
mhmhmhmhmhm	4
hmhm	4
titta	4
aa	4
jaha	2
vetduva	2
haa	2
mhmhmhmhmhmhmhmhm	1
här	1
aaa	1
hääeh	1
ärdufärdig	1
intefärdig	1
kucka	1

Table 4. Lexical candidates derived by the model under assumption of 20 sec. memory span and with recency bias.

Infant's age: 2.97 months
Number of lexical candidates found in "mamma" utterances: 14
Assumed memory span: 20 sec. and memory recency.
Total recorded duration: 974.449 sec.
Effective total utterance time: 651.407 sec.
Lexical items' rate: 0.01437 candidates/sec. (rel. total recorded duration)
Effective lexical items' rate: 0.02149 candidates/sec. (rel. effective utterance time)

Items (memory recency)	Repetitions within span
mmm	7
jaa	6
mhmhmhmhmhm	3
hmhm	3
va	3
titta	3
aa	2
här	1
aaa	1
hääeh	1
ärdufärdig	1
intefärdig	1
vetduva	1
haa	1

Table 5. Lexical candidates derived by the model under assumption of 20 sec. memory span and with no primacy or recency biases

Adult speaker
Number of lexical candidates found in "SwePolitician" utterances: 1
Assumed memory span: 20 sec.
Total recorded duration: 221.77 sec.
Effective total utterance time: 80.705 sec.
Lexical items' rate: 0.00451 candidates/sec. (rel. total recorded duration)
Effective lexical items' rate: 0.01239 candidates/sec. (rel. effective utterance time)

Items (no primacy or recency effects)	Repetitions within span
eh	9

old infant may look like. Factors like the speaker's communicative intention, speaking style, rhetoric aspects, cultural- and situation-dependent strategies, etc. will not be addressed here at all. The distribution of the pause durations is shown in Figure 4.

These pauses spread over a wide range of durations, compatible with the notion that the mother may be encouraging the infant to respond to her utterances. The distribution is heavily skewed and dominated by short duration pauses, with durations up to about 5 seconds. To have a closer look at pauses within this range, an expanded version of the pause distributions up to 5 s is shown in Figure 5.

As the histogram in Figure 5 shows, the typical pauses between utterances are about 1 s long. Although this pause length appears to suggest a rather intensive rate of utterances directed to the infant during the 16 minutes of this recorded interaction, it should be noted that the duration of these IDS pauses is indeed much more variable (and generally longer) than what was found in the sample of ADS speech produced in the context of the radio interview (Figure 6). There may be several reasons for this but one possibility is that the speaker's goal during the radio interview appears to be conveying information. There is an implicit assumption that the listener shares a good deal of life experiences with the speaker and there is no opportunity to check whether the listener has got the message or not. It is in fact broadcasting, and although the interviewer in the studio plays the role of the listener, both the speaker and the interviewer are aware that interruptions in their speech flow may have a different impact on the listener, who can only access the audio signal, than on the interlocutor in the studio, who obviously can see and interpret the facial expressions of the speaker. Pauses in this context are probably motivated by stylistic and rhetoric factors, along with the need to inhalate to maintain the speech flow, thus contrasting with the previous mother-child interaction. In IDS, the listener is not expected to understand much of the utterances' information content nor to have considerable life experience to establish a common ground for the interaction. Thus the pauses in IDS may be determined by the speaker's need to observe the infant and maintain the infant's interest on the communication by adjusting the utterance rate to the infant's responses. Future work will attempt to integrate in ETLA a component to deal with the potential significance that such pauses may have for early language acquistion.

3.6. Empirical support for the emergent lexicon model

The example presented above shows that it is in principle possible to derive plausible lexical candidates from the relatively short episodes of IDS using a model that does not have any initial linguistic knowledge. At this initial stage of the language acquisition process, the emergence of the lexical candidates can be accounted for by the interaction between typical IDS towards young infants, short-term memory representations and a similarity distance metric. But can this simple model predict empirical results from infant speech perception experiments? In fact it can. An experiment in which 8–10- and 14–16-month-old Swedish infants listened to declarative utterances, produced in unfamiliar or constructed languages, while being shown puppets that the utterances referred to, showed that the infants did pick up the relevant referential function conveyed by the audio-visual situation, after only 2 minutes of exposure (Lacerda et al.,

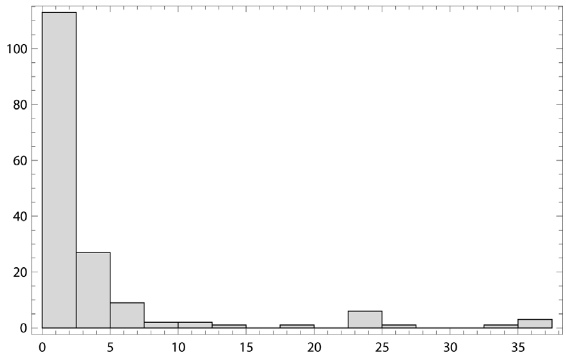


Figure 4. Distribution of pause durations (time in seconds) during the mother's IDS.

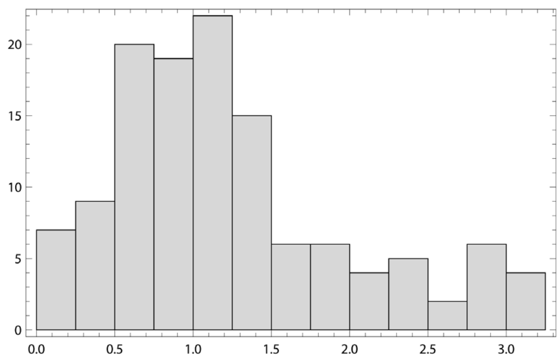


Figure 5. Distribution of the short pauses (up to 5 seconds) in IDS.

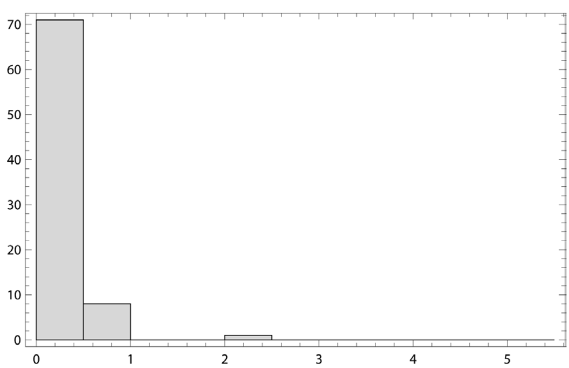


Figure 6. Distributions of pause durations up to 5 seconds. during a Swedish politician's radio interview.

forthcoming). The speech materials in these studies had target words in utterance final position. To address the influence of target word placement in the utterance and possible interactions with prosodic focal accent, Gustavsson et al. (2004) conducted a study using 3-word utterances produced in an artificial language and where the three possible positions of the target word were systematically combined with the three possible positions of the focal accent. Her results indicated that while placing the target word in stressed utterance-final position induced the most robust responses to the intended target, the next strongest response was conveyed by the placement of focus on the first (irrelevant for the inference of the referential function) word of utterances where the target word was in utterance-final position. Taken together these results suggest that, at least by 8 months of age, infants are able to pick up the linguistic referential function from the co-occurrence of utterances and possible referents. They also suggest that the infants are sensitive to prosodic variables that presumably draw the infant's attention to the target word – either directly or by perceptually capturing the infant's attention to the utterance. Memory recency and primacy effects seem to underline the process. The plausibility of the model of emergent lexicon proposed here is further enhanced by empirical results from other research groups (Jusczyk & Aslin, 1995; Mandel & Jusczyk, 1996; Mandel et al., 1996; Mandel et al., 1995; Mandel et al., 1994; Ganger & Brent, 2004; Brent & Siskind, 2001; Brent, 1999; Bahrick et al., 2002; Gogate et al., 2000). To be sure, IDS is not a homogeneous type of speech, partly because adults vary in their sensitivity to the infant's developing communicative needs and cognitive level. Indeed, the highly repetitive pattern of speech towards young infants tends to change towards a more adult-like, information-oriented communication style, as the infant shows signs of increased cognitive maturity.

4. Closing the language acquisition loop

The model of early language acquisition proposed here relies on three main factors: the repetitive characteristics of infant-directed speech, the infant's capacity to store auditory representations, and a similarity metric to find recurrent patterns among the auditory representations. The characteristics of IDS are relatively assessable from observations of adult–infant interactions under different communicative goals. For instance, the adult's infant-directed speech during the introduction of a new toy, interpretations of the current situations, reference to the infant's earlier experiences, etc., is typically repetitive. Regarding the infant's memory capacities, it has been demonstrated that by 8 months of age, infants who are exposed to words indicate that infants can remember words over a period of at least two weeks (Jusczyk & Hohne, 1997). Remembering words is by itself a manifestation of the infant's ability to detect similarity between the acoustic patterns of heard words and the auditory representations stored in memory; the similarity metric can be viewed as the extent to which two auditory representations overlap. Nevertheless such perceptual and representational capabilities have to be integrated in a broader model in order to achieve a more complete picture of the language acquisition process. To be sure, at least one other sensory component has to be considered in the model as well as the infant's production capabilities. The main implications of these two additional aspects will be highlighted in the following subsections.

4.1. Multi-sensory aspects of the early language acquisition process

A fundamental property of spoken languages is the use of sounds as symbols to refer to objects and events in the humans' ecological settings. According to ETLA, the infant's ability to retrieve the underlying referential principles of their ambient language is initially conveyed by the adult's recurrent use of words or expressions while interacting with the infant. From the purely auditory perspective, such a recurrent use of sound strings creates a situation that is closely related to what is usually meant by "statistical learning" (Saffran et al., 1996) but the adult–infant interaction in typical ecological settings introduces an important additional component to the "pure statistical learning", since the adult implicitly demonstrates the spoken language's referential function. Even though during adult–infant interaction the speech produced by the adult may actually refer to objects or events not available at the interaction site, focusing on the infant's immediate field of attention is highly probable. The adult tends to monitor their focus of interest and spontaneously adapts her speech in response to her perception of what might have triggered the infant's curiosity or what might be the infant's information needs for the moment (Murray & Trevarthen, 1986; Nelson, 1988). As stated above, from the infant's perspective, a narrator voice is added to the scene, speaking about whatever might be in the infant's field of attention or filling in comments or interpretations of the infant's state of mind, as interpreted by the adult. Under these circumstances, the implicit linguistic referential function emerges when the infant matches auditory representations of the adult's recurrent utterances with information simultaneously available through other sensory modalities, irrespective of the specific situational details of the adult–infant interaction. In other words, the recurrent acoustic patterns selected by the memory processes suggested in the model above tend to gain linguistic referential value as they are linked to relatively time-locked recurrent patterns available in other modalities. In this sense, meaning emerges from the relationship between the recurrent auditory representations and the correlated recurrent representations in at least one other sensory modality.

A direct implication of this view is that the early stages of the acquisition of the spoken language demand access to at least two sensory modalities: one, the auditory modality, must be present for representation of the acoustic forms of the lexical items; another sensory modality, a visual, olfactory, gustative, or tactile representation of the object or situation being referred to, must be available to be linked to the acoustic form. For instance, adding a visual component to the model's acoustic dimension and assuming that the dimensionality of the visual space is of about the same order of magnitude as that of the auditory space, leads to an immediate combinatorial explosion with important theoretical implications. The number of possible combinations of audio and visual stimuli is so large ($10^{20} * 10^{20} = 10^{40}$; continuing the previous example, this corresponds to the number of cells in the bodies of 10^{17} times the Earth's current population) that the random re-occurrence of a given specific combination of audio and visual stimuli is extremely improbable (10^{-40}). That single repetition can be taken as a highly significant indication that those particular auditory and visual stimuli must somehow be linked to each other. Given the number of audio-visual repetitions occurring in typical adult–infant interaction settings, the emergence of a referential linguistic function is almost inevitable as long as memory representations of these modalities last long enough for the comparison between the occurrences to be made.

It should be noted that this kind of initial association-induced referential function is neither the only nor even the most productive component of the language acquisition process as the child matures. Indeed, there is experimental evidence that importance of the recurrent occurrence of correlated multisensory stimuli is drastically reduced as the child develops what is often named a "theory of mind" (Tomasello & Farrar, 1986; Bloom, 2000). As the child achieves a better understanding of others' intentions (Behne et al., 2005) a more interpretative approach to the referential function starts to emerge. At this point, the child appears to start deriving the names of objects and events not from the immediate contingency between the sounds of the utterances and the items available in the neighbourhood but from what the child assumes to be the other person's labelling intentions (Tomasello & Farrar, 1986; Tomasello & Barton, 1994; Tomasello et al., 1996; Tomasello et al., 1997; Moll & Tomasello, 2004).

4.2. Production and action aspects

The infant's ability to produce vocalizations and spontaneous motor actions are two further aspects that must be considered in the context of the early language acquisition process. These aspects are important because they introduce a necessary link between the language learner and the speakers in the ambient language, creating opportunities for effective negotiation of the utterances' referential value. The adult's and the infant's responses to each other's utterances create opportunities for establishing and verifying equivalences between their acoustically different utterances which are created as the utterances are applied to commonly available external objects or events (Tomasello et al., 2005). At the same time, through the iterative application of its own utterances to refer to such objects or events the infant also exposes their hypotheses and beliefs about the environment or relations among things, opening for validation or correction of those beliefs.

An important consequence of referring to external referents using own vocalizations is that the infant tests and establishes equivalences between acoustically very different utterances and the heard adult utterances applied to the same referents. For the proposed ETLA model, where the acoustic form of the utterances is used to account for the initial emergence of potential lexical items, the possibility of establishing acoustic equivalence classes via sound-object links to external referents is an important and productive aspect.

Indeed, matching spectral patterns of adult utterances with spectral patterns from infants' utterances is not trivial. Because the ratio of pharynx length to vocal tract length is much lower in the infant than in the adult, sounds articulated at the same place of articulation by the infant and the adult will have different spectral characteristics. Conversely, when the sounds are spectrally similar, the places of articulation used by the infant and the adult are almost certainly different (Lacerda, 2003; Serkhane et al., 2007). This is a logical consequence of the non-linear scaling between the infant's and the adult's vocal tract and of the impact that the vocal tract's geometry has on its acoustic properties (Fant, 1960; Stevens, 1998). However in early language acquisition, it is probably reasonable to assume that the match between the infant and the adult will be acoustically rather than articulatory determined (Kent & Murray, 1982; Menard et al., 2004). From the ETLA's perspective, the infant's initial vocalizations, although essentially random, will enable the infant to establish acoustic-articulatory maps where biases towards "successful" vocalizations become increasingly stronger as the infant interacts with speakers of the ambient language. This process of convergence towards adequate acoustic-articulatory maps has been proposed in models like the "DIVA model" (Guenther, 1994), even though DIVA makes a stronger assumption than ETLA by postulating discrete speech sounds at the onset of the process. Recent results from experiments carried out at the Stockholm University's Phonetics Laboratory, where the phonetic variables influencing adult judgments of the similarity between infant and adult utterances were investigated (Gustavsson, forthcoming), shed some light on the mechanisms underlying the infant's convergence towards linguistically meaningful acoustic-articulatory maps. As suggested by ETLA, the adult judgments of similarity between the utterances produced by an infant and an adult model seem to be influenced by the judge's perception of the infant's developmental state. The results indicate a hierarchical structure in the impact of phonetic deviations in the infant's utterances as a function of the perceived age of the infant. If the infant is perceived as being young, then infant utterances matching the adult's in terms of general syllabic structure and timing tend to be accepted as successful matches of the adult model; when the infant is perceived to be older, further accuracy at the segmental level appears to be necessary in order for the infant utterances to be judged as successful matches of the adult's. The overall picture conveyed by Gustavsson's results appears to be compatible with the notion of ecologically and dynamically adjusted early language acquisition process positing that:

1. The adult in the infant's ecological context explores the infant's focus of attention and the affordances of the objects or events available in the infant's neighbourhood.
2. The adult provides the infant with speech input that is finely tuned to be correlated with whatever is perceived to be in the infant's current focus of attention.
3. The adult interprets the utterances produced by the infant in a developmentally adjusted fashion; this interpretation may effectively guide the infant into successfully converging toward the ambient language's phonetic targets.

Addressing the process of early language acquisition by creating an operational model of how such components may interact and trigger further language development is one of the research challenges in which the research team at Stockholm University's Phonetics Laboratory is currently engaged.

5. Acknowledgements

This research was conducted within the framework of MILLE (The Bank of Sweden Tercentenary Foundation, grant K2003:0867), and CONTACT (EU-NEST project n. 5010) as well as of the Knut and Alice Wallenberg Foundation (Grant no. KAW 2005.0115).

References

Aslin, R. N., Saffran, J. R. & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324.
Bahrick, L. E., Gogate, L. J. & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: The salience of actions over faces in dynamic events. *Child Development, 73,* 1629–1643.
Beckman, M. E. & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development, 71,* 240–249.
Behne, T., Carpenter, M. & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8,* 492–499.
Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.
Brent, M. (1999). Speech segmentation and word discovery: A computational perspective. *Report from Johns Hopkins University*.
Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33-B44.
Chugani, H. T., Phelps, M. E. & Mazziotta, J. C. (1987). Positron Emission Tomography study of human brain functional development. *Annals of Neurology, 22,* 487–497.
de Casper, A. & Fifer, W. P. (1980). On human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174–1176.
de Casper, A. & Spence, M. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9,* 133–150.
Fant, G. (1960). *Acoustic theory of speech production*. The Hague, Netherlands: Mouton.
Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60,* 1497–1510.
Fernald, A. & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209–221.
Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104–113.
Ganger, J. & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology, 40,* 621–632.
Gogate, L. J., Bahrick, L. E. & Watson, J. D. (2000a). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71,* 878–894.
Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernics, 72*(1), 43–53.

Gustavsson, L., Sundberg, U., Klintfors, E., Marklund, E., Lagerkvist, L. & Lacerda, F. (2004). Integration of audio-visual information in 8-month-old infants. In L. Berthouze, H. Kozima, C. Prince, G. Sandini, G., Stojanov, G. Metta & C. Balkenius (Eds.), *Proceedings of the Fourth International Workshop on Epigenetic Robotics (EPIROB 2004)*, *Lund University Cognitive Studies*, 117 (pp. 143–144).

Jusczyk, P. & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1–23.

Jusczyk, P. & Hohne, E. (1997). Infants' memory for spoken words. *Science, 277,* 1984–1986.

Kent, R.D. & Murray, A.D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America, 72,* 353–365.

Lacerda, F. (2003). Phonology: An emergent consequence of memory constraints and sensory input. *Reading and Writing: An Interdisciplinary Journal, 16,* 41–59.

Lacerda, F. & Sundberg, U. (2006). An ecological theory of language acquisition. *Linguística, 1,* 53–106.

Lacerda, F., Sundberg, U., Klintfors, E. & Gustavsson, L. (2009). Infants learn nouns from audio-visual contingencies. Manuscript.

Lecanuet, J.P., Granier-Deferre, C., DeCasper, A.J., Maugeais, R., Andrieu, A.J. & Busnel, M.C. (1987). [Fetal perception and discrimination of speech stimuli; demonstration by cardiac reactivity; preliminary results] (in French). *Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie, 305*, 161–164.

Lenneberg, E.H. (1967). *Biological Foundations of Language*. New York: Wiley.

Lenneberg, E.H., Rebelsky, F.G. & Nichols, I.A. (1965). The vocalizations of infants born to deaf and to hearing parents. *Human Development, 8,* 23–37.

MacWhinney, B. (1999). Commentary on Rispoli's review: Rethinking innateness. *Journal of Child Language, 26,* 237–239.

Mandel, D. & Jusczyk, P. (1996). When do infants respond to their names in sentences? *Infant Behavior and Development, 19,* 598.

Mandel, D., Jusczyk, P. & Kemler Nelson, D.G. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition, 53,* 155–180.

Mandel, D., Jusczyk, P. & Pisoni, D.B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6,* 314–317.

Mandel, D., Kemler Nelson, D.G. & Jusczyk, P. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development, 11,* 181–196.

Menard, L., Schwartz, J.L. & Boe, L.J. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *Journal of Speech Language and Hearing Research, 47,* 1059–1080.

Moll, H. & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science, 7,* F1-F9.

Moon, C. & Fifer, W.P. (1990). Syllables as signals for 2-day-old infants. *Infant Behavior and Development, 13,* 377–390.

Murphy, K.P. & Smyth, C.N. (1962). Response of foetus to auditory stimulation. *Lancet, 5,* 972–973.

Murray, L. & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of Child Language, 13,* 15–29.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 756–766.

Nelson, K. (1988). Constraints on word learning? *Cognitive Development, 3,* 221–246.

Pinker, S. (1994). *The language instinct: How the mind creates language* (1 ed.). New York: William Morrow and Company, Inc.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. London: Penguin Books.

Querleu, D., Renard, K., Versyp, F., Paris-Delrue, L. & Crepin, G. (1988). Fetal hearing. *European Journal of Obstetrics, Gynecology and Reproductive Biology, 29,* 191–212.

Roug, L., Landberg, I. & Lundberg, L. J. (1989). Phonetic development in early infancy: A study of four Swedish children during the first eighteen months of life. *Journal of Child Language, 16,* 19–40.

Saffran, J. R., Aslin, R. N. & Newport, E. (1996). Statistical learning by 8-month old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Senghas, A. & Trueswell, J. C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences USA, 98,* 12874–12875.

Saffran, J. R. & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39,* 484–494.

Serkhane, J. E., Schwartz, J. L., Boe, L. J., Davis, B. L. & Matyear, C. L. (2007). Infants' vocalizations analyzed with an articulatory model: A preliminary report. *Journal of Phonetics, 35,* 321–340.

Snow, C. E., Arlmanrupp, A., Hassing, Y., Jobse, J., Joosten, J. & Vorster, J. (1976). Mothers speech in 3 social-classes. *Journal of Psycholinguistic Research, 5,* 1–20.

Spence, M. J. & de Casper, A. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. *Infant Behavior and Development, 10,* 133–142.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, Mass.: MIT Press.

Sundberg, U. (1998). *Mother tongue – phonetic aspects of infant-directed speech.* Stockholm University.

Thiessen, E. D., Hill, E. & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*, 53–71.

Tomasello, M., Akhtar, N., Dodson, K. & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language, 24,* 373–387.

Tomasello, M. & Barton, Michelle (1994). Learning words in nonostensive contexts. *Developmental Psychology, 30,* 639–650.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences, 28,* 675–691.

Tomasello, M. & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57,* 1454–1463.

Tomasello, M., Strosberg, R. & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language, 23,* 157–176.

van de Weijer, J. (1999). *Language input for word discovery.* MPI Series in Psycholinguistics.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Cambridge, Mass.: Basil Blackwell.

Vihman, M. M. & McCune, L. (1994). When is a word a word? *Journal of Child Language, 21,* 517–542.

Vince, M. A., Billing, A. E., Baldwin, B. A., Toner, J. N. & Weller, C. (1985). Maternal vocalizations and other sounds in the fetal lamb's sound environment. *Early Human Development, 11,* 179–190.

Walker, D., Grimwade, J. & Wood, C. (1971). Intrauterine noise: A component of the fetal environment. *American Journal of Obstetrics and Gynecology, 109,* 91–95.

Cerebral activation patterns of speech perception in 4-month-olds and adults

Iris-Corinna Schwarz, Francisco Lacerda,
Heléne Walberg, Ellen Marklund
Department of Linguistics, Stockholm University

Abstract

Speech perception in infancy is language-general, but language-specific in adulthood. Is this reflected in Event-Related Potentials (ERPs)?

ERPs from 128 electrodes were recorded for three speech perception conditions in 4-month-old Swedish infants (N = 9) and adults (N = 13). The stimuli were simple sentences in infant-directed speech (IDS) in Swedish, familiar in semantic content and prosodic characteristics; in spectrally rotated Swedish with incomprehensible semantic content, but familiar prosodic characteristics; and in Portuguese, unfamiliar in both aspects.

As predicted, infants show bilateral activation across conditions, but also present two left-hemispheric advantages as potential first signs of native language specialisation. In adults, left-hemispheric activation shows condition-sensitive ERP patterns: processing rotated Swedish is similar to Swedish, but different to Portuguese. The right-hemispheric dominance in adults is explained by the strong prosodic and weak semantic characteristics of the IDS stimuli. Infants' language-general and adults' language-specific speech perception is reflected in the ERPs, although 4-month-olds show first signs of specialisation.

1. Introduction

Infant and adult speech perception differ considerably. Young infants' language skills are language-general, whereas adults have become specialised in their native tongue through development and present language-specific abilities as an outcome of that.

On the phonemic language level, infants are born as "citizens of the world" (Kuhl, 2004), equipped with global speech perception skills which enable them to discriminate minimal contrasts of every language they have been tested on so far (Polka & Werker, 1994). Adults on the other hand perceive speech contrasts language-specifically: their ability to discriminate non-native speech contrasts is greatly diminished and can only partially be retrained (Tees & Werker, 1984; Werker & Tees, 1999, 2002).

Differences in infant and adult speech perception continue on the suprasegmental language level: infants younger than six months are equally sensitive to all stress patterns, not only to the ones predominantly present in their native language to which older infants listen significantly longer (Jusczyk et al., 1993). In contrast, syllable discrimination on the basis of prosody has been shown in 2- to 3-month-olds (Karzon & Nicholas, 1989), demonstrating the salience of prosodic features early in life.

Are these differences between infant and adult speech perception represented in the brain and how can they be measured? In the case of phoneme discrimination, similarities in behavioural and ERP data from infants and adults confirm a common neural basis for phoneme processing from infancy to adulthood (Dehaene-Lambertz & Gliga, 2004). However, the natural speech signal encompasses prosodic aspects besides the phonemic level. Could cortical representation be an indication for the developmental differences in speech perception between early infancy and adulthood?

Lateralisation of potentially language-relevant hemispheric functions is widely accepted and well established. The left hemisphere is typically seen as dominant in language processing (e.g. Knecht et al., 2000; Ojemann et al., 1989), in memory encoding (e.g. Gabrieli et al., 1998), in categorisation (e.g. Gerlach et al., 2000; Seger et al., 2000), in analytic and logical skills (e.g. Atherton et al., 2003), and in numerical processing (e.g. Chochon et al., 1999). The right hemisphere is more active in tasks involving creativity (e.g. Carlsson et al., 2000; Weinstein & Graves, 2002), emotions (e.g. Canli et al., 1998), memory recall (e.g. Marsolek et al., 1994; Rypma & D'Esposito, 1999), and musical patterns (e.g. Nicholson et al., 2003; Platel et al., 1997).

Geschwind and Levitsky's seminal paper (1968) puts forth anatomical hemispheric differences in the temporal lobes. In particular, the left planum temporale, part of the temporal speech cortex, is in the majority of cases on average 1/3 longer than on the right (Geschwind & Levitsky, 1968). The planum temporale processes acoustically complex sounds including, but not restricted to, speech sounds (Obleser et al., 2007).

There are, however, differences between the processing of speech sounds and other complex signals. The anterolateral superior part of the left temporal cortex shows stronger activation to consonants than to spectrally matched nonspeech sounds, measured by monitoring blood oxygen flow with functional magnetic resonance imaging (fMRI) (Obleser et al., 2007). The processing of vowels compared to tones is also left-hemispheric dominant, as measured by magnetoencephalography (MEG) (Gootjes et al., 1999).

Cerebral activation in response to natural speech can be differentiated into effects of human voice perception which activate the right hemisphere in right-handed adults, and lexical-semantic effects which are processed in the left hemisphere (Koeda et al., 2006). When contrasting the activation of sentences, reverse sentences and nonvocal identifiable sounds, the right-hemispheric response due to human voice perception could mask the lexical-semantic processing activation in the left hemisphere, as measured by fMRI (Koeda et al., 2006).

What does the picture look like for infants? There are only few infant brain-imaging studies on speech perception, involving human infants and monkeys.

Similar to humans, adult rhesus monkeys display left-hemispheric lateralisation when processing vocalisations, while infant monkeys show no such asymmetry (Hauser & Andersson, 1994). The infant monkeys respond with bilateral activation to the signal.

Hemodynamic responses to intonation in normal speech and in a flattened version with removed pitch contour alteration indicate where the infant brain processes prosody in speech stimuli (Homae et al., 2006). Besides bilateral activation in 3-month-olds, the temporoparietal region in the right hemisphere was sensitive to the prosodic differences.

Cerebral activation in response to mismatches in auditory nonspeech signals (sine wave tones) and in syllables in 4-month-olds indicates the involvement of different neural networks within the temporal lobe, but does not yield an asymmetry with a left-hemispheric advantage for phonemic changes compared to tone variation (Dehaene-Lambertz, 2000).

Both studies, however, obtained their results with adult-directed speech, although infants typically prefer infant-directed speech (IDS) over adult-directed speech (Fernald, 1985, 1992). IDS is characterised by an elevated fundamental frequency with a broad range, as well as hyperarticulation of vowels and strong affect in the parental voice (Burnham et al., 2002; Kuhl et al., 1997; Sundberg, 1998). Infants at four months prefer IDS stimuli and even sine wave analogs to the fundamental frequency of IDS which demonstrates the salience of exaggerated pitch contours in infant perception (Panneton Cooper & Aslin, 1994). In response to IDS-stimuli, 6-month-old infants show higher neural activation than to adult-directed speech (Zangl & Mills, 2007). Thus, only IDS stimuli were used in the current study to optimally attract infant interest.

The study contrasts the cerebral activation patterns of speech perception at two developmental extremes: early infancy and adulthood. Speech perception lateralisation in the native language is not predicted for 4-month-olds, but for adults, exhibiting a left-hemispheric bias. The three stimulus conditions are expected to affect activation patterns differentially due to their variation of semantic and prosodic aspects in relation to the participants' native language. For infants, native Swedish, rotated Swedish, and Portuguese should all result in bilateral cerebral activation, as 4-month-olds' speech perception is not yet specialised in the native language; adults, however, should display left-hemispheric lateralisation for Swedish and bilateral activation for rotated Swedish and for Portuguese, since with decreasing comprehension level, the prosodic quality of the IDS-stimuli is expected to activate the right hemisphere more strongly.

2. Method

2.1. Participants

There were two groups of participants, 4-month-old Swedish-learning infants (*N* = 9, 7 female, 2 male; age range: 108–129 days) and adult native speakers of Swedish (*N* = 13, 7 female, 6 male; mean age: 26.3 years; age range: 21–40 years), all except one right-handed. The infant retention rate was 70%, as four of 13 originally tested infants dropped out due to fussiness, whereas only one adult's data was excluded due to a too high level of noise in the signal.

Adult participants and infants' parents answered a brief questionnaire on relevant aspects related to speech perception performance. For infants, we asked for information about date of birth, languages spoken at home, and history of otitis media. Eight infants grew up in a Swedish-only environment; only one infant had a bilingual home (Swedish and Greek). The adult participants reported which languages they spoke, understood, their parents' mother tongue, and any musical training they engaged in. The criterion of at least 3 years of regular musical training divided them into a group with musical experience (*N* = 5) and a musically naïve group (*N* = 8). Known languages other than Swedish were English, Spanish, German, Japanese, Chinese, and French, but none of the participants were familiar with Portuguese. Since 2007, newborn hearing tests are standard in Sweden, therefore normal infant hearing could be assumed. All adult participants reported normal hearing.

2.2. Stimuli

All three stimulus conditions were recorded in IDS; the stimuli comprised short, simple sentences in Swedish of an average length of 1 s, rotated Swedish, and lexically equivalent Portuguese utterances. To the listener, Swedish stimuli present familiar semantic content and familiar prosodic features; rotated Swedish stimuli are semantically incomprehensible, yet contain similar prosodic characteristics; and Portuguese stimuli have both inaccessible semantic content and unfamiliar prosody.

The Swedish stimuli consisted of 12 different recorded infant-directed sentences produced by two female native speakers of Swedish, whereas the six Portuguese infant-directed utterances were recorded with a female native speaker of Portuguese. All speakers had experience in IDS modulation, recording situations, and in the use of microphones.

Single sentence duration ranged from 0.8 to 1.8 s. The stimuli were recorded in an anechoic chamber with a ½-inch Brüel & Kjær free-field condenser microphone (4190) and a Brüel & Kjær amplifier (2209), and saved directly to a hard drive (Creative Audigy 2ZS sound card). They were cut (WaveSurfer 1.8.5), and manually matched for loudness (Cool Edit 2000). A second Swedish stimulus set, in which the target word *mamma* had been replaced by the Swedish nonword *kocka*, was spectrally inverted for the rotated version (Blesser, 1972). This modification was needed to ensure the rotated speech stimuli could not be deciphered, as *mamma* remained intelligible even after rotation.

Blesser's rotation method inverts the spectrum vertically without affecting the horizontal axis, and changes low frequencies to high, and high frequencies to low, while preserving intonation and temporal aspects fully. Since the cut-off frequency in Blesser's original program was set at telephone band level (200–3000 Hz), some of the information present in the speech signal was lost during rotation, rendering some cues, for example fricatives, ambiguous.

Our adaptation uses a 4 kHz upper cut-off frequency, and as such, more information in higher frequencies than with Blesser's cut-off frequency remains in the rotated signal. Speech rotation is useful since only spectral acoustic cues are affected by the rotation which means that all non-spectral cues, such as pauses, fricative noise and voicing, remain recognizable by listeners. Rotated Swedish therefore maintains Swedish prosody.

2.3. Apparatus

To measure stimulus-related brain activation, EEG-head-nets (Electrical Geodesics Inc.) with 128 sensor channels were used, size selected according to head circumference. The experiment was programmed and run using the software E-prime 1.2 and Net Station 4.2. Net Station tools cleaned the EEG data with a band pass filter from 0.3–30 Hz noise, segmented the stimulus responses into 360 segments of 1 s each, and removed unusable channels and segments (e.g. eye blinks) before collating segment averages across stimuli conditions for each participant, and referencing the EEG-voltage measurements to a baseline prior to stimuli onset.

2.4. Procedure

In a sound-attenuated room with dimmed lights, the EEG head-net was placed on the participants' heads. Adults were instructed to simply listen to the stimuli presented via loudspeaker and focus on a fixation mark in the middle of a white screen while sitting motionless in a chair. They were advised to reduce blinking as much as possible. Infants were seated on their parent's lap, kept as still as possible. Breastfeeding, looking at the parent's face or at a toy were legitimate ways of keeping the infant still. The session was aborted in case of continuous infant fussiness.

The stimuli were presented in 10 blocks with 36 randomised trials each; each of the six exemplars per stimulus condition occurring twice per block (SPL ranged from 55 to 65 dB). The interstimulus interval was 500 ms. Each session took approximately 18 minutes for adults, 8 to 18 minutes for infants.

3. Results

The topographic view of average brain activation gives a good first impression of where and to which intensity infants and adults perceive the speech stimuli (Figure 1). The selected times of 300 ms, 350 ms, and 400 ms after stimulus onset are relevant points of language processing (Federmeier et al., 2002; Hoen & Dominey, 2000).

Infants show high general activation with a frontal-parietal tendency. At 350 ms, the similarity between Swedish and rotated Swedish activation patterns appears closer than to Portuguese in infants, but not in adults. No palpable unilateral activation in infants arises at any time and condition, whereas adults seem to show lateralisation for Portuguese and rotated Swedish. Activation for Swedish and rotated Swedish in adults has ceased at 400 ms while it is still present for Portuguese.

To contrast left- and right-hemispheric activation, the head-net electrodes were divided into areas corresponding to frontal, temporoparietal, and occipital lobe (Figure 2). Only frontal and temporoparietal areas were considered language-related (Imada et al., 2006) and therefore included in the following analysis.

The study's exploratory character did not let us settle for the common significance tests between specific points in time only. Therefore the ERP responses for the complete signal length were plotted for the frontal and the temporoparietal lobes (see Figures 3 and 4). While the graphs show the mean activation curve per lobe, standard error confidence intervals specify the presence of a significant difference between the hemispheres.

Looking at the adults' frontal ERP graphs, right-hemisphere dominant lateralisation is significant between 300 and 500 ms in the rotated Swedish (at 316 ms: $M_{rotLF} = 0.4$ µV; $SD_{rotLF} = 1.0$; $M_{rotRF} = 1.1$ µV; $SD_{rotRF} = 1.1$; $p < .05$) and between 200 and 400 ms in the Portuguese condition (at 400 ms: $M_{portLF} = -0.2$ µV; $SD_{portLF} = 1.0$; $M_{portRF} = 0.6$ µV; $SD_{portRF} = 1.1$; $p < .05$). It is not surprising to see no differences between left and right hemisphere activation in the Swedish condition, as the reported differences in rotated Swedish and Portuguese seem to exist because of a reduced left-hemispheric activation level compared to a maintained right-hemispheric activation level across the conditions. The left-hemispheric activation graph describes the course of native Swedish speech perception with a modest N300. This is repeated, although less defined, but with a similar activation level, in the perception of rotated Swedish, indicating that the rotated stimuli in fact match the original in perceptual response. Portuguese perception seems to differ both in activation level and in the course of the graph, reflecting the semantic and phonotactic differences of the stimulus conditions.

Infants' frontal ERPs of Swedish, rotated Swedish, and Portuguese speech are quite similar. As already indicated by the topographic view, besides a significant pre-stimulus left-hemispheric advantage for rotated Swedish, there is only one other speech condition that presents left-hemispheric dominance: rotated Swedish. As the first difference begins prior to stimulus onset, it is possible that it is an artefact although all stimuli were randomised. However, the second left-hemispheric advantage between 450 and 600 ms in the rotated Swedish condition is significant (at 500 ms: $M_{rotLF} = 6.4$ µV; $SD_{rotLF} = 2.5$; $M_{rotRF} = 4.7$ µV; $SD_{rotRF} = 1.7$; $p < .05$). The familiar prosody of rotated Swedish elicits a higher left-hemispheric activation, possibly reflecting a semantic decoding attempt. The fact that greater stimulus familiarity results in higher activation is further substantiated by the fact that the overall activation level for Swedish and rotated Swedish appears higher than for Portuguese. The infant brain at four months seems to respond more intensely to the familiar Swedish and to its semantically deprived rotated form than to unfamiliar Portuguese.

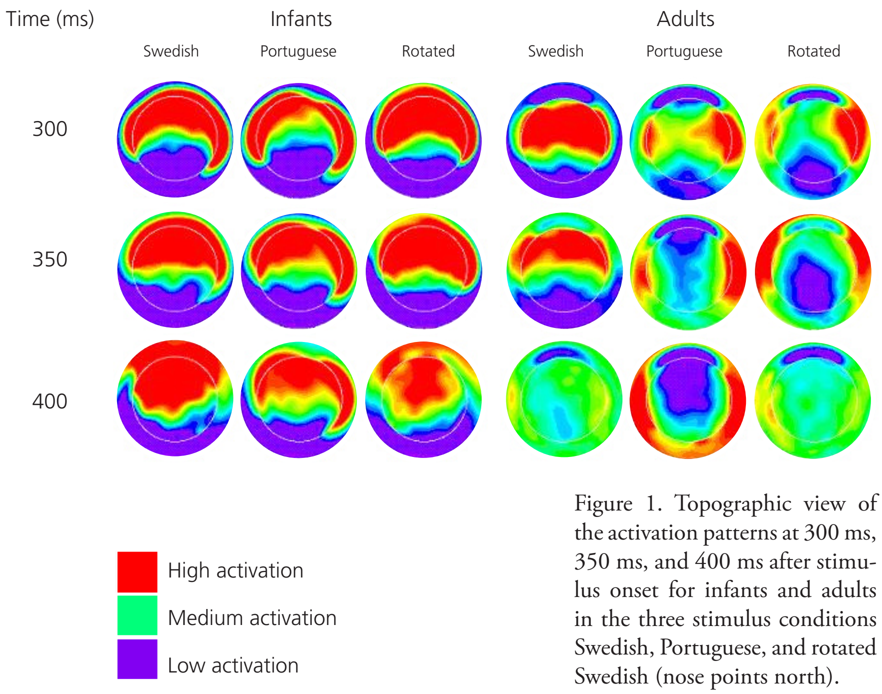


Figure 1. Topographic view of the activation patterns at 300 ms, 350 ms, and 400 ms after stimulus onset for infants and adults in the three stimulus conditions Swedish, Portuguese, and rotated Swedish (nose points north).

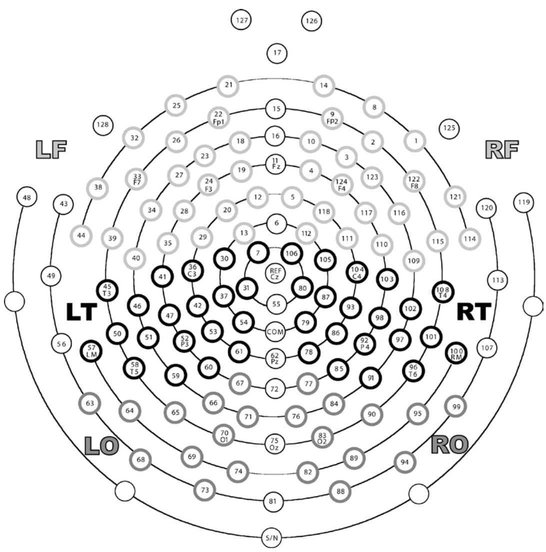


Figure 2. Division of the head-net into frontal, temporoparietal, and occipital lobe per hemisphere (nose points north): left- and right-frontal (LF & RF), left- and right-temporoparietal (LT & RT), left- and right-occipital (LO & RO). Only frontal and temporoparietal lobe activation was analysed. White-ringed electrodes are reference points on the scalp and/or belong to the sagittal plane and were therefore not included in the measurement of lateral activation.

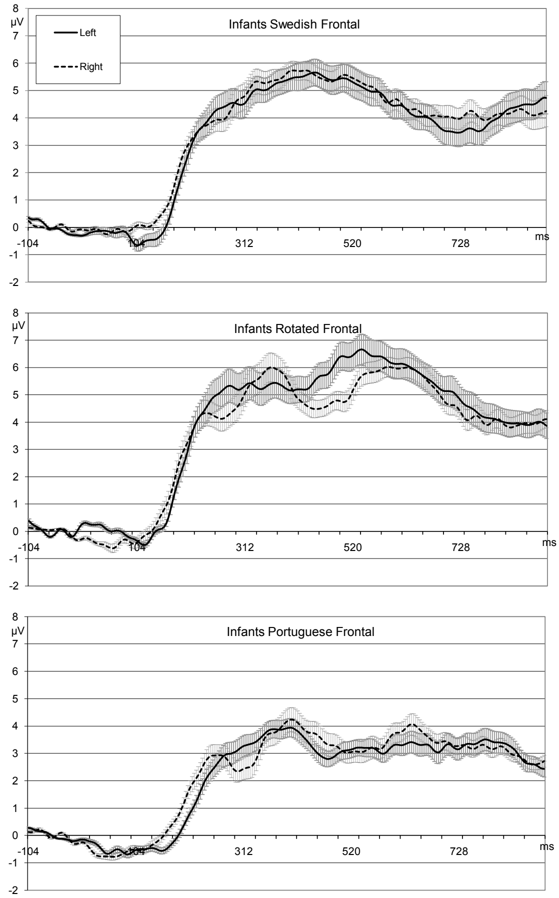


Figure 3a. Frontal lobe activation (in μV) for the left and right hemisphere in infants (N = 9) for the three stimulus conditions Swedish, rotated Swedish, and Portuguese. Stimulus onset is at 0 ms. The confidence interval is shown by standard error bars.

In the temporoparietal lobe, adults' left-hemispheric activation at a late P100 can be described as most distinct for Swedish, then for rotated Swedish, and least for Portuguese. After 300 ms, left-hemispheric activation in Swedish and in rotated Swedish continues with a late N300 whereas the graph flattens for Portuguese. At this point in time, there is a significant difference between left- and right-hemispheric activation in the rotated Swedish (at 344 ms: M_{rotLF} = –0.4 μV; SD_{rotLF} = 0.6; M_{rotRF} = 0.3 μV; SD_{rotRF} = 0.8; p < .05) and the Portuguese condition (at 344 ms: M_{rotLF} = 0 μV; SD_{rotLF} = 0.4; M_{rotRF} = 0.5 μV; SD_{rotRF} = 0.6; p < .05), but not for Swedish. In Swedish, lateralisation difference is strong just before 100 ms, again with right-hemispheric dominance.

The infants show a significant right-hemispheric dominant difference in activation between left and right temporoparietal lobe at 100 ms when listening to Swedish and rotated Swedish, but not for Portuguese. At the early response times, the ERP patterns

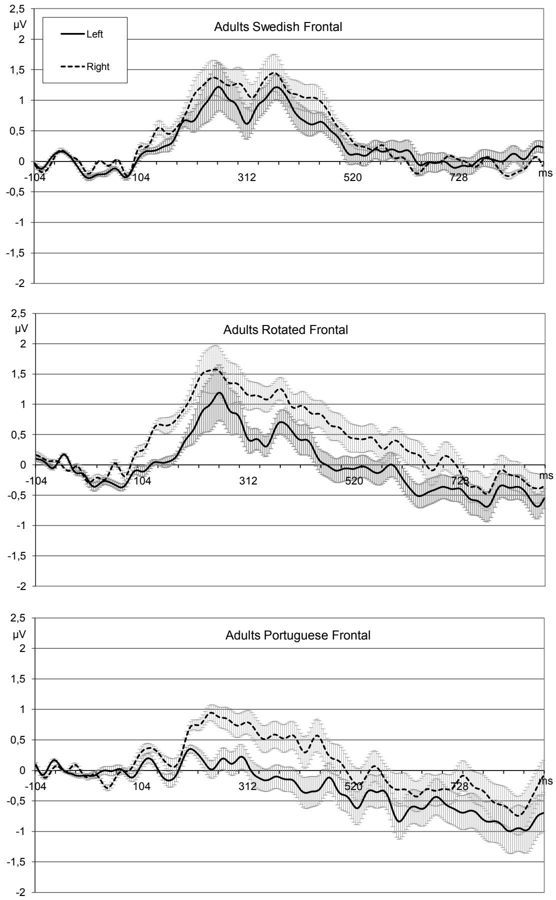


Figure 3b. Frontal lobe activation (in µV) for the left and right hemisphere in adults (N = 13) for the three stimulus conditions Swedish, rotated Swedish, and Portuguese. Stimulus onset is at 0 ms. The confidence interval is shown by standard error bars. Note that the scale is different from Figure 3a.

are especially homogenous with only little variation, as the narrowness of the confidence interval denotes. Another significant difference lies between 500 and 600 ms solely in the Portuguese condition (at 600–604 ms: M_{portLF} = –1.7 µV; SD_{portLF} = 2.8; M_{portRF} = –0.1 µV; SD_{portRF} = 2.5; p < .05). A N600 in the left hemisphere coincides here with steady right-hemispheric activation.

In adults, both in frontal and temporoparietal ERPs after 300 ms, hemispheric differences appear only for rotated Swedish and Portuguese, not for Swedish. Lateralisation dominates mostly in the right hemisphere.

In infants, the ERPs demonstrate mainly bilateral activation apart from two incidents, both of which concern the left hemisphere; a frontal P500 for rotated Swedish and a temporoparietal N600 for Portuguese. Lateralisation, if any, is dominant mostly in the left hemisphere.

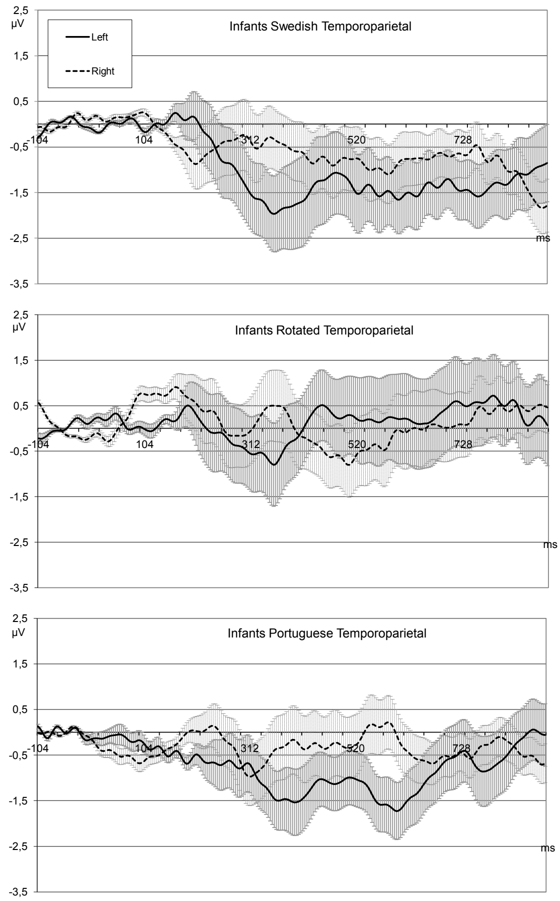


Figure 4a. Temporoparietal lobe activation (in μV) for the left and right hemisphere in infants for the three stimulus conditions: Swedish, rotated Swedish, and Portuguese. Stimulus onset is at 0 ms. The confidence interval is shown by standard error bars.

4. Discussion

Adults, but not 4-month-olds, exhibited lateralisation for speech perception in general as predicted. However, the expected left-hemispheric bias for the native language condition could not be shown in adults, whereas infants showed unpredicted left-hemispheric effects in addition to a general bilateral activation across conditions.

The right-hemispheric dominance in adults can be explained by the IDS nature and the semantic content of the speech stimuli. While semantic and syntactic language aspects are processed in the left hemisphere, prosodic elements mainly activate the right hemisphere: dynamic intonation variation, a key characteristic of IDS, is processed in the right hemisphere (Scott et al., 2000). IDS is further characterised by high affectional content (Kitamura & Burnham, 1998; Sundberg, 1998). In adults, prosodic speech without semantic content and semantically neutral sentences spoken

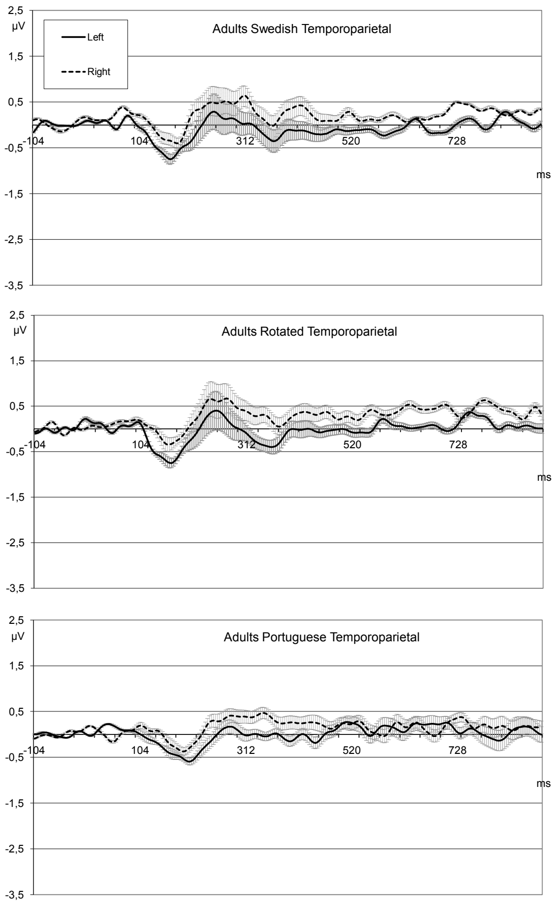


Figure 4b. Temporoparietal lobe activation (in µV) for the left and right hemisphere in adults for the three stimulus conditions: Swedish, rotated Swedish, and Portuguese. Stimulus onset is at 0 ms. The confidence interval is shown by standard error bars. The temporoparietal activation level differences between adults and infants are not as pronounced as in the frontal lobe, so the same scale can be used as in Figure 4a.

with positive, negative, or neutral intonation elicits bilateral frontal activation (Kotz et al., 2003). There is equivocal support that emotional speech perception occurs in the right hemisphere, but should be differentiated over time as bilateral mechanisms appear to be involved (Schirmer & Kotz, 2006).

The other important aspect of the IDS stimuli in this study was the simplicity of their semantic content. The semantics were possibly not sufficiently challenging for adults to evoke left-hemispheric dominance. Left-hemispheric activation in adults ceases gradually at around 400 ms in the current study, but higher-order semantics can continue left-hemispheric activation until 600 ms (Hinojosa et al., 2001). The high emotional content and the semantic simplicity of the stimuli corroborate the right-hemispheric advantage found in adults.

Infants' results were generally in line with the findings of other speech perception studies which stated bilateral activation for 3- and 4-month-olds in conditions involving natural speech, a-prosodic speech, sine wave tones, and syllables (Dehaene-Lambertz, 2000; Homae et al., 2006) and followed therefore the predictions. Apart from two instances which are discussed in the following, the infants did not show, either in Swedish, or rotated Swedish, or in Portuguese, a general hemispheric bias.

The three stimulus conditions with their variation of semantic and prosodic aspects affected the ERPs in both adults and infants differentially, therefore we need to specify further than our predictions go.

In adults' frontal and temporoparietal ERPs after 300 ms, hemispheric differences appear for rotated Swedish and Portuguese, but not for Swedish. The differences seem to develop as the left-hemispheric activation flattens rapidly after this time and opens up a gap between the right-hemispheric processing of prosody and emotion on the one hand and the abandoned left-hemispheric search for content and phonetic units in rotated Swedish and Portuguese on the other. The lack of semantic content in rotated Swedish and Portuguese cannot compete with the dominant characteristics of IDS and foreign prosody after 300 ms of listening and reduces the left-hemispheric activation significantly, while the right hemisphere still processes information.

Therefore, differences between the stimuli conditions are best visible not by contrasting the hemispheric activation, but by observing the left-hemispheric curves only. They seem to change from a distinct, although non-dominant ERP pattern for Swedish to a reduced version in rotated Swedish to a flattened graph in Portuguese.

Although no differences between speech stimuli conditions in the infant group were predicted, left-hemispheric ERPs showed a frontal P500 for rotated Swedish and a temporoparietal N600 for Portuguese. Rotated Swedish with its native-like prosody could have challenged the infants' semantic activation to elicit the P500 in the left-hemisphere of the frontal lobe in an effort to recognise familiar words such as *mamma*. *Mamma* occurred in each of the Swedish sentences, is understood by 6-month-old infants (Tincoff & Jusczyk, 1999), and can potentially be recognised by 4 ½-month-olds (Mandel et al., 1995). Similarly, the temporoparietal N600 effect for Portuguese in the left hemisphere can also be a sign for increased efforts to pick up semantic content, even though or especially because the foreign prosody signals a non-native language.

According to speech perception development theory, 4-month-olds should show the same discrimination abilities to foreign languages as they do to their native tongue, although specialisation occurs earlier for vowels than for consonants and is said to begin at around four months (Polka & Werker, 1994; Werker & Tees, 2002). Finding indication of lateral processing differences means that infants could begin to specialise well within the language-general perception phase – earlier than hitherto assumed on the basis of behavioural data. These two significant ERP effects indicate such early specialisation.

Of course, the low participant numbers in the infant group do not permit more than just to hint that 4-month-olds could in fact show first signs of differentiation between Swedish and Portuguese. Independent of semantics, the infants respond similarly to the prosody of their native language and its rotated derivate to an extent that a right-hemispheric bias occurs at around 100 ms. No such right-hemispheric activation can be seen for Portuguese. This is in line with the finding of prosodic discrimination at the age of two to three months (Karzon & Nicholas, 1989) and can be seen as a further indication of the beginning differentiation between the native and foreign languages in 4-month-olds.

5. Conclusions and further work

This study has explorative character and provides as such a basis for future research. It is one of the first to use infant-directed stimuli to investigate speech perception ERPs in infants and adults which led to unexpected right-hemispheric biases in the adults. This could be improved by using sufficiently challenging semantic content in the stimuli for the adult group. However, differences between Swedish, rotated Swedish, and Portuguese were found both in infant and adult ERPs. There are similarities in the semantic processing for Swedish and rotated Swedish after 300 ms in adults and in the prosodic processing for Swedish and rotated Swedish at 100 ms in infants. Furthermore, left-hemispheric infant ERPs showed a frontal P500 for rotated Swedish and a temporoparietal N600 for Portuguese, indicating a native language bias not only in adults, but also in 4-month-olds.

6. Acknowledgements

This study is the joint effort of the language development research group at the Phonetics Lab at Stockholm University and was funded by the Swedish Research Council (VR 421-2007-6400), the Knut and Alice Wallenberg Foundation (Grant no. KAW 2005.0115), the Bank of Sweden Tercentenary Foundation (MILLE, RJ K2003-0867), and the 6th European Framework Programme (CONTACT, NEST 5010), whose support we gratefully acknowledge. We would also like to thank all our participants without whose input, the research presented here would not have been possible.

References

Atherton, M., Zhuang, J., Bart, W. M., Hu, X. & He, S. (2003). A functional MRI study of high-level cognition: The game of chess. *Cognitive Brain Research, 16*(1), 26–31.

Burnham, D., Kitamura, C. & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435.

Canli, T., Desmond, J. E., Zhao, Z., Glover, G. H. & Gabrieli, J. D. E. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. Motivation, emotion, feeding, drinking, sexual behaviour. *Neuroreport, 9*(14), 3233–3229.

Carlsson, I., Wendt, P. E. & Risberg, J. (2000). On the neurobiology of creativity. Differences in frontal activity between high and low creative subjects. *Neuropsychologia, 38*(6), 873–885.

Chochon, F., Cohen, L., van de Moortele, P. F. & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience, 11*(6), 617–630.

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience, 12*(3), 449–460.

Dehaene-Lambertz, G. & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience, 16*(8), 1375–1387.

Federmeier, K. D., McLennan, D. B., de Ochoa, E. & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133–146.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181–195.

Fernald, A. (1992). Meaningful melodies in mother's speech to infants. In H. Papousek, U. Jürgens & M. Papousek (Eds.), *Nonverbal vocal behaviour*. Cambridge: Cambridge University Press.

Gabrieli, J. D. E., Poldrack, R. A. & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 906–913.

Gerlach, C., Law, I., Gade, A. & Paulson, O. B. (2000). Categorization and category effects in normal object recognition: A PET study. *Neuropsychologia, 38*(13), 1693–1703.

Geschwind, N. & Levitsky, W. (1968). Human brain: Left-right asymmetries in temporal speech region. *Science, 161*(3837), 186–187.

Gootjes, L., Raij, T., Salmelin, R. & Hari, R. (1999). Left-hemisphere dominance for processing of vowels: A whole-scalp neuromagnetic study. *Neuroreport, 10*(14), 2987–2991.

Hauser, M. D. & Andersson, K. (1994). Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: Field experiments. *Proceedings of the National Academy of Sciences of the United States of America, 91*(9), 3946–3948.

Hinojosa, J. A., Martín-Loeches, M. & Rubia, F. J. (2001). Event-related potentials and semantics: An overview and an integrative proposal. *Brain and Language, 78*, 128–139.

Hoen, M. & Dominey, P. F. (2000). ERP analysis of cognitive sequencing: A left anterior negativity related to structural transformation processing. *Neuroreport, 11*(14), 3187–3191.

Homae, F., Watanabe, H., Nakano, T., Asakawa, K. & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research, 54*(4), 276–280.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A. & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *Neuroreport, 17*(10), 957–962.

Jusczyk, P. W., Cutler, A. & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.

Karzon, R. G. & Nicholas, J. G. (1989). Syllabic pitch perception in 2- to 3-month-old infants. *Perception and Psychophysics, 45*(1), 10–14.

Kitamura, C. & Burnham, D. (1998). The infant's response to maternal vocal affect. In C. Rovee-Collier, L. P. Lipsitt & H. Hayne (Eds.), *Advances in Infancy Research* (pp. 221–236). Stamford: Ablex Publishing Corporation.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., et al. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain, 123*(12), 2512–2518.

Koeda, M., Takahashi, H., Yahata, N., Asai, K., Okubo, Y. & Tanaka, H. (2006). A functional MRI study: Cerebral laterality for lexical-semantic processing and human voice perception. *American Journal of Neuroradiology, 27*(7), 1472–1479.

Kotz, S. A., Meyer, M. S., Alter, K., Besson, M., Von Cramon, D. Y. & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language, 86*(3), 366–376.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience, 5*(11), 831–843.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, E. I., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686.

Mandel, D. R., Jusczyk, P. W. & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science, 6*(5), 314–317.

Marsolek, C. J., Squire, L. R., Kosslyn, S. M. & Lulenski, M. E. (1994). Form-specific explicit and implicit memory in the right cerebral hemisphere. *Neuropsychology, 8*(4), 588–597.

Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G. & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition, 52*(3), 382–389.

Obleser, J., Zimmermann, J., Van Meter, J. & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex, 17*(10), 2251–2257.

Ojemann, G., Ojemann, J., Lettich, E. & Berger, M. (1989). Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery, 71*(3), 316–326.

Panneton Cooper, R. & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development, 65*(6), 1663–1677.

Platel, H., Price, C., Baron, J.-C., Wise, R., Lambert, J., Frackowiak, R. S., et al. (1997). The structural components of music perception: A functional anatomical study. *Brain, 120*(2), 229–243.

Polka, L. & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421–435.

Rypma, B. & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences of the United States of America, 96*(11), 6558–6563.

Schirmer, A. & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*(1), 24–30.

Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400–2406.

Seger, C. A., Poldrack, R. A., Prabhakaran, V., Zhao, M., Glover, G. H. & Gabrieli, J. D. E. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia, 38*(9), 1316–1324.

Sundberg, U. (1998). *Mother tongue – Phonetic aspects of infant-directed speech.* Perilus XXI. Published PhD thesis. Department of Linguistics, Stockholm University, Stockholm, Sweden.

Tees, R. C. & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579–590.

Tincoff, R. & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172–175.

Weinstein, S. & Graves, R. E. (2002). Are creativity and schizotypy products of a right hemisphere bias? *Brain and Cognition, 49*(1), 138–151.

Werker, J. F. & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50*, 509–535.

Werker, J. F. & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 25*, 121–133.
Zangl, R. & Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6- and 13-month-old infants. *Infancy, 11*(1), 31–62.

Rapid, automatic, and parallel language processing in the brain

Neurophysiological data using the mismatch negativity

Yury Shtyrov, Friedemann Pulvermüller
Cognition and Brain Sciences Unit, Medical Research Council, Cambridge

.

Abstract

We review recent studies investigating spoken language processing in the brain using the mismatch negativity (MMN) brain response. They found that size and topography of the MMN reflect activation of memory traces for language elements in the human brain, even under attentional withdrawal. Familiar sounds of one's native language may elicit a larger brain activation than unfamiliar sounds, and at the level of meaningful language units, words elicit a larger response than meaningless word-like sounds. Unattended word stimuli elicit an activation sequence starting with superior-temporal cortex and rapidly progressing to left-inferior-frontal lobe. The fine-grained spatio-temporal patterns of such cortical activations depend on lexical and semantic properties of spoken stimuli, providing further information about neural configuration of respective memory traces. At the phrase level, we can see MMN reflections of grammatical regularities and contextual semantics in word strings. This growing body of results suggests that the earliest stages of phonological, lexical, semantic and syntactic processing can be carried out by the central nervous system outside the focus of attention in an largely automatic manner. Analysis of such attention-independent responses to speech suggested that a variety of linguistic information types are accessed by our brain rapidly (within ~200 millisecond) in a near-simultaneous, if not parallel, fashion.

1. Introduction

1.1. Background: Psycholinguistics

Traditional psycholinguistic models of language comprehension (Morton, 1969; Fromkin, 1973; Garrett, 1980; Dell, 1986; MacKay, 1987), also reflected in current approaches to speech comprehension and production (Levelt et al., 1999; Norris et al., 2000), distinguished a few types of information involved in these processes. Even though these theories may be rather different, they mostly agree on the plausibility of (a) a *phonological* processing level, at which speech sounds are analysed for their phonetic/phonological features, following the level of basic acoustic analysis; (b) a *lexical* processing level, sometimes conceptualised as the lookup of an item in a "mental lexicon", which lists only word forms but not their meaning or other related information; (c) a *semantic* processing level, where the item's meaning is accessed; and (d) a *syntactic* processing level, where grammatical information linking words in sentences to each other is analysed. These different "modules" of processing can sometimes be merged, omitted, or extended to include more types of information. A great debate in psycholinguistics is, however, between models according to which processing of these information types is consecutive, and access to them commences at substantially different times, and models implying that these processes take place nearly simultaneously, even in parallel. This issue could potentially be resolved on the basis of neurophysiological data, using EEG or MEG which can track brain processes with unprecedented temporal resolution, on a millisecond range.

1.2. Background: Neurophysiology

Neurophysiological research did not fail to live up to this promise. A body of research done on various linguistic materials formed an established view on the serial order of linguistic information access in the brain. With basic acoustic feature extraction commencing at 20–50 ms after the stimulus onset (Krumbholz et al., 2003; Lutkenhoner et al., 2003), phonological tasks modulated responses with latencies of 100–200 ms (Poeppel et al., 1996; Obleser et al., 2004). "Higher" levels of information appeared to take substantially longer to be appreciated by the human brain: the N400 component, which peaks around 400 ms after the onset of a critical visual word (Kutas & Hillyard, 1980) is traditionally seen as the main index of semantic processes. A slightly earlier (350 ms) peaking sub-component of it is sometimes separated as an index of lexical processing (Bentin et al., 1999; Embick et al., 2001; Pylkkanen et al., 2002; Stockall et al., 2004) aiding the elusive distinction between the lexical and semantic levels of processing. An even more complicated situation emerged for syntactic processing, with an early component (Early Left Anterior Negativity, ELAN) appearing at ~100 ms and reflecting the phrase's grammaticality (Neville et al., 1991; Friederici et al., 1993); this is complemented by later grammatically related frontal negativities with longer (> 250 ms) latencies (Münte et al., 1998; Gunter et al., 2000), and, finally, by the P600, a late positive shift reaching its maximum at ~600 ms at centro-parietal sites (for review, see Osterhout et al., 1997; Osterhout & Hagoort, 1999).

1.3. Conflicting evidence

Such data suggested a stepwise access to different types of linguistic information, forming a mainstream view in the neurophysiology of language (for a review, see Friederici, 2002). This, however, came into a sharp conflict with a body of behavioural evidence collected in a number of psycholinguistic studies, which indicated parallel processing of crucial information about incoming words and their context very early, within the first ~200 ms after a critical word can be recognized (Marslen-Wilson, 1973; Marslen-Wilson, 1987; Rastle et al., 2000; Mohr & Pulvermüller, 2002). For example, subjects can already make reliable button-press motor responses to written words according to their evaluation of aspects of phonological and semantic stimulus properties within 400–450 ms after their onset (Marslen-Wilson & Tyler, 1975). Therefore, the earliest word-related psycholinguistic processes as such must take place substantially before this, as considerable time is required for the motor response preparation and execution. Surprisingly, at these early latencies, linguistic processing was already influenced by syntactic and semantic information (Marslen-Wilson & Tyler, 1975). Furthermore, studies using the shadowing technique suggested that the language output must be initiated by 150 to 200 ms already after the input onset (Marslen-Wilson, 1985). Early behavioural effects reflecting semantic processing and context integration were documented in cross modal priming, where specific knowledge about an upcoming spoken word could be demonstrated to be present well before its end, within 200 ms after the acoustic signal allows for unique word identification or even well before this word recognition point (Zwitserlood, 1989; Moss et al., 1997; Tyler et al., 2002). Eye-Tracking experiments demonstrated that a range of psycholinguistic properties of stimuli influences short-latency eye movement responses (Sereno & Rayner, 2003). The view imposed by such studies is that near-simultaneous access to the different types of linguistic information may commence within 200 ms after the relevant information becomes available. This calls into question the stepwise models of language processing also put forward in psycholinguistics and supported, at least partially, by early ERP experiments.

2. Mismatch negativity

It has also been a matter of debate whether the distinctively human capacity to process language draws on attentional resources or is automatic. The ease with which we can perceive the entire complexity of incoming speech and seemingly instantaneously decode syntactic, morphological, semantic and other information while doing something else at the same time, prompted suggestions that linguistic activity may be performed by the human brain in a largely automatic fashion (Fodor, 1983; Garrett, 1984; Garrod & Pickering, 2004; Pickering & Garrod, 2004). To address this here, we will focus on data obtained outside the focus of attention in a passive oddball paradigm using mismatch negativity (MMN) brain response. MMN is an evoked brain response elicited by rare (so-called deviant) acoustic stimuli occasionally presented in a sequence of frequent (standard) stimuli (Alho, 1995; Näätänen, 1995). Importantly, MMN can be elicited in the absence of the subject's attention to the auditory input (Tiitinen et al., 1994; Schröger, 1996). It therefore has become considered to reflect the brain's automatic discrimination of changes in the auditory sensory input and thus to be a unique indicator of automatic cerebral processing of acoustic events (Näätänen, 1995). More recently, the MMN has been increasingly utilised for investigating the neural processing of speech and language (Näätänen, 2001; Pulvermüller & Shtyrov, 2006). Let us first consider the main motivations for applying MMN to explore the brain's processing of language. We would like to suggest the following main reasons, further detailed in the rest of this section, for this: (a) MMN is early; (b) MMN is automatic; (c) MMN is a response to individual sounds; (d) MMN is a response to change. The first two points are based on the known properties of MMN response tentatively related to its neural mechanisms, whereas the second two are important from technical and methodological points of view related to recording brain responses to language.

2.1. The earliness of the MMN response

As mentioned, a range of behavioural data have suggested that linguistic processes in the brain commence within 200 ms after the relevant acoustic information is available. Should these processes be reflected in the dynamics of event-related activity recorded on the scalp surface, the respective brain responses must have similar, if not earlier, latencies in order to be considered directly related to these processes rather then being their remote consequences. This clearly eliminates late shifts (N400, M350, P600) as potential indicators of early processing. Early obligatory auditory responses (P1, N1) have not been found sensitive to linguistic variables. ELAN, the early syntax-related response, is well within the time window, but cannot be seen for most types of linguistic (even syntactic) tasks. MMN, on the contrary, is both early (usually reported as having latencies of 100–200 ms) and firmly linked to such cognitive processes as memory, attention allocation, and primitive auditory system intellect (Näätänen, 1995; Näätänen & Alho, 1995; Näätänen, 2000; Näätänen et al., 2001). It is known to be sensitive to highly abstract features of auditory signals and is therefore a good candidate for identifying putative early linguistic activations.

2.2. Automaticity of the MMN

At this stage, we would like to leave aside the debate of whether or not, and to what extent, the MMN is independent of or modulated by attention. It is uncontroversial that MMN can be easily elicited in the absence of attention to the stimulus input and therefore does not require active voluntary processing of stimulus material by the individual, who may be engaged in an unrelated primary task while MMN responses are being evoked by random infrequent changes in auditory stimulation. At least in this sense the MMN can be considered an automatic brain response. This, in turn, has an important implication for language research. Typically, in language experiments, subjects are asked to attend to presented words or sentences (e.g. Neville et al., 1991; Osterhout & Swinney, 1993; Friederici et al., 2000). Often, the task is to make a judgment of the stimulus material (e.g. familiar/unfamiliar, correct/incorrect) or even perform a specific linguistic task (e.g. lexical decision, grammar assessment). When attention is required, one can not be sure to what extent the registered responses are influenced by brain correlates of attention rather than by the language-related activity as such. Attention-related phenomena are known to modulate a variety of the brain's evoked responses involving a number of brain structures including those close to, or overlapping with, the core language areas (see e.g. Picton & Hillyard, 1974; Alho, 1992; Woods et al., 1993a; Woods et al., 1993b; see e.g. Tiitinen et al., 1997; Escera et al., 1998; Yamasaki et al., 2002; Yantis et al., 2002). It is also likely that subjects pay more attention to unusual or incorrect stimuli (pseudowords, nonsense sentences, and grammatical violations are among the most widely used stimulus types) as they try to make some sense of them, or that they process proper and malformed items using different strategies. Such different stimulus-specific strategies and attention variation may be reflected in the event-related measures, overlapping with true language-related activity. Limiting such attention- and task-related effects is therefore highly important. The MMN provides a straightforward solution to this, as it can be recorded when the subjects are distracted from the stimuli and are not engaged in any stimulus-oriented tasks.

2.3. MMN as a single-item response

Whereas early psycholinguistic processes have been suggested, such putative early activity has remained mostly undetected neurophysiologically. One reason for the failure of most studies to detect any early language activation may be of methodological nature. In most brain studies of language, large groups of stimuli are investigated and compared with each other, and average responses are used to draw general conclusions on all materials falling into a certain category. This leads to the problem of physical stimulus variance, with stimuli having different physical features (e.g. duration, spectral characteristics, distribution of sound energy, etc.). Differences even in basic physical features may lead to differential brain activation (Näätänen & Picton, 1987; Korth & Nguyen, 1997) that could in principle overlap with, mask, or be misinterpreted as language-related effects. This also leads to psycholinguistic variance problems, with stimuli differing in their linguistic features, e.g. the frequency of their occurrence in the language or their word recognition parameters. The latter may be especially difficult to control, as different words, even of identical length, become uniquely recognized from their lexical competitor environment at different times, in extreme cases shortly after their onset or only after a substantial post-offset period (Marslen-Wilson, 1987). Here, the traditional strategy of matching average parameters across stimulus categories is undoubtedly useful, but still has a serious caveat: if the brain responses reflecting early linguistic processes are small and short-lived (as all known early ERPs peaks are), the variance in the stimulus group may reduce or even remove any effects in the average brain responses (Pulvermüller, 1999; Pulvermüller & Shtyrov, 2006). Later responses (N400, P600), on the other hand, will likely survive such averaging, as they are large in amplitude and span across hundreds of milliseconds. Therefore, to locate the putative early effects with any certainty, stimulus variance should be maximally reduced. As MMN is typically a response to a single deviant item presented randomly a large number of times in order to optimize signal-to-noise ratio (SNR) of the ERP, this offers an ultimate control over the stimulus variance by removing it altogether.

2.4. MMN as a difference response

By definition, the MMN is obtained as a difference between responses to standard and deviant stimuli and is elicited by contrasts between the two. This offers a unique opportunity to strictly control acoustic stimulus properties in language experiments. This can be done by using the same identical acoustic contrasts in different experimental conditions, while manipulating their linguistic properties. Consider the following hypothetic experimental conditions: (a) the word *bay* as standard stimulus vs. *bays* as deviant; (b) *lay* vs. *lays*; (c) *gray* vs. *graze*; (d) *tay* vs. *taze*. In all of these, the MMN responses would be elicited by the presence of stimulus final sound [z] in the deviant as opposed to silence in the standard one. So, a purely acoustic MMN elicited by deviant–standard stimulus contrast should stay the same. However, the four conditions differ widely in their linguistic context while incorporating the identical acoustic contrast: in (a), this contrast constitutes a noun inflection (number change); in (b), a verb is inflected (3rd person) instead. In (c), the same acoustic contrast signifies a complete change of the stimulus' lexical and semantic features when both part-of-speech information and meaning diverge. Finally, in (d), two meaningless pseudowords are contrasted offering an additional acoustic control for any effects that can be obtained in the other three conditions. So, by recording MMNs to such identical acoustic contrasts, one may focus on effects of the different linguistic contexts without the usual confound of diverging stimulus features. Additionally, the same identical deviant stimuli can be presented in a separate set of conditions as frequent standards, and the MMN can then be computed by using responses to physically identical items presented as both deviants and standard, offering an ultimate control over the physical stimulus features. The use of such strictly controlled stimulation proved to be very successful in a number of studies. Let us now briefly review this evidence.

3. Experimental data to date

At the time of presenting this review at the 1st Birgit Rausing Conference in Linguistics, experimental evidence had been collected that demonstrates MMN sensitivity to linguistic processes in all major domains of language information: phonological, lexical, semantic, and syntactic (Figure 1).

3.1. Phonological processes

Single phonemes and their simple combinations, syllables, were the first linguistic materials to be studied using MMN paradigm. These experiments showed that native language sounds, e.g. vowels, elicit larger MMN responses than their analogues that do not have corresponding representations in one's phonological system (Dehaene-Lambertz, 1997; Näätänen et al., 1997). This happened irrespective of the magnitude of acoustic contrasts between the standards and deviants. Furthermore, while MMN is usually a bilateral or even right-dominant response (Paavilainen et al., 1991), such phonetic MMNs showed left-hemispheric dominance, potentially linking their origin to the

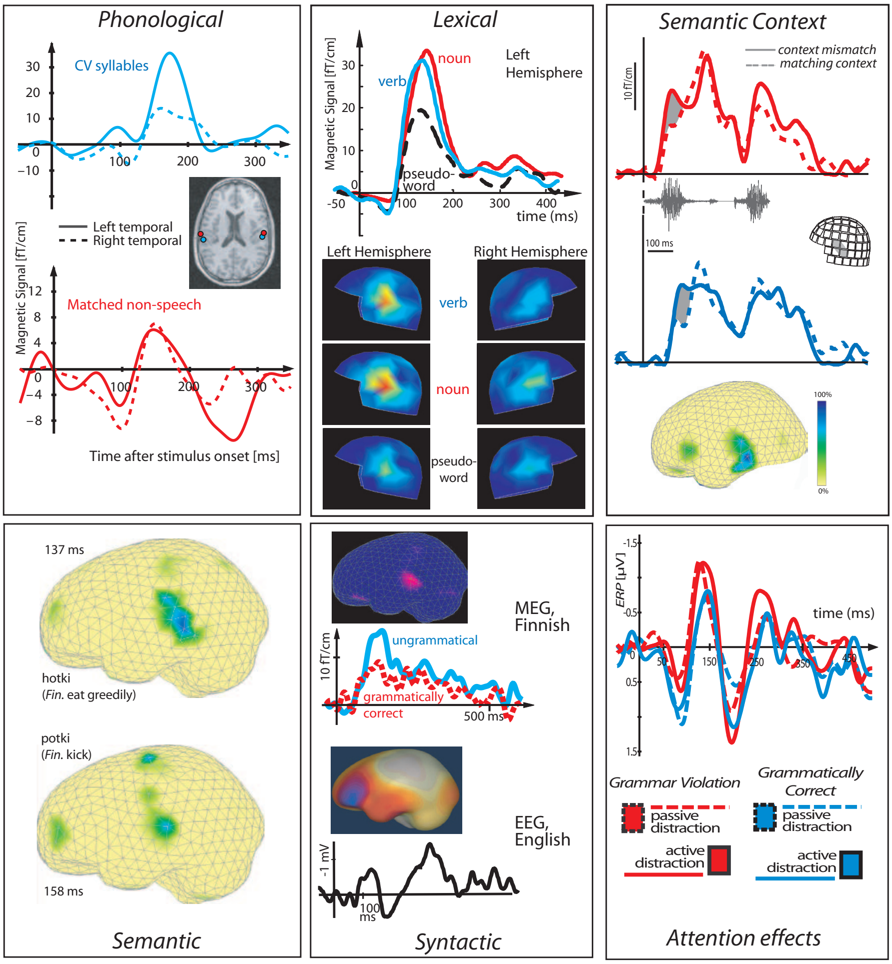


Figure 1. MMN reflections of different types of linguistic information. Left to right, top to bottom: (a) phonological enhancement for syllables (adopted from Shtyrov et al., 2000); (b) lexically enhanced MMN for meaningful words (adopted from Shtyrov et al., 2005); (c) semantic context integration (adopted from Shtyrov & Pulvermüller, 2007); (d) single-word category-specific semantic effects (adopted from Pulvermüller et al., 2005); (e) MMN in response to syntactic violations in spoken phrases of different languages (adopted from Pulvermüller & Shtyrov, 2003, and Shtyrov et al., 2003); (f) absence of attention effects on early stages of syntactic processing (adopted from Pulvermüller & Shtyrov, 2008).

language-specific structures housed in the left hemisphere (Näätänen et al., 1997; Shtyrov et al., 1998; Shtyrov et al., 2000). Further experiments even showed how changes in the pattern of the MMN response may reflect development of a new phonological representation in the process of learning a language by children or adults (Cheour et al., 1998; Winkler et al., 1999). Later experiments demonstrated MMN sensitivity to phonotactic probabilities (Dehaene-Lambertz et al., 2000; Bonte et al., 2005; Bonte et al., 2007), stress patterns (Honbolygo et al., 2004; Weber et al., 2004), audio-visual integration of phonetic information (Colin et al., 2002; Colin et al., 2004), and various other phonological variables. The timing of these phonetic and phonological mismatch negativities ranged from close to 100 ms (Rinne et al., 1999) to 200–300 ms (D'Arcy et al., 2004; Connolly et al., 1995).

To explain these results it was suggested that, in addition to change-detection and short-term memory processes, MMN is sensitive to long-term memory traces preformed in the subject's neural system in the process of their previous experience with spoken language (Näätänen et al., 1997; Shtyrov et al., 2000; Näätänen, 2001). Importantly, this implied that such long-term memory traces for language elements can become activated in the brain by a deviant stimulus in an odd-ball sequence and this specific activation can be recorded neurophysiologically even without attention to the stimulus or any stimulus-oriented task. This led to further MMN experiments with language stimuli that targeted "higher-order" linguistic processes.

3.2. Lexical processes

Similar to the phonological enhancement of the MMN in response to the native language's phonemes and syllables, we found that mismatch negativity evoked by individual words was greater than for comparable meaningless word-like (i.e., obeying phonological rules of the language) stimuli. In a series of studies, we presented subjects with sets of acoustically matched word and pseudoword stimuli and found an increased MMN response whenever the deviant stimulus was a meaningful word (Pulvermüller et al., 2001; Shtyrov & Pulvermüller, 2002; Pulvermüller et al., 2004; Shtyrov et al., 2005). This enhancement, typically peaking at 100–200 ms, is best explained by the activation of cortical memory traces for words realised as distributed strongly connected populations of neurones (Pulvermüller & Shtyrov, 2006). The lexical enhancement of the MMN, repeatedly confirmed by our group, was also demonstrated by other groups using various stimulus set-ups and languages (Korpilahti et al., 2001; Kujala et al., 2002; Sittiprapaporn et al., 2003; Endrass et al., 2004; Pettigrew et al., 2004). Our studies indicated that the lexical status of the deviant stimulus was relevant for eliciting the MMN, but the lexical status of the standard stimulus did not significantly affect MMN amplitude (Shtyrov & Pulvermüller, 2002). Other reports suggested that the event-related brain response evoked by the speech standard stimulus may also be affected by its lexical status (Diesch et al., 1998; Jacobsen et al., 2004).

Scrutinising word-elicited MMN activation using state-of-the-art technology, such as whole-head high-density MEG combined with source reconstruction algorithms, allowed for detailing word-related processes in the brain at unprecedented spatio-temporal scale. Using MMN responses to single words, we could demonstrate how the activation appears first in the superior temporal lobe shortly (~130 ms) after the information allows for stimulus identification, and, in a defined order (with a delay of ~20 ms), spreads to inferior-frontal cortices, potentially reflecting the flow of information in the language-specific perisylvian areas in the process of lexical access (Pulvermüller et al., 2003). Furthermore, we found a correlation between the latency of MMN responses in individual subjects and the stimulus-specific word recognition points determined behaviourally, which suggests that word-elicited MMN reflects the processes of individual word recognition by the human brain and the two processes may therefore be linked to a common part (Pulvermüller et al., 2006).

As a result of such growing body of evidence, the pattern of MMN response to individual word deviants has been considered as indicating a word's "signature", its memory trace activated in the brain, which offers a unique opportunity to investigate neural processing of language in non-attention demanding and task-independent fashion. Indeed, investigations of language function using MMN that followed were able to provide a great level of detail on the spatio-temporal patterns of activation potentially related to processing of semantic information attached to individual words.

3.3. Semantic processes

The first experiments targeting semantic word properties using MMN utilised predictions of the somatotopy of an action words model (Pulvermüller, 1999, 2005). The latter is a part of a general framework maintaining that word-specific memory traces exist in the brain as distributed neuronal circuits formed in the process of mutual connection strengthening between different (even distant) areas, as actions, objects, or concepts are experienced in conjunction with the words used to describe them (Hebb, 1949; Pulvermüller, 1999). The referential meaning is an integral part of a word's semantics (Frege, 1980). Action words' regular usage for referring to arm/hand actions (words such as *pick, write*) or leg/foot actions (*kick, walk*) is therefore an essential characteristic of their meaning (even though their semantics may not be exhausted by it). If lexical representations become manifest cortically as neuronal assemblies and the actions referred to by these words are somatotopically represented in motor areas of the brain (Penfield & Rasmussen, 1950), the semantic links between neuronal sets in these cortical regions should realise the semantic relationship between the word forms and their actions (cf. the somatotopy of action word model in Pulvermüller, 2001). Crucially, this leads to the specific prediction that action words with different reference domains in the body also activate the corresponding areas of motor cortex. This claim, which received strong support from conventional neuroimaging studies (Hauk et al., 2004; Hauk & Pulvermüller, 2004), was put to test in MMN experiments using different methods (EEG, MEG) and different stimuli from two languages (English, Finnish).

In both experiments (Shtyrov et al., 2004; Pulvermüller et al., 2005), in addition to the usually observed superior temporal MMN sources, activation elicited by the words referring to face and/or arm movements involved inferior fronto-central areas likely including the cortical representation of the upper body. Furthermore, the leg-related words elicited a stronger superior central source compatible with the leg sensorimotor representation. This leg-word specific superior fronto-central activation was seen later

(~170 ms) than the more lateral activation for face- and arm-related words (~140 ms). These spatio-temporal characteristics suggest that MMN sources in perisylvian areas along with near-simultaneous activation distant from the Sylvian fissure can reflect access to word meaning in the cortex. The minimal delays between local activations may be mediated by conduction delays caused by the traveling of action potentials between cortical areas.

These data suggested that processing of semantic features of action words is reflected in the MMN as early as 140–180 ms after acoustic signals allow for word identification. Similar to lexical access and selection, meaning access may therefore be an early brain process occurring within the first 200 ms, and the two processes can be concluded to be near-simultaneous.

Even more intriguing results were produced by MMN experiments on semantic context integration. In these, complete phrases or word combinations were presented to subjects in a passive non-attend oddball design while the physical stimulus properties were strictly controlled for. Sometimes, the deviant combinations included a semantic mismatch between the words, similar to the established semantic violation paradigm known to produce an N400 response. Surprisingly, these contextual sentence-level violations modulated the MMN response rather early in its time course and, at the same time, elicited no N400-like response in the passive oddball paradigm. In one of the studies, this modulation was seen as early as ~115 ms after the words could be recognised as different and was shown to be mediated by the superior-temporal and inferior-frontal cortices in the left hemisphere (Shtyrov & Pulvermüller, 2007). In another study (Menning et al., 2005), the semantic abnormalities were detected by MMN responses at 150–200 ms. Crucially, these separate proofs of MMN reflections of semantic context integration were obtained in different laboratories which utilised diverging techniques, stimulation protocols and even stimulus languages (Finnish, German). These MMN studies showed for the first time higher-order processes of semantic integration of spoken language occurring well before the N400 time range, within 200 ms. This strongly supports the notion of parallel access to different types of linguistic information. Additional support for this view was found in recent visual studies (Sereno et al., 2003; Penolazzi et al., 2007) indicating that the semantic context integration affects visual brain responses at latencies under 200 ms.

3.4. Syntactic processes

Finally, only one domain remains uncovered by this review: that of morpho-syntactic information processing in the brain. This was addressed in a number of experiments by our group and was also independently verified in other laboratories. To control exactly for physical stimulus properties, we once again presented identical spoken stimuli in different contexts. In these experiments, the critical word occurred after a context word with which it matched or mismatched in syntactically; this has been a standard approach used in neurophysiological studies on syntax (Osterhout, 1997). We again used different methods (EEG and MEG) and languages (Finnish, German, and English), enabling us to draw generalised conclusions which is especially important in MMN studies when only a very limited set of stimuli is used. The first experiment looked at the neurophysiological responses to Finnish pronoun-verb phrases ending in a verb suffix which did or did not agree with the pronoun in person and number; this was done in an orthogonal design in which physically identical stimuli marked a syntactically congruent or incongruent event, thus fully controlling for physical stimulus features.

The results showed an increase of the magnetic MMN to words in an ungrammatical context compared with the MMNs to the same words in a grammatical context. Similar results were also seen in English and German experiments (Pulvermüller & Shtyrov, 2003; Pulvermüller & Assadollahi, 2007). The latencies where grammaticality effects were found varied somewhat between studies but responses were generally present within 200 ms after the word recognition point, sometimes starting as early as at 100 ms (Pulvermüller & Shtyrov, 2003; Shtyrov et al., 2003). The cortical loci where the main sources of the syntactic MMN were localized also varied. MEG results indicated a distributed superior temporal main source with possibly some weak effects in inferior frontal cortex whereas the EEG suggested the opposite, a pronounced grammaticality effect in the left inferior frontal cortex with possibly a minor superior-temporal contribution. This divergence reflects the previous neuroimaging literature on the cortical basis of syntax, where this module is sometimes localized in frontal areas and sometimes in temporal lobes (e.g. Kaan & Swaab, 2002; Bornkessel et al., 2005; Kircher et al., 2005). It is likely therefore that different areas in perisylvian cortex contribute to grammatical and syntactic processing and their activity may be differentially reflected in EEG and MEG recordings due to the known specificity of these techniques to positioning and orientation of current sources in the brain.

Our data were further supported by findings of other groups which used similar paradigms and found a specific MMN response peaking at 150–200 ms whenever the presented sentence stimulus contained a syntactic mismatch (Menning et al., 2005; Hasting et al., 2007).

The early syntactic MMN resembles the ELAN component (Neville et al., 1991; Friederici et al., 1993) which has been suggested to index early syntactic structure building (Friederici, 2002). Syntactic MMN results support this interpretation. Importantly, they also show that the early syntactic processing in the brain does not require focused attention on the language input. In this sense, early syntactic processing seems to be automatic. This appears true even for agreement violations, which are considered more demanding computationally and which did not elicit ELAN in the past. The late positivity (P600), which is abolished in the passive oddball task, may in turn reflect secondary controlled attempts at parsing a string after initial analysis has failed (Osterhout & Holcomb, 1992; Friederici, 2002).

3.5. Attentional control or automaticity?

One may argue that demonstrating language related effects in a paradigm where subjects are instructed to attend to a video film or book while language stimuli are presented does not control strictly for attentional withdrawal. To draw conclusions on the automaticity of the processes investigated from such studies, subjects must strictly follow the instruction to try to ignore the speech stimuli. It would be desirable to control for the attentional withdrawal in each subject throughout the experiment. Therefore, we performed an experiment to further investigate the role of attention in language processing by comparing the classic MMN paradigm with its moderate attentional withdrawal by a silent video film with a distraction task where subjects had to continuously perform an acoustic detection task. Language stimuli were only played through the right ear while acoustic stimuli were delivered to the left ear. In a streaming condition, subjects had to press a button to a "deviant" acoustic stimulus in the left ear while, at the same time, the language stimuli were played to the right ear. In the other condition, subjects were allowed to watch a video as usual, without further distraction, while the same stimuli, sounds and language stimuli, were played. Results showed a replication of the grammaticality effect, i.e. stronger MMNs to ungrammatical word strings than to ungrammatical ones up to a latency of ~150 ms. There was no difference between task conditions varying attentional withdrawal. Only later did we find significant interactions of the task and attention factors indicating that at these later stages the grammar processes revealed by the MMN were influenced by attention and task demand. We interpret this as strong evidence for the attention independence of the early part of the MMN and for the automaticity of early syntactic analysis (Pulvermuller et al., 2008).

4. Conclusions and challenges

We have considered the premises for neurophysiological research into the earliest stages of language perception and how the MMN could be used to help elucidate them. Four major motivations for such use of this brain response have been presented: MMN earliness, its automaticity, the possibility to record responses to individual language stimuli using MMN design, and the unique feature of the MMN as a difference response that allows one to control for physical stimulus properties. We have then seen how the MMN has been used over the recent years to assess the neural processing of phonological, lexical, semantic (including both single-word semantics and higher-order semantic context integration at a phrase level) and syntactic information. Not only has the MMN been found sensitive to these different types of linguistic information (Figure 1), but it has also proved useful for disentangling the neural correlates of corresponding processes from those related to attention or stimulus-oriented tasks and strategies; even more importantly, the MMN studies allowed for the strictest possible control over the stimulus properties thus minimizing the possibility that the obtained effects are simply due to acoustically or psycholinguistically unbalanced stimulus sets.

Most crucially, these studies shed new light on the time course and, at least in some cases, on the cerebral structural basis of access to linguistic information. Throughout these experiments, different types of stimuli, including phonetic/phonological, lexical, semantic, and syntactic contrasts, modulated MMN responses at latencies within 100–200 ms after the relevant acoustic information was present in the input. These data supported, and very strongly so, parallel or near-simultaneous access to different types of linguistic information commencing very rapidly in the brain shortly after the words could be uniquely identified. Such access was envisaged in some earlier psycholinguistic studies and remained hotly debated until now.

In the spatial domain, the MMN results provided strong support for the existence of distributed neural circuits which may underlie the processing of the incoming linguistic information. For example, an interplay of activations between superior-temporal and inferior-frontal cortices was shown as occurring in the process of word perception (Pulvermüller et al., 2003, 2005). Depending on the exact referential semantics of the words presented, constellations of areas in temporal, inferior-frontal, fronto-central dorsal cortices were sparked by the stimuli, sometimes even spreading to the right hemisphere (Pulvermüller et al., 2004). The opportunities offered by the MMN to explore such complicated spatio-temporal patterns linked to language comprehension are certainly far from being exhausted at this stage.

The MMN results by no means falsify the earlier studies which delineated the later neurophysiological indices of language processing. Instead, they add a new dimension to neurolinguistic research, suggesting that these later processes reflected in e.g. N400, P600 or syntactic negativities (LAN) may build upon the early automatic ones and are possibly related to secondary mental processes that can be triggered by the earlier access to linguistic memory representations, but depend on attention and task-related strategies.

The MMN research into language still has some open issues. One of most frequent criticisms of the MMN design concerns the issue of attention control in MMN experiments. Traditional passive MMN paradigms, in which the subjects' primary task is to watch a videofilm, play a computer game or read a book, indeed cannot guarantee that the subject cannot sometimes "sneak a little listen" to the experimental stimuli (Carlyon, 2004). One of the possible directions for future research, therefore, is to further explore more stringent protocols, including distracters in the same modality, in order to be able to fully separate attention-independent, and thereby possibly automatic, linguistic processes from those that require the individual's active participation. As reviewed above, we have been successful in doing this to validate the syntactic MMN as being attention-independent at its earliest stages. Although this, combined with other results, strongly suggests that the earliest linguistic access is automatic, this approach needs to be extended to different types of linguistic information processing.

Another obvious critical issue of the MMN approach to language lies with the inherently small selection of stimuli that can be used in the context of an MMN experiment (the obvious benefits of this approach were covered earlier in this review). The small number of tokens and the need for their multiple repetitions inevitably lead to difficulties in generalizing results obtained in this way to the entirety of natural language. Here, replication of results using different stimuli as well as different languages and experimental setups proved useful. A possible future direction is using larger groups of stimuli thus introducing a certain stimulus variability typical of natural language settings; this has already been done for single phonemes (Shestakova et al., 2004). However, applying this approach to more complicated stimuli, even at the level of words, requires stringent control over their physical and, importantly, psycholinguistic features as well as over the variance of these dimensions in each stimulus group which will undoubtedly prove to be a challenge for researchers willing to tackle this problem.

Finally, there is an obvious conflict between the traditional interpretation of the MMN as related to auditory change-detection mechanisms and short-term acoustic memory processes on one hand, and the nature of MMN responses to linguistic stimuli on the other. The latter require long-term memory traces to be present for their elicitation, and exhibit complex spatio-temporal patterns frequently involving areas beyond the auditory cortices. This has led to a suggestion that linguistic activation in the MMN paradigm is a separate process which takes place in addition to the classical change-detection MMN (Pulvermüller & Shtyrov, 2006). It may in theory be that invoking such processes need not necessarily involve the time-inefficient oddball paradigm and may be effectively done under different conditions. Exploring further conditions for recording memory-trace activations in the brain in an attention-free and strictly controlled manner may not only shed further light on the brain processes underlying them but may also lead to improved experimental paradigms facilitating further studies of the brain mechanisms of language functions.

References

Alho, K. (1992). Selective attention in auditory processing as reflected by event-related brain potentials. *Psychophysiology*, *29*, 247–263.

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear*, *16*, 38–51.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F. & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*, 235–260.

Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H. & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology, 116*, 2765–2774.

Bonte, M. L., Poelmans, H. & Blomert, L. (2007). Deviant neurophysiological responses to phonological regularities in speech in dyslexic children. *Neuropsychologia*, *45*, 1427–1437.

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y. & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage*, *26*, 221–233.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Science*, *8*, 465–471.

Cheour, M., Alho, K., Ceponiene, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O. & Näätänen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology*, *29*, 217–226.

Colin, C., Radeau, M., Soquet, A. & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology*, *115*, 1989–2000.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F. & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*, 495–506.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport*, *8*, 919–924.
Dehaene-Lambertz, G., Dupoux, E. & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635–647.
Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.
Diesch, E., Biermann, S. & Luce, T. (1998). The magnetic mismatch field elicited by words and phonological non-words. *Neuroreport*, *9*, 455–460.
Embick, D., Hackl, M., Schaeffer, J., Kelepir, M. & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Brain Research. Cognitive Brain Research*, *10*, 345–348.
Endrass, T., Mohr, B. & Pulvermüller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, *19*, 1653–1660.
Escera, C., Alho, K., Winkler, I. & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*, 590–604.
Fodor, J.A. (1983). *The modulatity of mind*. Cambridge, MA: MIT Press.
Frege, G. (1980). Über Sinn und Bedeutung (first published in 1892). In G. Patzig, (Ed.), *Funktion, Begriff, Bedeutung* (pp. 25–50). Göttingen: Huber.
Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78–84.
Friederici, A.D., Pfeifer, E. & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*, 183–192.
Friederici, A.D., Wang, Y., Herrmann, C.S., Maess, B. & Oertel, U. (2000). Localization of early syntactic processes in frontal and temporal cortical areas: A magnetoencephalographic study. *Human Brain Mapping*, *11*, 1–11.
Fromkin, V.A. (1973). Introduction. In V.A. Fromkin (Ed.), *Speech errors as linguistic evidence* (pp. 11–45). The Hague, Paris: Mouton.
Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production I* (pp. 177–220). London: Academic Press.
Garrett, M. (1984). The organization of processing structures for language production. In D. Caplan, A.R. Lecours & A. Smith (Eds.), *Biological perspectives on language* (pp. 172–193). Cambridge, MA: MIT Press.
Garrod, S. & Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Science*, *8*, 8–11.
Gunter, T.C., Friederici, A.D. & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*, 556–568.
Hasting, A.S., Kotz, S.A. & Friederici, A.D. (2007). Setting the stage for automatic syntax processing: The Mismatch Negativity as an indicator of syntactic priming. *Journal of Cognitive Neuroscience*, *19*, 386–400.
Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.
Hauk, O. & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Humain Brain Mapping*, *21*, 191–201.
Hebb, D.O. (1949). *The organization of behavior. A neuropsychological theory*. John Wiley, New York.

Honbolygo, F., Csepe, V. & Rago, A. (2004). Suprasegmental speech cues are automatically processed by the human brain: A mismatch negativity study. *Neuroscience Letters*, *363*, 84–88.
Jacobsen, T., Horvath, J., Schroger, E., Lattner, S., Widmann, A. & Winkler, I. (2004). Pre-attentive auditory processing of lexicality. *Brain and Language*, *88*, 54–67.
Kaan, E. & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Science*, *6*, 350–356.
Kircher, T. T., Oh, T. M., Brammer, M. J. & McGuire, P. K. (2005). Neural correlates of syntax production in schizophrenia. *British Journal of Psychiatry*, *186*, 209–214.
Korpilahti, P., Krause, C. M., Holopainen, I. & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, *76*, 332–339.
Korth, M. & Nguyen, N. X. (1997). The effect of stimulus size on human cortical potentials evoked by chromatic patterns. *Vision Research*, *37*, 649–657.
Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C. & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, *13*, 765–772.
Kujala, A., Alho, K., Valle, S., Sivonen, P., Ilmoniemi, R. J., Alku, P. & Näätänen, R. (2002). Context modulates processing of speech sounds in the right auditory cortex of human subjects. *Neuroscience Letters*, *331*, 91–94.
Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.
Lutkenhoner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinstrater, O. & Patterson, R. D. (2003). Localization of primary auditory cortex in humans by magnetoencephalography. *Neuroimage*, *18*, 58–66.
MacKay, D. G. (1987). *The organization of perception and action. A theory of language and other cognitive skills*. New York: Springer.
Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.
Marslen-Wilson, W. & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature*, 257, 784–786.
Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*, 55–74.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.
Menning, H., Zwitserlood, P., Schoning, S., Hihn, H., Bolte, J., Dobel, C., Mathiak, K. & Lutkenhoner, B. (2005). Pre-attentive detection of syntactic and semantic errors. *Neuroreport*, *16*, 77–80.
Mohr, B. & Pulvermüller, F. (2002). Redundancy gains and costs in cognitive processing: Effects of short stimulus onset asynchronies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 1200–1223.
Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, *76*, 165–178.
Moss, H. E., McCormick, S. F. & Tyler, L. (1997). The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes*, *12*, 695–731.
Münte, T. F., Schiltz, K. & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*, 71–73.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear Hear*, *16*, 6–18.

Näätänen, R. (2000). Mismatch negativity (MMN): Perspectives for application. *International Journal of Psychophysiology*, *37*, 3–10.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*, 1–21.

Näätänen, R. & Alho, K. (1995). Mismatch negativity – a unique measure of sensory processing in audition. *International Journal of Neuroscience*, *80*, 317–337.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J. & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Näätänen, R. & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P. & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, *24*, 283–288.

Neville, H., Nicol, J.L., Barss, A., Forster, K.I. & Garrett, M.F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151–165.

Norris, D., McQueen, J.M. & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–370.

Obleser, J., Lahiri, A. & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, *16*, 31–39.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494–522.

Osterhout, L. & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, *14*, 1–14.

Osterhout, L. & Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Osterhout, L., McLaughlin, J. & Bersick, M. (1997). Event-related brain potentials and human language. *Trends in Cognitive Sciences*, *1*, 203–209.

Osterhout, L. & Swinney, D.A. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*, *22*, 273–286.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M. & Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalography and Clinical Neurophysiology*, *78*, 466–479.

Penfield, W. & Rasmussen, T. (1950). *The cerebral cortex of man*. New York: Macmillan.

Penolazzi, B., Hauk, O. & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, *74*, 374–388.

Pettigrew, C.M., Murdoch, B.E., Ponton, C.W., Finnigan, S., Alku, P., Kei, J., Sockalingam, R. & Chenery, H.J. (2004). Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear Hear*, *25*, 284–301.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behaviorial and Brain Sciences*, *27*, 169–190; discussion 190–226.

Picton, T. & Hillyard, S. (1974). Human auditory evoked potentials: II. Effects of attention. *Electroencephalography and Clinical Neurophysiology*, *36*, 191–200.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K. & Marantz, A. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Research. Cognitive Brain Research*, *4*, 231–242.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–336.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, *5*, 517–524.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576–582.

Pulvermüller, F. & Assadollahi, R. (2007). Grammar or serial order? Discrete combinatorial brain mechanisms reflected by the syntactic mismatch negativity. *Journal of Cognitive Neuroscience*, *19*, 971–980.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J. & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *Neuroimage*, *14*, 607–616.

Pulvermüller, F. & Shtyrov, Y. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *Neuroimage*, *20*, 159–172.

Pulvermüller, F. & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*, 49–71.

Pulvermuller, F., Shtyrov, Y., Hasting, A. S. & Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, *104*, 244–253.

Pulvermüller, F., Shtyrov, Y. & Ilmoniemi, R. (2003). Spatiotemporal dynamics of neural language processing: An MEG study using minimum-norm current estimates. *Neuroimage*, *20*, 1020–1025.

Pulvermüller, F., Shtyrov, Y. & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience*, *17*, 884–892.

Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R. J. & Marslen-Wilson, W. D. (2006). Tracking speech comprehension in space and time. *Neuroimage*, *31*, 1297–1305.

Pulvermüller, F., Shtyrov, Y., Kujala, T. & Näätänen, R. (2004). Word-specific cortical activity as revealed by the mismatch negativity. *Psychophysiology*, *41*, 106–112.

Pylkkanen, L., Stringfellow, A. & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language*, *81*, 666–678.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D. & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*, 507–537.

Schröger, E. (1996). On detection of auditory diviations: A pre-attentive activation model. *Psychophysiology*, *34*, 245–257.

Sereno, S. C., Brewer, C. C. & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, *14*, 328–333.

Sereno, S. C. & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Science*, *7*, 489–493.

Shestakova, A., Brattico, E., Soloviev, A., Klucharev, V. & Huotilainen, M. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Brain Research. Cognitive Brain Research*, *21*, 342–350.

Shtyrov, Y., Hauk, O. & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words. *European Journal of Neuroscience*, *19*, 1083–1092.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J. & Näätänen, R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neuroscience Letters*, *251*, 141–144.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J. & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *Neuroimage*, *12*, 657–663.

Shtyrov, Y., Pihko, E. & Pulvermüller, F. (2005). Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? *Neuroimage*, *27*, 37–47.

Shtyrov, Y. & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport*, *13*, 521–525.

Shtyrov, Y. & Pulvermüller, F. (2007). Early MEG activation dynamics in the left temporal and inferior frontal cortex reflect semantic context integration. *Journal of Cognitive Neuroscience*, *19*, 1633–1642.

Shtyrov, Y., Pulvermüller, F., Näätänen, R. & Ilmoniemi, R. J. (2003). Grammar processing outside the focus of attention: An MEG study. *Journal of Cognitive Neuroscience*, *15*, 1195–1206.

Sittiprapaporn, W., Chindaduangratn, C., Tervaniemi, M. & Khotchabhakdi, N. (2003). Preattentive processing of lexical tone perception by the human brain as indexed by the mismatch negativity paradigm. *Annals of the New York Academy of Science*, *999*, 199–203.

Stockall, L., Stringfellow, A. & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language*, *90*, 88–94.

Tiitinen, H., May, P. & Näätänen, R. (1997). The transient 40-Hz response, mismatch negativity, and attentional processes in humans. *Progress in Neuropsychopharmacology and Biological Psychiatry*, *21*, 751–771.

Tiitinen, H., May, P., Reinikainen, K. & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*, 90–92.

Tyler, L., Moss, H. E., Galpin, A. & Voice, J. K. (2002). Activating meaning in time: The role of imageability and form-class. *Language and Cognitive Processes*, *17*, 471–502.

Weber, C., Hahne, A., Friedrich, M. & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. *Brain Research. Cognitive Brain Research*, *18*, 149–161.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R. J. & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*, 638–642.

Woods, D. L., Alho, K. & Algazi, A. (1993a). Intermodal selective attention: Evidence for processing in tonotopic auditory fields. *Psychophysiology*, *30*, 287–295.

Woods, D. L., Knight, R. T. & Scabini, D. (1993b). Anatomical substrates of auditory selective attention: Behavioral and electrophysiological effects of posterior association cortex lesions. *Brain Research. Cognitive Brain Research*, *1*, 227–240.

Yamasaki, H., LaBar, K. S. & McCarthy, G. (2002). Dissociable prefrontal brain systems for attention and emotion. *Proceedings of the National Academy of Sciences of the USA*, *99*, 11447–11451.

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J. & Courtney, S. M. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience*, *5*, 995–1002.
Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64.

Modelling brain activity understanding center-embedded sentences

Bengt Sigurd
Department of Linguistics, Lund University

Bengt Nilsson
Department of Neurology, Lund University

Abstract

The present paper presents grammatical analyses of complex sentences, which demonstrate their characteristics and suggest what the brain has to handle in processing them. Previous brain imaging studies as well as the authors' Eye-Tracking experiments are reported. The importance of the restricted working memory, visual attention and eye movements during reading is stressed and the split of the subject–predicate connection is suggested as a major source of difficulty. Two computer programs modelling different parsing strategies are presented in an appendix.

1. Introduction

Modern imaging techniques have generated new knowledge about the higher brain functions and possibilities of mapping anatomical and functional patterns. The left frontal lobe, in particular the Broca area, is clearly involved in language processing and complex sentences are assumed to put a great load on verbal working memory (vWM). The visual attention and eye movements during reading of complex sentences may also shed light on the processes of reading and understanding.

As particularly conspicuous examples of complexity, *center-embedded sentences* (CES), have long interested linguists and psycholinguists ([1]).

[1] The rat that the cat that the dog chased bit squeaked.

The present paper presents a computerized generalized phrase structure grammar (GPSG) which can analyze such sentences, pinpoint their complications, and help to predict their difficulties. We also report from some preliminary Eye-Tracking experiments. The results show intensive movements over the middle of the sentences and processing problems increasing with the number of embeddings where the subject–predicate connection is split and the distance subject–predicate and predicate–object is great, cf. Gibson (1998).

2. Center-embedded sentences

Figure 1 shows a tree diagram of the Swedish equivalent to [1]. Note the nested relative clauses (relc) and the final series of three verbs (vt, vt, vi). CES may involve other syntactic categories; an adverbial center-embedding is *Om dig har jag då du då jag nös fått en ofördelaktig uppfattning* '(lit.) About you have I as you as I sneezed laughed got an unfavourable impression.'

2.1. Why CES are difficult

Various theories have been presented to explain the processing problems of embedded sentences. Explanations suggested include (see References):

1. A clause cannot be processed in the same clause type (self-embedding).
2. The brain has to process long noun-verb patterns, e.g. N1 N2 N3 V1 V2 V3 and combine the N, V (as subject, predicate, object) correctly which overloads vWM.
3. Several (pending) subjects without immediate predicates are difficult.

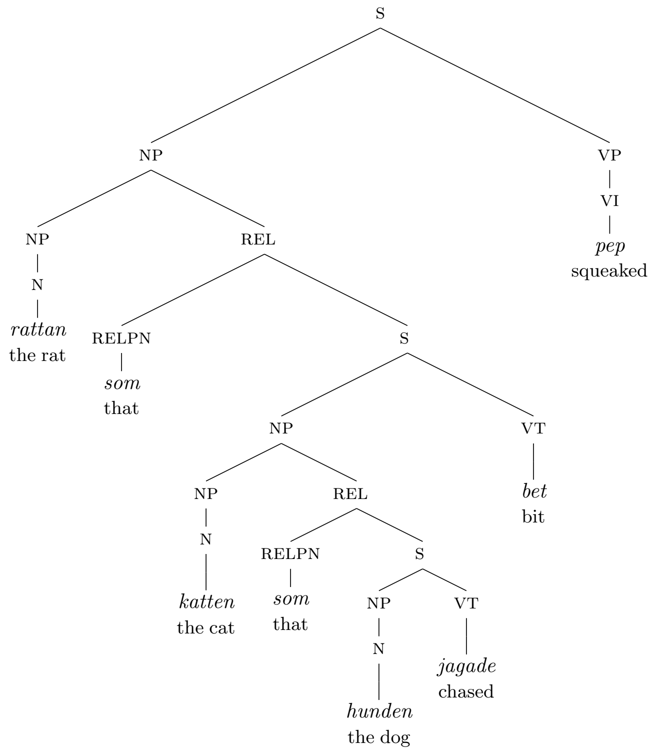


Figure 1. Center-embedded sentence *Råttan som katten som hunden jagade bet pep* 'The rat that the cat that the dog chased bit squeaked'.

4. Several predicates at the end cause problems: which predicate belongs to which subject?
5. Every subject–predicate split by another subject–predicate increases difficulty. Great distances subject–predicate and predicate–object cause parsing problems.

2.2. Brain activation and CES

Stromsvold et al. (1996) studied the brain activation difference between the sentence *The juice that the child spilled stained the rug* and the sentence *The child spilled the juice that stained the rug*. The first sentence took 3.00 s to perceive and the second sentence 3.40 s. The grammatical difference stated in the paper is that the second sentence is right branching while the first is center-embedded. The second sentence could also be defined as not center-embedded. Furthermore, the relative clause in the second sentence is a subject relativizing clause and such clauses are known to be easier to process than object relativizing clauses, where the object has to be found at a distance in the correlate before the clause. Embedded sentences with pronominal subjects instead of full noun phrases are also known to be easier to process (Warren & Gibson, 2002). The PET technique employed by Stromswold et al. (1996) shows specific rCBF (regional Cerebral Blood Flow) for center-embedding minus rCBF right branching) activation of the opercular portion of Broca's area.

3. Grammatical analysis of CES

For automatic grammatical analysis of CES, we have implemented a Generalized Phrase Structure Grammar (GPSG) in the programming language Prolog. Our GPSG rules, after Sigurd (1978), are more complex than Chomsky's in his classical Syntactic Structures (1957), as they are enriched by information to the left of the arrow where functional (S) and phrase structure (P) representations are built. The grammar is called by the command `esent(S, P, X, [])` which processes X and binds a functional (subject–predicate) representation to S and a phrase structure representation to P (Figure 2).

A complexity value can be calculated by the grammar on the basis of the use of center-embedded rules such as relative clause rules and adverbial clause rules.

```
esent([subj(Np), pred(V)],[np(Nx),vi(Vx)]) --> enp(Np,Nx),evifin(V,Vx).

% rule causing self embedding
enp([N,Rs],[n(N),rels(Rf)]) --> en(N,Nf), [that],erdef(N,Rs,Rf).


% sample run
?- esent(S, P,
 [the, rat, that, the, cat, that, the, dog, chased, bit, squeaked], []).

S = [subj([therat, [subj([thecat, [subj(thedog), pred(chased),
 obj(thecat)]]), pred(bit), obj(therat)]]), pred(squeaked)],

P = [np([n(therat), rels([that, [n(thecat), rels([that, n(thedog),
 vt(chased)])], vt(bit)])]), vi(vi(squeaked))]
```

Figure 2. Sample GPSG rules and a sample run. S is subject–predicate representation, P is phrase representation.

4. Eye movements in CES reading

Several imaging studies (PET and fMRI) showing that the processing of complex sentences induces hemodynamic responses specifically in the opercular portion of Broca's area have supported the hypothesis of a load on verbal working memory (vWM). The concept of WM may be based on several models, but WM is usually considered to have a limited capacity of 7 ± 2 units ("the magical number 7", Miller & Chomsky, 1963). Center-embedded sentences seem useful for the study of vWM.

In order to further evaluate the possible load of center-embedded sentences on WM including attentional and visual mechanisms, we have undertaken a preliminary study of eye movements during reading of different types of center-embedded sentences. The study was carried out at the Humanities' Laboratory, Lund University.

4.1. Method

Ten sentences, five grammatically correct, five incorrect, were presented to five "naive" students. Eye movements were registered during the time the participants took to reach the judgment correct/incorrect.

The diagrams in Figures 3–5 show individual recordings of saccadic eye movements/fixations during reading and attempting to understand center-embedded sentences.

Circles mark fixation points during reading – a bigger circle represents longer time. The lines between circles are saccadic movements. Green circles mark fixation points during the first second. A blue circle marks the end of eye movements.

4.2. Results

The sentence *Av segling hade Karl då han då han var ung bodde på landet ingen erfarenhet* '(lit.) Of sailing had Karl as he as he was young lived in the countryside no experience', with three pending subjects (Karl, he, he) and the pending prepositional phrase *Av segling* belonging to *erfarenhet*, was the most difficult (Figure 3a); it was marked as correct by one subject only. Note the great eye activity over *då han då han* 'as he as he.' The attention scan diagram (Figure 3b) shows that the eyes focused on the middle of the sentence with the adverbial clauses and the two subjects *han* 'he' the most. The colour scale goes from blue to red.

The correct sentence *Bollen som Karl då Per missade tog gick i mål* '(lit.) The ball that Karl as Per missed took hit the goal' (Figure 4) was marked as correct by 3 out of the 5 subjects.

The correct sentence *Glaset som kvinnan som ägde hunden tappade sprack* '(lit.) The glass that the woman who owned the dog dropped split' (Figure 5) with only two pending subjects (*glaset* 'the glass', *kvinnan* 'the woman') was understood and marked as correct by all participants.

Figure 3a. Visualization of Eye-Tracking data from a reading of the sentence *Av segling hade Karl då han var unge bodde på landet ingen erfarenhet* '(lit.) Of sailing had Karl as he as he was young lived in the countryside no experience'.

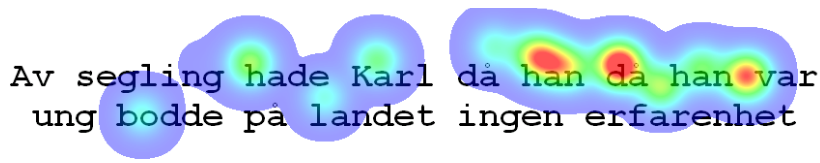


Figure 3b. Attention scan diagram of the sentence in 4a.

Figure 4. Visualization of Eye-Tracking data from a reading of the sentence *Bollen som Karl då Per missade tog gick i mål* '(lit.) The ball that Karl as Per missed took hit the goal'.

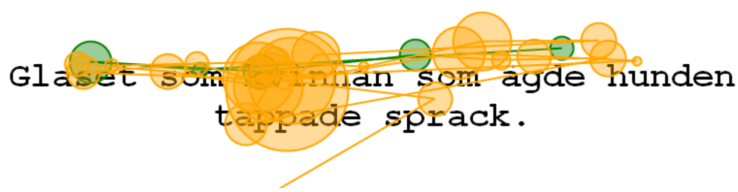


Figure 5. Visualization of Eye-Tracking data from a reading of the sentence *Glaset som kvinnan som ägde hunden tappade sprack* 'The glass that the woman who owned the dog dropped split'.

5. Conclusions

As a pilot study, the experiment was performed only on a small number of subjects, using very few sentences. Its value lies first and foremost in implying hypotheses. Nevertheless, there are reasons to comment on the preliminary results.

The incorrect sentences were identified in almost all readings. The findings indicate that an increase in syntactic complexity by more pending constituents and subject–predicate splits induce a marked increase in working memory load, as evidenced by eye movements and negative judgements. In contrast to eye movement patterns during reading of common text, our recordings show high antero- and retrograde activities with generally more than twenty saccades/fixations per sentence. The subjects' reaction times varied a lot and it is not possible to find out whether wrong or correct sentences were generally identified faster.

Following Gibson (1998), processing difficulty may be estimated by the number of words between subject and predicate plus the number of words between object and predicate. The existing data does at least not contradict this idea: the number of correct answers increases as the subject–predicate and predicate–object distance decreases.

The Appendix presents two computer models which are inspired by the eye movements. WMparse is based on the idea of a restricted working memory where the sequence of words is stored while the operations assigning the words to functional roles are performed. The distances between related functional roles are calculated to indicate the difficulty of the sentence. The program Nexsucc reflects the idea that the innermost clause is first identified, then the next, etc.

6. Acknowledgements

We are grateful to Richard Andersson, Humanities' Laboratory, Lund University, for practical assistance.

References

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gibson, E., Desmet, T., Grodner, D., Watson, D. & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics, 16*, 313–53.

Gouvea, A. C. (2003). *Processing syntactic complexity: Cross-linguistic differences and ERP evidence*. PhD Dissertation, University of Maryland, College Park.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics, 43*, 365–392.

Miller, G. & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce et al. (Eds.), *Handbook of mathematical psychology* (pp. 419–491). New York: Wiley.
Roll, M., Frid, J. & Horne, M. (2007). Measuring syntactic complexity in spontaneous spoken Swedish. *Language and Speech 50*(2), 227–245.
Sigurd, B. (1978). Referent Grammar (RG). A generalized phrase structure grammar with built-in referents. *Studia Linguistica 41*(2), 115–135.
Stromswold, K., Caplan, D., Alpert, N. & Rauch, S. (1996). Localization of syntactic comprehension by Positron Emission Tomography. *Brain and Language, 52*(3), 452–473.
Warren, T. & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*, 79–112.

Appendix

6.1. WMPARSE

Program modelling the use of restricted working memory and operations which assign functional roles to the categories N(oun), V(erb).

The total functional distance between subj(ect)–pred(icate), pred(icate)–obj(ect) in the clauses is calculated. It indicates the difficulty of processing.

```
% Functional dist:N-V2:2+ N-V1:1. Total:3
func1(X,Y) :-
 X=[N,that,V1,V2], % 1 itr rel clause recognized
 n(N),vi(V1),vi(V2), % gram categories
 Y=[subj(N),pred(V2),subj(N),pred(V1)],
 print(3),nl.

% Functional dist: N1-V3: 6 + N2-V2:3 +V2-N1:5+V1-N2:2. Total = 16
func2(X,Y) :-
 X=[N1,that,N2,that,N3,V1,V2,V3], % 2 tr obj rel clauses
 n(N1),n(N2),n(N3),vi(V3),vt(V1),vt(V2),
 Y=[subj(N1),pred(V3),subj(N2),pred(V2),obj(N1),subj(N3),pred(V1),obj(N2)],
 print(16),nl.


%sample run

?- func1(X, Y)
X = [dog, that, dog, bit, ran],
Y = [subj(dog), pred(ran), subj(dog), pred(bit), obj(dog)]

?- func2([boy, that, dog, that, cat, chased, bit, ran], Y)
Y = [subj(boy), pred(ran), subj(dog), pred(bit), obj(boy), subj(cat),
 pred(chased), obj(dog)]
```

6.2. NEXSUCC

Program which “understands” sentences such as the Swedish *Råttan som katten som hunden jagade pep* ‘The rat that the cat that the dog chased squeaked’, in the sense that it identifies the subject–predicate combination (nexus) successively starting from the innermost clause. The eye movement studies indicated intensive focusing over the middle of the sentences.

```
nex(X,Y) :-
 % searching for N1, som, N2,V
 % I is any initial part, F is any remaining final part
 append(I,[N1,som,N2,V|F],X),

 % checking that N1 and N2 are nouns and V is a verb
 n(N1),n(N2),v(V),

 % prints sentence with N2 as subject, N1 as object and V as predicate
 print([N2,V,N1]),nl,

 % adds N1 to the initial part I, constructing I2
 append(I,[N1],I2),

 % appends I2 to F resulting in a new string to which the
 % rule can be applied again
 append(I2,F,Y).

% extracts an intransitive sentence
nex(X,Y) :- X=[N1,V], n(N1),v(V),R=[N1,V],print(R),nl,Y=R.



% sample run

?- nex([råttan,som,katten,som,hunden,jagade,bet,pep],Y),nex(Y,Z),nex(Z,W)

Y = [hunden,jagade, katten]
Z = [katten, bet, råttan]
W = [råttan, pep]
```

Hemifield asymmetries in parafoveal word processing

Jaana Simola, Kenneth Holmqvist
Humanities Laboratory, Lund University

Magnus Lindgren
Department of Psychology, Lund University

Abstract

In reading, both foveal (currently fixated) and parafoveal words (outside the current eye fixation) are important. Having only the fixated word available slows reading, but when the next word is available, reading is almost as fast as when the whole line is seen.

The region of effective vision, the perceptual span, sets the outer limits for parafoveal processing. For readers of Western languages, the perceptual span extends 3 to 4 letters to the left and up to 15 letters to the right of current fixation. An opposite bias is observed for readers of right-to-left script.

Previous research has shown that words presented to the right visual hemifield are recognized faster. This article discusses three theories that have been suggested for the hemifield differences in word recognition. Furthermore, we introduce an insightful method for investigating the hemifield differences in parafoveal information extraction.

1. Introduction

During an eye fixation in reading, the area from which useful information can be acquired is limited, extending about 14–15 letter spaces to the right of fixation and 3–4 letter spaces to the left of fixation (Rayner, 1998). A rightward bias in the area of effective vision, i.e. the *perceptual span*, has been observed for readers of Western languages, whereas Hebrew readers exhibit a leftward bias in their perceptual span (Pollatsek et al., 1981). This asymmetry suggests that linguistic material on one side of the eye fixations is processed more than on the other side. Therefore, in written language processing, it matters on which side of the visual field the words appear. Most previous studies suggest a right visual field (RVF) advantage in word recognition (for a review, see *Brain & Language*, 2004, *88*, pp. 259–370). Three explanations have been proposed for these hemifield asymmetries. *Hemispheric asymmetries, attentional bias,* or long-term *perceptual learning* effects from reading in a specific direction are considered to account for the way in which information is acquired from either side of an eye fixation.

Another critical factor in reading is that words are not processed in isolation, but rather information from parafoveal words is extracted while the current word is still fixated. According to the E-Z reader model (Pollatsek et al., 2006; Reichle et al., 2006), attention moves ahead of the currently fixated word after it has been identified, resulting in parafoveal preview benefits when the parafoveal words are subsequently fixated. This shifting of attention provides the text comprehension system with a strictly ordered sequence of successive words similar to auditory speech. Moreover, parafoveal-on-foveal effects (Kennedy & Pynte, 2005; Kennedy et al., 2002) occur when the processing of the currently fixated word is affected by the parafoveal words.

Hemifield asymmetries in word recognition and the parafoveal information extraction during reading have been widely examined separately. However, relatively little is known about hemifield asymmetries in parafoveal information usage. This article discusses the theoretical background and importance of studying hemifield asymmetries in parafoveal processing. Furthermore, results from an experiment using a technique that combines eye movement and electroencephalograpy (EEG) recordings are discussed.

2. Hemifield asymmetries in word recognition

2.1. The hemispheric asymmetries account

The standard model of the neural background of reading postulates that visual information is initially processed by occipitotemporal areas contralateral to the stimulated visual hemifield. From there, the information is transferred to the visual word form (VWF) system in the left fusiform gyrus that is specifically devoted to the processing of printed words (Cohen et al., 2000; Cohen et al., 2002). According to the *hemispheric asymmetry* account, the words in the RVF are processed faster, because the RVF initially projects to the left cerebral hemisphere that is dominant for language processing for most individuals. The information presented in the left visual field (LVF) projects initially to the right hemisphere and must be transferred to the left hemisphere via a longer transcallosal pathway. The time needed for transferring information via the splenium of the corpus callosum is referred to as an interhemispheric transfer time which has been estimated to take around 10–15 ms. Due to faster conducting cortico-cortical myelinated axons in the right hemisphere, the right-to-left transfer is faster than the left-to-right transfer in right-handed participants (see Martin et al., 2007). With letter string stimuli, the ipsilateral information transfer time ranged from 26 to 42 ms (Martin et al., 2007), and around 180–200 ms post-stimulus, the information from both hemispheres has reached the visual word form area (Cohen et al., 2000).

Previous studies have consistently shown a RVF advantage for English readers, but the results concerning readers of right-to-left scripts such as Hebrew or Yiddish have been less clear (see Nazir et al., 2004). Some studies have shown a LVF advantage for right-to-left readers, whereas in other studies, only those readers who had learned Yiddish as their first and native tongue showed the LVF advantage. Moreover, in some studies, the LVF advantage was observed only for left-handed Hebrew readers while the right-handed readers were better for words presented in the RVF. The differences between left- and right-handed readers was taken to reflect functional differences between the two hemispheres, and to date most researchers have accepted that the visual field asymmetries in printed word recognition are primarily due to functional differences between the two cerebral hemispheres.

2.2. The attentional bias account

Covert attention improves discriminability in a variety of visual tasks and also accelerates the rate of visual information processing (Carrasco & McElree, 2001). According to the *attentional bias* account, the hemifield differences result from biased distribution of attention across the visual field. The attention theory proposes that left-to-right readers process the words presented in the RVF faster, because their attention during reading moves towards the right visual field. This assumption is supported by the serial attention shift models of reading. According to the original work by Morrison (1984), attention begins at the point of fixation in order to allow detailed processing of the foveal object, but once the foveal processing reaches some threshold of information processing, attention moves ahead to the next position. The eye then follows the attention to the new position once the motor programming of a saccadic eye movement is accomplished.

Henderson et al. (1989) studied the influence of covert attention on extrafoveal information acquisition with a setup comprising four objects presented at the corners of an imaginary square. Their primary interest was whether all areas of the visual field beyond the fovea are equally available for the analysis prior to a saccadic eye movement. They showed that a substantial amount of information was acquired from extrafoveal locations even when the participants had to identify foveal objects. However, the extrafoveal information extraction was restricted to locations that were about to be fixated next. Their results supported the serial model of attention allocation, in which the extrafoveal preview is derived only from the single position to which attention has moved. Since the extrafoveal preview is derived from any location that is about to be fixated, the attention account predicts no permanent differences between left and right visual hemifields.

Inhoff et al. (1989) compared reading of normal and spatially transformed (mirror reversed) text in order to examine the allocation of covert attention in reading. The sequence of letters within words was either congruent or incongruent with the sequence of words in sentences and with the reading direction, which was either from left to right or from right to left. They also varied the amount of visible text so that in some conditions only the fixated word and the words preceding it were visible, whereas in other conditions the fixated word and the succeeding words were both visible. Their results showed that reading was faster and readers were able to extract more parafoveal information when the words were normal than when the letters were mirror reversed. They also found that the congruity of the word and letter order had no effect on readers' ability to extract parafoveal information, indicating that covert attention does not work on a letter-by-letter basis, but rather the letters within a word are processed in parallel inside an asymmetric spotlight. While reading normal text in an unusual direction (from right to left), readers were about 25 ms slower moving their eyes, the forward saccades were shorter and the words received more (approximately 0.2 times) fixations compared to the normal left-to-right reading, reflecting reading difficulties in an unusual reading direction. Interestingly, there were little differences in parafoveal information extraction between left-to-right and right-to-left reading of the normal text conditions. This result suggested that readers were able to extract an equal amount of parafoveal information to the left and to the right of fixation, supporting the conclusion that covert attention is closely locked to the eye movements. The authors also concluded that the perceptual span is not constant, but rather that when readers of English were required to read from right-to-left, they were able to acquire more information to the left of fixation.

2.3. The perceptual learning account

The *perceptual learning* account suggests that left-to-right readers process the RVF words faster because they have learned to acquire useful information from the RVF. Although learning usually involves higher order cortical areas, perceptual learning can occur at early stages of visual information processing (Gilbert et al., 2001) and provides a means for rapid and efficient stimulus recognition. Psychophysical studies with non-verbal stimuli have shown that target discrimination can improve dramatically with continuous training (e.g. Sigman & Gilbert, 2000). However, this improvement seems to depend on the specific stimulus location, configuration, and context during training. For example, an improvement related to a stimulus composed of lines does not transfer to a stimulus composed of dots or to a stimulus at another retinal location that was not used during training. This lack of generalization suggests that perceptual learning involves early stages of cortical processing where retinotopic organization and orientation tuning are still refined (Gilbert et al., 2001).

According to Nazir et al. (2004), the visual training associated with the regularity of reading eye movements provides a necessary condition for development of location-specific pattern memories for letter stimuli. The statistics about where the eyes land in words during natural reading show a very stable pattern. In left-to-right reading, this landing site distribution is skewed with a maximum slightly to the left of the word center (Hunter et al., 2007; McConkie et al., 1988; Vitu et al., 1990) while for scripts read from right-to-left, the maximum is to the right of the word center (Deutsch & Rayner, 1999). The location where the eyes land within a word is mainly determined by low-level oculomotor factors that are relatively independent of the text content and the on-going linguistic processing. Given that during reading, most words are fixated only once (Rayner & Pollatsek, 1989), the landing site distribution provides an estimation of the relative frequency with which words are processed from different locations on the retina. Nazir et al. showed that for French readers, the viewing position (fixation zone within a word) affected word recognition in such a way that the lexical decision accuracy was best at locations that are used most during natural reading, whereas the viewing position had no effect on the processing of non-words of equal length. Furthermore, Nazir et al. compared letter recognition performance for Roman and Hebrew scripts that are read in opposite directions, and showed that visual hemifield effects varied with script.

The findings by Nazir et al. supported the assumption that factors additional to the hemispheric differences or the uneven distribution of attention are needed to explain visual hemifield dependent variations in the perception of print. This is because the neural network of reading also comprises regions prior to the visual word form area. Depending on the characteristics of a specific writing system, the landing site distribution of saccades within words possibly modifies the early stages of the visual pathway in such a way that word recognition improves in regions of the retina that fall on the side of reading direction. Thus, the reading-related visual training constrains the way in which we perceive printed words (Nazir et al., 2004).

3. Intake of information from the parafovea

The written language signal contains a multitude of static symbols presented at the same time. In order to extract the text content, the reader must actively follow word order by moving her eyes along the print. At least a partial analysis of the extrafoveal words that are not currently fixated is obtained, and provides information that speeds the analysis of the words when they are subsequently fixated. A *moving window* technique is often used to study the amount of benefit derived from extrafoveal preview of a word. This technique utilises on-line eye movement recording in order to present text on a computer-controlled display contingent upon where the reader is fixating. The text characteristics are mutilated except for an experimenter-defined window region around the reader's fixation. Each time the reader moves her eyes, a new region of the display is exposed and a different region is mutilated.

Another often used technique for studying extrafoveal information extraction is the *invisible boundary* technique, where the preview is changed after the reader's eyes have crossed an invisible boundary embedded in a text. An estimate of the amount of benefit is obtained through a comparison of performance when the preview is identical with the stimulus that is seen after crossing the boundary with the performance when the preview stimulus is changed after the boundary. Previous research suggests that the subsequent processing of the parafoveally previewed words benefits for example from orthographic (e.g. word vs. non-word) and context-related information (Inhoff et al., 2000). However, the usefulness of a parafoveal preview depends on the ease with which the directly fixated word is processed. As Henderson & Ferreira (1990) have shown, the parafoveal preview is reduced when the fixated word is difficult and requires extra processing resources. An alternative possible explanation for the observation that preview advantage changes as a function of foveal load is the idea that foveal and parafoveal words are processed in parallel.

Parallel processing models can predict so-called parafoveal-on-foveal effects, indicating that the processing of a currently fixated word is affected by the properties of a following word (Kennedy, 2000; Kennedy & Pynte, 2005; Kennedy et al., 2002). What information from the parafovea can then influence the viewing times of the currently fixated word? Kennedy (1998) obtained effects of parafoveal word frequency and length on the processing of the currently fixated word in a word comparison task. However, in a normal reading task, neither the frequency nor the length of the parafoveally visible word affected the time spent viewing the fixated word, but the current fixation times were influenced by the visuospatial (e.g. UPPERCASE vs. lowercase) and orthographic (e.g. word vs. non-word) information of the following word (Inhoff et al., 2000). Moreover, Inhoff et al. showed that the meaning of the following word affected processing of the foveal word when the current fixations landed closer to the end of the word and near the following word.

In summary, the serial attention shift models of reading account well for the parafoveal preview benefits, whereas only the models accepting parallel processing of words can predict parafoveal-on-foveal effects, suggesting that parafoveally available information can influence processing of the foveal word.

4. EFRPs and parafoveal processing

Baccino & Manuta (2005) used a technique that combines the recordings of eye movements and event-related potentials (ERPs) to study the parafoveal-on-foveal effects. They presented participants with two words – a prime word in the middle of a screen and a target to the right of the prime – and measured eye-fixation-related potentials (EFRPs) while the participants were fixating the foveal prime word. Properties of the target (parafoveal) word were manipulated, so that the target either was semantically related or unrelated to the prime, or was a non-word. Their results suggested that the semantic relatedness between the prime and target word influenced a P2 ERP component, suggesting that semantic information was acquired from the parafovea during foveal processing.

In order to examine the hemifield differences in parafoveal processing, we (Simola et al., submitted) used an otherwise similar paradigm to that of Baccino & Manuta, but extended it by presenting the target words either to the left or to the right of the foveal word. By doing so, we were able to measure the visual hemifield differences in parafoveal processing. Previous research has addressed either the parafoveal information extraction or hemifield differences in word recognition separately, while little is known about parafoveal information extraction at distinct locations in the visual field. So far, only studies related to the attention theory of hemifield asymmetries have looked at the parafoveal information extraction at different parts of the visual field. Since these studies suggest that parafoveal information is extracted only from the single location that is about to be fixated, they assume no differences between distinct visual field locations. The studies suggesting hemispheric asymmetries or long-term perceptual learning effects as explanations for the hemifield differences have usually looked at word recognition of laterally or centrally presented words while, to the best of our knowledge, parafoveal information extraction has not been previously studied in relation to the location of the parafoveal stimuli in the visual field. Investigating these effects, however, may help in clarifying the mechanism of the hemifield asymmetries in the word recognition process. Moreover, these effects may also reveal important aspects concerning the neural background of the parafoveal information processing during reading.

5. Results and conclusions

Our current research (Simola et al., submitted) addresses the differences in parafoveal processing between left and right visual fields by measuring EFRPs while the foveal word is fixated. Preliminary results show that during foveal processing, parafoveal information related to the orthography of extrafoveal word is extracted, that is, the orthographically legal words are differentiated from illegal non-words after 200 ms from the fixation onset. This is indicated by smaller responses of a bilateral occipital P2 EFRP component for orthographically illegal non-words. Interestingly, the parafoveal information about the orthography of the next word is obtained only from the right visual field, indicating that either hemispheric asymmetries or perceptual learning would account for the observed visual field difference.

The hemispheric asymmetry account suggests that visual hemifields differ because the cortical structures underlying language processing are lateralized in the left hemisphere. Therefore, word information from the left visual field (the right hemisphere) must be transferred to the left hemisphere and to the visual word form area, a system located in the left inferior temporal region (Cohen et al., 2000; Cohen et al., 2002). According to previous research, this information transfer occurs by 180–200 ms post-stimulus; after that point, the processing states are assumed to be identical irrespective of the stimulated hemifield. However, the findings in Simola et al. (submitted) show that hemifield differences only begin to occur around 200 ms after stimulus onset. Moreover, the P2 EFRP component differentiating between non-word and word responses for RVF targets occurs bilaterally, and shows no difference between the cerebral hemispheres, whereas an earlier occipital negativity between 140 and 200 ms post-stimulus is emphasized over the left hemisphere and most likely corresponds to processing at the visual word form area. These findings are consistent with the perceptual learning account, suggesting that long-term reading in a certain direction improves word recognition in that direction. The results further show that parafoveal information relevant for reading was primarily obtained from the right visual field, i.e. from the normal reading direction of the participants, while no difference between word conditions was found when identical words were presented in the left visual field.

6. Acknowledgements

This work was supported by Helsingin Sanomat Centennial Foundation, Wihuri Foundations, and NordForsk. The authors would like to thank Richard Andersson and Marcus Nyström for technical support, Gerd Waldhauser, Petter Kallioinen, Sverker Sikström and Mikael Roll for help in stimulus preparations, and Devin Terhune and Anne-Cécile Treese for help with the statistics.

References

Baccino, T. & Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology, 19*(3), 204–215.

Carrasco, M. & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proc. Nat. Acad. Sci., 98*(9), 5363–5367.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., et al. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*, 291–307.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S. & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain, 125*, 1054–1069.

Deutsch, A. & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes, 14*, 393–421.

Gilbert, C. D., Sigman, M. & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron, 31*, 681–697.

Henderson, J. M. & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 417–429.

Henderson, J. M., Pollatsek, A. & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & Psychophysics, 45*(3), 196–208.

Hunter, Z. R., Brysbaert, M. & Knecht, S. (2007). Foveal word reading requires interhemispheric communication. *Journal of Cognitive Neuroscience, 19*(8), 1373–1387.

Inhoff, A. W., Pollatsek, A., Posner, M. I. & Rayner, K. (1989). Covert attention and eye movements in reading. *Quarterly Journal of Experimental Psychology, 41 A*, 63–89.

Inhoff, A. W., Starr, M. & Shindler, K. L. (2000). Is the processing of words during eye fixations in reading strictly serial? *Perception & Psychophysics, 62*(7), 1474–1484.

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: Evidence for a processing trade-off. In *Eye guidance in reading and scene perception* (pp. 149–179). Amsterdam: Elsevier Science Ltd.

Kennedy, A. (2000). Parafoveal processing in word recognition. *The Quarterly Journal of Experimental Psychology, 53A*(2), 429–455.

Kennedy, A. & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research, 45*(2), 153–168.

Kennedy, A., Pynte, J. l. & Ducrot, S. p. (2002). Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology: Human Perception and Performance, 55A*, 1307–1337.

Martin, C. D., Thierry, G., Démonet, J.-F., Roberts, M. & Nazir, T. A. (2007). ERP evidence for the split fovea theory. *Brain Research, 1185*, 212–220.

McConkie, G. W., Kerr, P. W., Reddix, M. D. & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research, 28*, 1107–1118.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 667–682.

Nazir, T. A., Ben-Boutayab, N., Decoppet, N., Deutsch, A. & Frost, R. (2004). Reading habits, perceptual learning, and recognition or printed words. *Brain and Language, 88*, 294–311.

Pollatsek, A., Bolozky, S., Well, A. D. & Rayner, K. (1981). Asymmetries in the perceptual space for Hebrew readers. *Brain and Language, 14*, 174–180.

Pollatsek, A., Reichle, E. D. & Rayner, K. (2006). Tests of the E-Z reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*, 1–56.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, New Jersey: Prentice-Hall.

Reichle, E. D., Pollatsek, A. & Rayner, K. (2006). E–Z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*, 4–22.

Sigman, M. & Gilbert, C. D. (2000). Learning to find a shape. *Nature neuroscience, 3*(3), 264–269.

Simola, J., Holmqvist, K. & Lindgren, M. (submitted). Right visual field advantage in parafoveal processing: Evidence from eye-fixation related potentials. *Brain and Language.*

Vitu, F., O'Regan, J. K. & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics, 47*(6), 583–600.

Clinical aspects

Cognitive subtypes of dyslexia[1]

Stefan Heim
Department of Psychiatry and Psychotherapy
RWTH Aachen University

Katrin Amunts, Marcus Wilms, Simone Vossel, Elisabeth Bay
Institute for Neurosciences and Biophysics (INB3-Medicine), Research Centre Jülich

Julia Tschierse, Marion Grande, Walter Huber
Dept. of Neurolinguistics, RWTH Aachen University

Klaus Willmes
Dept. of Neuropsychology, RWTH Aachen University

Anna Grabowska
Dept. of Psychophysiology, Nencki Institute of Experimental Biology, Warsaw

Abstract

Different theories conceptualise dyslexia as either a phonological, attentional, auditory, magnocellular, or automatisation deficit. Such heterogeneity suggests the existence of yet unrecognised subtypes of dyslexics suffering from distinguishable deficits. The purpose of the study was to identify cognitive subtypes of dyslexia. Out of 642 children screened for reading ability 49 dyslexics and 48 controls were tested for phonological awareness, auditory discrimination, motion detection, visual attention, and rhythm imitation. A combined cluster and discriminant analysis approach revealed three clusters of dyslexics with different cognitive deficits. Compared to reading-unimpaired children cluster #1 performed worse in the phonological, auditory, and magnocellular tasks; cluster #2 had worse phonological awareness; and cluster #3 had higher attentional costs. These results indicate that dyslexia may result from distinct cognitive impairments. As a consequence, prevention and remediation programmes should be specifically targeted for the individual child's deficit pattern.

[1] Full citation and author details appear in Appendix.

1. Introduction

Developmental dyslexia is the disability to learn and perform reading sufficiently in spite of average or above-average intelligence and adequate education. Up to 17.5% of all children are affected (Shaywitz, 1998). The cognitive mechanisms underlying dyslexia are still a matter of debate. Numerous theoretical approaches have identified different potential causes of dyslexia.

The *phonological theory* (e.g. Liberman, 1973; Snowling, 2000), which is the most influential account for reading problems, relates dyslexia to a deficit in phonological awareness, i.e. the ability to segregate and manipulate the speech sounds that form a word (e.g. deleting the first sound from *pearl* gives *earl*).

In contrast, the *auditory processing deficit theory* (e.g. Tallal, 1980) assumes that dyslexics have a deficit in (rapid) auditory processing. It is argued that due to this more basic deficit, no adequate phonological representations can be built, resulting in additional phonological impairments. Thus, according to this theory, phonological problems are only secondary to the auditory deficits.

Yet other researchers conceptualise dyslexia as a visual processing deficit arising from the impairment of the visual *magnocellular system* in the brain (e.g. Stein & Walsh, 1997). This system supports the processing of rapidly moving visual stimuli and is thus important for vision during saccadic eye movements. Dysfunction of the magnocellular system is supposed to result in blurred visual representations of e.g. letters which, as a consequence, are more difficult to distinguish.

The role of *attentional deficits* for the development of dyslexia is also discussed (e.g. Facoetti et al., 2001, 2003; Hari & Renvall, 2001). Attentional deficits are thought to interfere with the encoding of a sequence of letters, resulting in the confusion of letters and visual word forms. Interestingly, attentional deficits can be dissociated from phonological deficits, and both types of deficits are valid predictors of reading (dis)ability (Valdois et al., 2003).

Finally, the *cerebellar theory* (e.g. Nicolson et al., 2001; Nicolson & Fawcett, 1999, 2005) argues that reading disabilities are a consequence of the impaired ability to *automatise* processes. It is assumed that the cerebellum supports the automatisation of basic articulatory and auditory abilities, which are relevant for the grapheme-phoneme correspondence. Moreover, the control of eye-movements during reading is controlled (among others) by the cerebellum.

The latter four theories (auditory, visual-magnocellular, attentional, and cerebellar/automatisation) have been subsumed in the *general magnocellular theory* of dyslexia (Stein, 2001), discussing impairment of the magnocellular system as the common cause for isolated or combined cognitive deficits resulting in reading disability. The general magnocellular theory regards phonological impairments as secondary to the other deficits. In contrast, the neurobiological model by Ramus (2004) assumes microlesions to the perisylvian cortex (mainly affecting phonological processing) as the primary cause of dyslexia. These lesions may, but need not, extend further into sub-cortical structures, additionally impairing other (e.g. magnocellular) cognitive functions.

All of these theories have seen supporting empirical evidence (for reviews see e.g. Ramus, 2003; Démonet, 2004). Interestingly, however, not all dyslexics suffer from deficits in all cognitive domains or profit equally from all remediation techniques (Ramus, 2003). Thus, it is possible that distinguishable phenotypes of dyslexia exist on the cognitive level (cf. Démonet et al., 2004; Ramus, 2004; Morris et al., 1998; Lachmann et al., 2005; Ho et al., 2004, 2007) for which universal or distinct genetic (Olson, 2002; Schulte-Körne et al., 2007) and neurobiological (Ramus, 2004; Paulesu et al., 2001) causes are controversially discussed. Unravelling different subtypes of dyslexia would be an essential prerequisite for developing or applying specifically targeted and thus more efficient remediation strategies (Rüsseler, 2006). These might even be administered to pre-school children (Démonet et al., 2004) before reading instruction begins.

The present study aimed at investigating subtypes of dyslexics that have specific and distinguishable deficits in one or more of the five cognitive domains, i.e., phonological awareness, auditory processing, visual-magnocellular processing, attention, and automatisation. Moreover, the relationship of phonological processing and the other cognitive abilities subsumed in the general magnocellular theory were investigated in order to obtain evidence for the evaluation of current neuro-cognitive models of dyslexia (e.g. Stein, 2001; Ramus, 2004).

2. Materials and methods

2.1. Participants

In order to acquire volunteers for the study, we sent a short project description to the headmasters and headmistresses of 40 primary schools in Aachen, Germany. Out of these, 21 agreed to cooperate. In a next step, detailed project descriptions containing information about the procedures of the study and an informed consent sheet created according to the Declaration of Helsinki (World Medical Association, 2000) were sent to the schools and forwarded by the teachers to the parents of their 3rd grade pupils. Of the parents, 642 agreed to let their children participate. Of these children, 104 were later selected for further examination according to criteria listed below, from which 97 complete data sets were obtained and analysed.

2.2. Procedure

Testing was performed in two successive phases. First, the reading abilities were assessed in the schools as group tests. Second, children with normal and with deficient reading scores were further tested individually during or after school in a classroom or at home for their non-verbal intelligence, phonological awareness, auditory sound discrimination, automatisation, magnocellular functions, and visual attention. The non-verbal intelligence test was always administered first since the value was used as an independent variable for the inclusion or exclusion of the subject into the group of dyslexics or controls. The other tests providing the dependent variables of the study were administered in a pseudo-randomised order. The order of the tests was balanced over subjects to exclude systematic influences of sequence. All tests were administered by one co-author (JT) in the last third of the term. The tests and paradigms employed for the assessment of the cognitive functions were standardised psychometric tests for which norms were available, or computerised tests involving well-established and previously published paradigms that validly tap the processes under investigation.

2.3. Psychometric tests and questionnaires

Reading test: "Würzburger Leise Leseprobe" (WLLP)
Reading ability was assessed with the Würzburger Leise Leseprobe (WLLP) (Küspert & Schneider, 1998), a standard German test for reading speed. Participants read words at the beginning of a row and mark with a pencil the one out of four pictures displayed in the same row that is denoted by the word. The names of the three distractor pictures may be semantically related, phonologically related, or semantically and phonologically related to the target word (e.g. *Blatt* 'leaf': *Blatt* (target); *Bett* 'bed' (phonologically related); *Ast* 'branch' (semantically related); *Baum* 'tree' (semantically and phonologically related). Separate norms (of 1997) are available for boys and girls in the first, second, third, and fourth grade. According to the research criteria of the International Classification of Diseases (WHO, 2006) children performing at or below the 10^{th} percentile were considered as dyslexic if the subsequent non-verbal intelligence test yielded an average or above-average IQ. Children performing above the 25^{th} percentile and with comparable IQ were included in the control group.

Non-verbal intelligence: "Grundintelligenztest Skala 2" (CFT-20)
The non-verbal IQ was assessed with the German version of the Cattell Culture Fair Test 20 (CFT 20; Weiß, 1998). In this test, participants are shown series of pictures generated according to a particular logic. In subtest 1, "Series", subjects have to complete a series of pictures (e.g. white squares containing increasingly long black bars) by marking the one out of five alternatives that is the correct continuation. Subtest 2, "Classifications", requires the participants to indicate which out of five pictures had not been generated after the same principle as the others (e.g. a white square with a vertical black bar among four white squares with horizontal black bars). In subtest 3, "Matrices", participants have to identify which out of five alternative pictures completes a set of three pictures (e.g. find a white square with two horizontally oriented black circles). Subtest 4, "Topologies", is supposed to assess logical reasoning. Subjects have to identify the one out of four alternatives that was created according to the same principle as a sample stimulus (e.g. where a dot can be inserted in a circle without placing it in a square). The performance in each subtest was assessed as the number of items that were correctly marked in a pre-defined time period (4 minutes for subtests 1 and 2; 3 minutes for subtests 3 and 4). The CFT-20 was administered individually in its short form (Part 1) for reasons of time economy. Norms for the short form are available for different age-groups (8;7–70 years), grades (5^{th}–10^{th}), or school types (primary school grade 3 and 4; vocational school years 1 and 2) as IQ scores, *T* values, and percentiles.

In the present study, age-related IQ scores were used. Children were only included in the study if they had at least average non-verbal intelligence (IQ $\geq$ 85) in order to ensure that no children with learning disabilities participated.

Phonological awareness

Phonological awareness, i.e. the ability to segregate and manipulate phonemes from given words, was tested with the German "Basiskompetenzen für Lese-Rechtschreibleistungen" (BAKO 1–4; Stock et al., 2003). From the seven subtests included in BAKO, one productive (Test 4, Phoneme Exchange) and one receptive test (Test 6, Vowel Length Discrimination) were selected, for which separate norms (of 2002; *T* values and percentiles) for grades 1–4 are available. In the productive test, subjects heard spoken stimulus words played from a CD and had to utter the pseudo-word that results from exchanging the first two phonemes (e.g. *Masse* → "amsse"). In the receptive test, they had to identify the one out of four auditorily presented pseudo-words in which the vowel was of a different length (e.g. *maar – raas – dack – laat*). Since the scores of both tests were highly correlated ($r = .41$, $p < .001$), the average *T* value was calculated for each child, which entered the further analyses as the measure for phonological awareness.

Auditory sound discrimination

Auditory sound discrimination was assessed with subtest 1 of the Heidelberger Lautdifferenzierungstest (H-LAD; Brunner et al., 1998). The children indicated if minimal sound pairs (e.g. /ba/ – /pa/) were identical or different. Stimulus materials included both non-lexicalised syllables (Test 1B) and words (Tests 1A and 1C). The stimuli were played from a CD at 60 dB. The H-LAD provides norms (*T* values and percentiles) for the second and fourth grade. Since the test was administered to third-graders in the last term of the school year, the *T* values for the fourth grade were calculated.

Questionnaire "FBB-HKS"

The FBB-HKS (Fremdbeurteilungsbogen für hyperkinetische Störungen) is a questionnaire for the assessment of attentional deficit and hyperactivity disorder. It is part of the "Diagnostik-System für Psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-IV" (DSYPS-KJ; Döpfner & Lehmkuhl, 2000). Separate norms for boys and girls are available (Brühl et al., 2000) for the three symptom categories "attentional deficit", "hyperactivity", and "impulsiveness". In the present study, only the ratings for the attentional deficit syndrome (ADS) were considered because these might be related to reading performance. Children classified as having comorbid ADS symptoms were not excluded from the sample. Instead, the number of children with additional ADS was included into the analysis in order to test whether ADS symptoms had a substantial influence on the performance.

2.4. Computerised tests

The presentation of the stimuli during the computerised tests and the registration of the button presses were performed with Presentation™ software (Version 0.70; Neurobehavioral Systems, Albany, CA, USA).

Automatisation

For the assessment of the automatisation abilities, a rhythm imitation paradigm was chosen (Tiffin-Richards et al., 2001). In this paradigm, the children had to click a button synchronously to an auditorily presented rhythm consisting of 5–6 beats with stimulus-onset asynchronies of 300 or 600 ms. For the present study we chose the complex rhythms #1, #4, and #5 from Tiffin-Richards et al. (2001) which had best differentiated between dyslexic and control children in that study. Each rhythm was first presented once in order to familiarise the child with it. Then, each rhythm was repeated five times while the subjects had to imitate it by synchronously clicking the left mouse button with the right index finger. The total number of correct rhythms was used as the indicator of automatisation ability. Only those rhythms were judged as correct for which (i) the number of mouse clicks equalled the number of beats and (ii) a required click was made before the next beat was played.

Magnocellular functions

Wilms et al. (2005) recently presented a visual paradigm that activated area V5/MT+ as part of the magnocellular system in a functional magnetic resonance imaging study. In this paradigm, the participants are presented with a radially expanding, static, or contracting random dot pattern which is well controlled for visual properties (for details, see Wilms et al., 2005). It has been demonstrated that such moving stimuli are processed differently by dyslexics and by normal controls: dyslexics have attenuated motion-onset related visual evoked potentials (Schulte-Körne et al., 2004). Therefore, in the present study, the paradigm of Wilms et al. (2005) was adopted. The random dot pattern changed its type of motion (e.g. expand → static, expand → contract, static → expand, etc.) after a variable time interval of 1 s, 1.5 s, 2 s, 2.5 s, or 3 s. The time intervals and motion direction changes were pseudo-randomised. The transition probabilities were equal for all types of motion changes. The participants had to indicate the changes in motion by clicking the left mouse button as quickly as possible. The average reaction times for correct responses were taken as measures for magnocellular functioning.

Visual attention

In the Posner paradigm (Posner, 1980) subjects had to indicate, by clicking the left or right mouse button as quickly as possible, at which of two positions in the left and right periphery of the computer screen a target stimulus occurred. A target stimulus could be preceded by a cue which was either informative ("valid"), neutral, or misleading ("invalid"). In the case of a valid cue, the participant could correctly prepare the reaction. In the neutral condition, he or she was alerted that a stimulus would be presented soon but could not prepare a left or right button click. In the invalid condition, attention to the shadowed position had to be redirected to the correct position of the target in order to perform the required reaction. The "cue validity effect", which is the reaction time difference between invalid and valid trials, is taken to reflect how quickly attention can be shifted to a new location. Smaller effects indicate better performance.

The Posner paradigm has been successfully applied for the identification of attentional deficits in dyslexic children (Heiervang & Hughdahl, 2003). In the present study, we used a version of the Posner paradigm that was previously applied by Vossel et al. (2006) with central cues and a ratio of 80:20 for valid versus invalid cues. Only correct trials were analysed. The average cue validity effect was calculated for each participant and included in the subsequent analyses.

2.5. Data analysis

In all analyses, only subjects that had complete data sets, i.e. values for all five tested cognitive functions, were considered ($N = 97$). All analyses were conducted with SPSS 12.0.1 for Windows (SPSS Inc., Chicago, Illinois, USA).

Multiple regression for reading

In order to understand the overall impact of the tested cognitive functions on reading ability, a multiple regression analysis was performed for the entire sample.

Two-step cluster analysis

The existence of sub-types within the dyslexic sample was tested with a two-step cluster analysis which provides the optimum number of clusters in a given data set. The analysis was run allowing for a maximum of 15 clusters, log-likelihood distance estimation, Akaike's information criterion as clustering criterion, no noise-handling for outlier treatment, initial distance change threshold of 0, a maximum of eight branches per leaf node, and a maximum of three tree depth levels. All variables were standardised during the clustering procedure. A Bonferroni-correction was applied.

Discriminant analyses

First, a discriminant analysis was conducted in order to assess in respect to which of the five cognitive variables the dyslexics as one homogeneous group differed from the controls. The variables were entered step-wise with an inclusion criterion of $p < .05$ and an exclusion criterion of $p \geq .10$. All priors were set equal. Wilks' lambda (λ) was calculated for each step. For the analysis of the correct assignment of participants to a particular group, the more conservative cross-validated statistics were reported. For the subsequent pair-wise discriminant analyses conducted in order to compare the clusters among themselves and against the control group, the same settings were selected.

3. Results and discussion

3.1. Results

Description of the sample

Data from 97 children were included in the analysis. From these, four children were excluded because no age norms for the H-LAD were available. Among the remaining 93 children (48 girls, 45 boys), there were 45 dyslexics (24 girls, 21 boys; mean age: 9.3 years) and 48 controls (24 girls, 24 boys; mean age 9.2 years). The average non-verbal IQ in the dyslexic group was 108.4, that in the controls 113.9. The reading score in the dyslexics was below the fifth percentile, whereas that of the controls was below the 63rd percentile.

Reading score and cognitive functions

The multiple regression for the entire sample revealed that phonological awareness and attention were the two significant predictors for the reading score ($R^2 = .36$; phonological awareness: $\beta = .42$, $p < .001$; attention: $\beta = -0.39$, $p < .001$).

Dyslexics versus controls

The two groups differed in phonological awareness ($\lambda = .713$; $F(1, 91) = 36.59$, $p < .001$) and attention ($\lambda = .578$; $F(2, 90) = 32.81$, $p < .001$). 83.7% of the dyslexics (41/49) and 79.2% of the controls (38/48) were correctly assigned to their groups by their scores on these two functions.

Clusters of dyslexics

The two-step cluster analysis of the data of the dyslexic sample for all five cognitive functions revealed three distinguishable clusters (Figure 1). Four subjects for which no age-related *T* values for auditory sound discrimination were available were excluded. The subsequent discriminant analysis over the identified clusters and the controls supported the clustering by the correct assignment of 15/16 (93.8%), 11/15 (73.3%), and 13/14 (92.9%) dyslexics to clusters #1–#3, respectively, yielding an overall correct assignment of 86.7%. The three subgroups of dyslexics and the control group differed significantly (all $p < .001$) from one another with respect to sound discrimination ($\lambda = .600$; $F(3, 89) = 19.77$), attention ($\lambda = .373$; $F(6, 176) = 18.70$), phonological awareness ($\lambda = .270$; $F(9, 211.89) = 16.80$), and magnocellular functions ($\lambda = .232$; $F(12, 227.83) = 14.01$). Figure 1 displays the average scores for reading ability and the five cognitive functions separately for each cluster.

Characterisation of the dyslexia clusters

The absolute cognitive profiles characterising each cluster were identified in pair-wise discriminant analyses of each cluster versus the control group (all $p < .001$). Cluster #1 performed worse than the control group in the tasks tapping phonological awareness

(λ = .551; $F(1, 62)$ = 50.58), sound discrimination (λ = .459; $F(2, 61)$ = 35.97), and magnocellular functions (λ = .423; $F(3, 60)$ = 27.23). Children in cluster #2 had lower scores than the controls only in phonological awareness (λ = .791; $F(1, 61)$ = 16.16), whereas sound discrimination was even better than in the controls (λ = .679; $F(2, 60)$ = 14.19). Cluster #3 was characterised by a high cue validity effect indicating slow attentional reorienting (λ = .587; $F(1, 60)$ = 42.25).

The relative differences between the dyslexics in one cluster compared to each other cluster were also assessed with pair-wise discriminant analyses (all $p < .001$). Cluster #1 performed worse in sound discrimination and magnocellular functions than both other clusters (cluster #2: sound discrimination: λ = .231; $F(1, 29)$ = 96.41; magnocellular: λ = .197; $F(2, 28)$ = 57.02; cluster #3: sound discrimination: λ = .190; $F(2, 27)$ = 57.57; magnocellular: λ = .428; $F(1, 28)$ = 37.48). Participants in cluster #2 had lower phonological awareness than cluster #3 (λ = .206; $F(3, 25)$ = 32.14) but performed better than cluster #3 in the attention and sound discrimination tasks (attention: λ = .325; $F(1, 27)$ = 56.20; sound discrimination: λ = .264; $F(2, 26)$ = 36.28).

Analysis of other factors
In addition to the previous analyses, potential gender differences between the dyslexic clusters were analysed. A one-factorial ANOVA yielded no effect of gender ($F(3, 89) < 1$).

Finally, the number of children with potential comorbid ADS (according to the FBB-HKS) were identified for each subgroup of children. There were two children in the control group, one child in cluster #1, three children in cluster #2, and two children in cluster #3 who were classified as suffering from ADS. Over the entire sample, the FBB-HKS classification was not significantly correlated with the size of the cue-validity effect ($r = -0.178$, $p = .120$).

3.2. Discussion

We identified three sub-groups of dyslexics with distinct cognitive patterns (Figure 1). Cluster #1 performed worse than the controls in the phonological, auditory, and magnocellular tasks. In contrast, cluster #2 only had impaired phonological awareness, whereas cluster #3 had increased attentional shift costs relative to the controls.

Distinct cognitive causes for dyslexia
The results reveal that differentiating among dyslexia subtypes with specific impairments allows a more fine-grained understanding of the disorder than simply comparing “dyslexics” against controls. This is evident from the fact that the three dyslexic clusters in total differ from the controls in more cognitive functions than the dyslexic group as a whole. Moreover, the correct assignment of the dyslexics improved by differentiating between clusters of dyslexics with distinct cognitive function patterns.

These findings may explain the heterogeneity of results in previous studies (for a review, see Ramus, 2003) which reported deficits in some but not all of the cognitive functions for some but not all dyslexics. The data from the present cluster analysis ap-

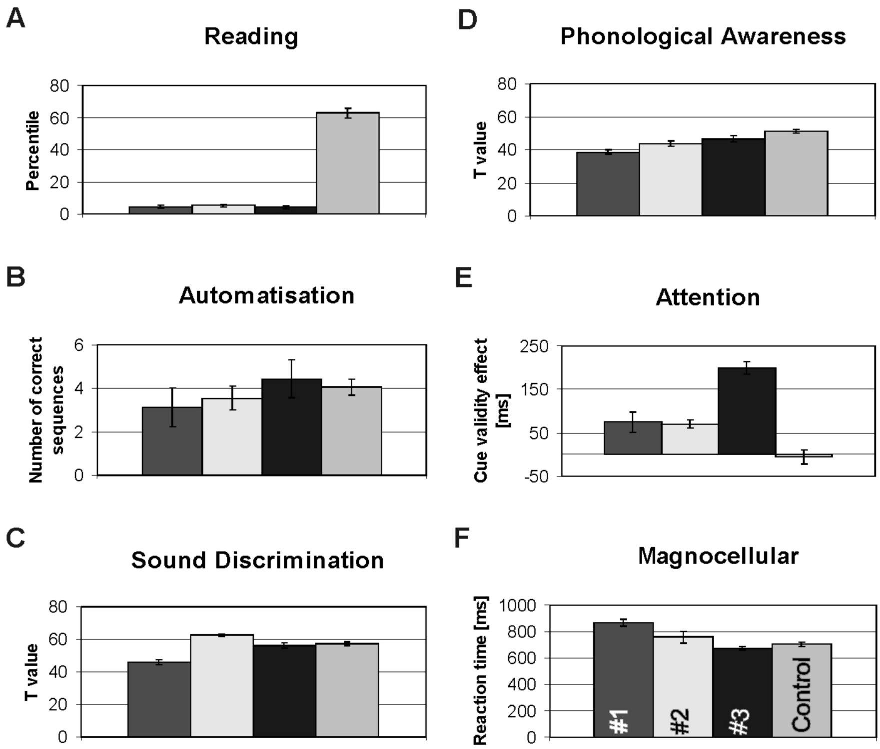


Figure 1. Description of the dyslexia clusters (columns 1–3) and the controls (column 4). A. Average percentile (± SEM) of the reading score. B. Total number of correctly imitated rhythms (± SEM). C. Average *T* value (± SEM) for sound discrimination in the H-LAD. D. Average *T* value (± SEM) for phonological awareness in the BAKO. E. Average size of the cue validity effect (± SEM) in the Posner paradigm. F. Average reaction time (± SEM) for the detection of motion changes indicative of magnocellular functions.

proximately reflect the proportions of dyslexics with corresponding deficits reported in the literature (calculated by Ramus, 2003; see also Ramus et al., 2003). A magnocellular decrease was observed in 35.6% of the dyslexics (Ramus, 2003: 29%). 35.6% of the clustered cases had an auditory deficit that was accompanied by a phonological deficit (Ramus, 2003: 39%). In the present study, 33.3% of the dyslexics had an isolated phonological effect (Ramus et al., 2003: 31.3%).

The relationship of phonological awareness and magnocellular functions

These considerations may be important for theoretical accounts for dyslexia. Whereas the neurobiological model presented by Ramus (2004) considers phonological deficits as primary and magnocellular impairments as secondary causes for dyslexia, the reversed order is assumed in the general magnocellular theory (Stein, 2001). In our study, bad phonological performance occurred in combination with decreased magnocellular and auditory performance (cluster #1), which is in line with both theories. However, cluster #2 only showed a phonological deficit, whereas the auditory abilities were even better than in the normal readers. Moreover, no cluster had isolated magnocellular deficits without phonological impairment. This pattern of results would be predicted by the Ramus (2004) model but is not compatible with the general magnocellular account. The data also go beyond the findings by Morris et al. (1998) who observed a phonological core deficit for all dyslexic children they examined.

No automatisation deficit?

With respect to automatisation abilities, the present data are not unequivocal. The dyslexics in cluster #1 showed a trend towards an automatisation deficit, but this result was not significant. No other dyslexia cluster revealed an automatisation deficit either. This is in contrast to some earlier studies (Tiffin-Richards et al., 2001; Nicolson et al., 1996) that did find impaired automatisation abilities in dyslexics. In particular, we did not replicate the results of Tiffin-Richards et al. (2001) from whose study the rhythms were adapted.

However, the findings are in line with the results of Ramus et al. (2003) who found no difference between dyslexics and controls in five different cerebellar/automatisation tasks (balance, bimanual finger tapping, repetitive finger tapping, bead threading, finger-to-thumb movement) in a sample of equal size to ours. Clearly, this discrepancy between studies that do or do not obtain automatisation deficits in dyslexics requires further investigation.

With respect to the relationship of automatisation and other cognitive causes of dyslexia, it is important to note that the dyslexics in cluster #1, who had the worst automatisation scores, were also impaired in three other cognitive functions. This finding might suggest that the automatisation deficit is associated with a larger, multiple cognitive deficit as proposed by the general magnocellular theory (Stein, 2001).

Attention and reading

In the present study, children with a positive ADS classification were not excluded. This comorbidity may present a potential confound in the data. However, the respective analyses carried out in the present study do not support this view. First, the number of children with a positive ADS score was comparable over all clusters. Even more importantly, there were only two ADS children in cluster #3 which was characterised by increased attentional shift costs. This result was further corroborated by the correlation analysis that revealed no significant relationship between the ADS classification and the performance in the Posner task. Thus, in the present study, we found no evidence that the data were significantly influenced by a comorbid ADS. Nonetheless, it is possible and potentially interesting to systematically investigate the relationship of dyslexia and ADS in future studies that may also include other tasks tapping different aspects of attention, e.g. vigilance or divided attention.

Cognitive versus language-related impairments in dyslexia

The present study was not the first to observe that dyslexics may be classified into subgroups. Earlier work demonstrated that dyslexics differ with respect to their performance in language-related tasks. Morris et al. (1998) identified seven clusters of dyslexics that were either globally deficient, impaired in phonological awareness in combination with rapid naming, or showed a deficit in processing rate. Similar results for Chinese were reported by Ho et al. (2004) whose subjects were deficient either globally, in an orthographic task, in phonological memory, or rapid naming. King et al. (2007) investigated the relationship of phonological awareness and rapid naming in dyslexia. The authors found four groups that had deficits in either phonological awareness or rapid naming, both tasks, or none. Lachmann et al. (2005) combined behavioural measures of word and non-word reading with a mismatch negativity (MMN) paradigm. They observed two groups of dyslexics, one impaired in word-reading and the other in non-word reading, with a reduced MMN amplitude in the first relative to the second group.

Taken together, these studies provide good insight into the language-related functions that may be affected jointly or distinctly in dyslexia. In contrast to these findings which only focus on aspects of language processing, a study by Valdois et al. (2003) investigated the relationship of attention and phonological awareness during reading, thus considering not only linguistic but also other cognitive influences (i.e. attention). The present data support this finding since in the multiple regression, both phonological awareness and attention significantly predict reading performance. Moreover, it was in these two functions that the dyslexic group as a whole differed from the controls. However, the present study goes beyond the approach by Valdois et al. (2003) by including a number of cognitive functions such as attention, automatisation, and magnocellular functioning which are all discussed as being substantial for reading. The data show the merit of this multivariate approach since many of the investigated cognitive functions prove differentially relevant for reading difficulties in dyslexia. The choice of cognitive functions included in the present study was guided by the leading theories of dyslexia but is certainly not exhaustive. Future research may include further aspects such as working memory or rapid naming in order to describe the cognitive basis of dyslexia in an even more fine-grained manner.

4. Conclusion

The present study demonstrates that dyslexic children can be classified into different groups with distinct cognitive patterns. There were three clusters that performed worse than the controls, either only in the phonological or only in the attention tasks, or differed from the controls in phonological, auditory, and magnocellular scores. As demonstrated above, these findings may account for the heterogeneity of results in the literature concerning cognitive deficits in dyslexics.

More generally, the data reveal that a cluster-analytic approach to multiple cognitive deficits in dyslexia is advantageous. The cognitive functions tested in the present study were selected in accordance with recent cognitive theories of dyslexia and may only be a fraction of all relevant functions. Future studies may use the cluster-analytic approach to include additional cognitive functions such as working memory or lexical retrieval during picture naming. Finally, the results may inspire the identification of dyslexia subtypes on the neurobiological and genetic level.

As a consequence for diagnostics and intervention, a refined view on specific deficits of dyslexic children may motivate the more targeted use of already existing tests and remediation strategies. In particular, pre-school diagnostics before reading instruction may include cognitive functions like the ones investigated in the present study, since none of these functions require the processing of writing. The combination of early diagnostics and deficit-specific intervention may present a significant advance for ameliorating reading problems in children

.

5. Acknowledgements

This paper was originally published in *Acta Neurobiologiae Experimentalis* (see Appendix for full citation and author details). We wish to thank the Editor-in-Chief, Professor Krzysztof Turlejski, for the kind permission to reprint the article.

We also wish to thank John Stein, Angela Fawcett, Maria Luisa Lorusso, Andrea Facoetti, Richard Olson, Margaret Snowling, Michel Habib, Piotr Jaskowski, Cecilia Marino, Marco Battaglia, Aryan Van der Leij, Peter de Jong, Heikki Lyytinen, Michel Maziade, and Marcin Szumowski for discussing some of the basic ideas of this research with us. Moreover, we appreciate the support with the data acquisition by Helen Schreiber.

This research was supported by the German Federal Ministry of Education and Research (BMBF 01GJ0613 to SH and BMBF 01GJ0614 to MG), the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and the National Institute of Mental Health (KA). Further support by Helmholtz-Gemeinschaft and the Brain Imaging Center West (BMBF 01GO0204) is gratefully acknowledged. We thank all primary schools and in particular all parents and children who participated in this study.

References

Brunner, M., Seibert, A., Dierks, A. & Körkel, B. (1998). *Heidelberger Lautdifferenzierungstest (H-LAD)*. Göttingen: Hogrefe.

Démonet, J. F., Taylor, M. J. & Chaix, Y. (2004). Developmental dyslexia. *Lancet*, *363*, 1451–1460.

Döpfner, M. & Lehmkuhl, G. (1998). *DISYPS-KJ. Diagnostik-System für psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-IV* (Manual). Bern: Huber.

Facoetti, A., Lorusso, M. L., Paganoni, P., Cattaneo, C., Galli, R., Umiltà, C. & Mascetti, G. G. (2003). *Brain Research. Cognitive Brain Research 16*, 185–191.

Facoetti, A., Turatto, M., Lorusso, M. L. & Mascetti, G. G. (2001). Orienting of visual attention in dyslexia: Evidence for asymmetric hemispheric control of attention. *Experimental Brain Research*, *138*, 46–53.

Fawcett, A. J., Nicolson, R. I. & Dean, P. (1996). Impaired performance in children with dyslexia on a range of cerebellar tasks. *Annals of Dyslexia*, *46*, 259–283.

Hari, R. & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends in Cognitive Science*, *5*, 525–532.

Heiervang, E. & Hugdahl, K. (2003). Impaired visual attention in children with dyslexia. *Journal of Learning Disabilities*, *36*, 68–73.

Heim, S., Tschierse, J., Amunts, K., Vossel, S., Wilms, M., Willmes, K., Grabowska, A. & Huber, W. (2008). Cognitive subtypes of dyslexia. *Acta Neurobiologiae Experimentalis*, *68*, 73–82.

Ho, C. S., Chan, D. W., Chung, K. K., Lee, S. H. & Tsang, S. M. (2007). In search of subtypes of Chinese developmental dyslexia. *Journal of Experimental Child Psychology*, *97*, 61–83.

Ho, C. S., Chan, D. W., Lee, S. H., Tsang, S. M. & Luan, V. H. (2004). Cognitive profiling and preliminary subtyping in Chinese developmental dyslexia. *Cognition*, *91*, 43–75.

King, W. M., Giess, S. A. & Lombardina, L. J. (2007). Subtyping of children with developmental dyslexia via bootstrap aggregated clustering and the gap statistic: Comparison with the double-deficit hypothesis. *International Journal of Language and Communication Disorders*, *42*, 77–95.

Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe*. Göttingen: Hogrefe.

Lachmann, T., Berti, S., Kujala, T. & Schröger, E. (2005). Diagnostic subgroups of developmental dyslexia have different deficits in neural processing of tones and phonemes. *International Journal of Psychophysiology*, *56*, 105–120.

Liberman, I. Y. (1973). Segmentation of the spoken word. *Bulletin of the Orton Society*, *23*, 65–77.

Lovegrove, W. J., Bowling, A., Badcock, B. & Blackwood, M. (1980). Specific reading disability: Differences in contrast sensitivity as a function of spatial frequency. *Science*, *280*, 439–440.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., Katz, L., Francis, D. J. & Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology*, *90*, 347–373.

Nicolson, I. R., Fawcett, A. J. & Dean, P. (2001). Dyslexia, development and the cerebellum. *Trends in Neurosciences*, *24*, 515–516.

Olson, R. K. (2002). Dyslexia: Nature and nurture. *Dyslexia, 8*, 143–159.

Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S. F., Cossu, G., Habib, M., Frith, C. F. & Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science*, *291*, 2165–2167.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, *13*, 212–218.
Ramus, F. (2004). Neurobiology of dyslexia: A reinterpretation of the data. *Trends in Neurosciences, 27*, 720–726.
Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S. & Frith, U. (2003). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, *126*, 841–865.
Rüsseler, J. (2006). Neurobiologische Grundlagen der Lese-Rechtschreib-Schwäche. *Zeitschrift für Neuropsychologie*, *17*, 101–111.
Schulte-Körne, G., Bartling, J., Deimel, W. & Remschmidt, H. (2004). Motion-onset VEPs in dyslexia. Evidence for visual perceptual deficit. *Neuroreport*, *15*, 1075–1078.
Schulte-Körne, G., Ziegler, A., Deimel, W., Schumacher, J., Plume, E., Bachmann, C., Kleensang, A., Propping, P., Nöthen, M. M., Warnke, A., Remschmidt, H. & König, I. R. (2007). Interrelationship and familiality of dyslexia related quantitative measures. *Annals of Human Genetics*, *71*, 160–175.
Shaywitz, S. E. (1998). Dyslexia. *New England Journal of Medicine*, *338*, 307–312.
Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Oxford: Blackwell.
Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia*, *7*, 12–36.
Stein, J. & Walsh, V. (1997). To see but not to read: The magnocellular theory of dyslexia. *Trends in Neurosciences*, *20*, 147–152.
Stock, C., Marx, P. & Schneider, W. (2003). *BAKO 1–4. Basiskompetenzen für Lese-Rechtschreibleistungen*. Göttingen: Beltz.
Tallal, P. (1980). Auditory temporal perception, phonics, and reading disability in children. *Brain and Language*, *9*, 182–198.
Tiffin-Richards, M. C., Hasselhorn, M., Richards, M. L., Banaschweski, T. & Rothenberger, A. (2001). Time reproduction in finger tapping tasks by children with attention-deficit hyperactivity disorder and/or dyslexia. *Dyslexia*, 10, 299–315.
Valdois, S., Bosse, M. L., Ans, B., Carbonnel, S., Zorman, M., David, D. & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing*, *16*, 541–572.
Vossel, S., Thiel, C. M. & Fink, G. R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. *Neuroimage*, *32*, 1257–1264.
Weiß, R. H. (1998). *Grundintelligenz Skala 2. CFT 20*. Göttingen: Hogrefe.
Wilms, M., Eickhoff, S. B., Specht, K., Amunts, K., Shah, N. J., Malikovic, A. & Fink, G. R. (2005). Human V5/MT+: Comparison of functional and cytoarchitectonic data. *Anatomy and Embryology*, *210*, 85–495.
World Health Organization. (2006). *International Classification of Diseases ICD-10* (10th revision), http://www.who.int/classifications/apps/icd/icd10online/.
World Medical Association. (2000). *Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects*. Edinburgh.

Appendix

This paper was originally published as Heim, S., Tschierse, J., Amunts, K., Vossel, S., Wilms, M., Willmes, K., Grabowska, A. & Huber, W. (2008). Cognitive subtypes of dyslexia. *Acta Neurobiologiae Experimentalis, 68,* 73–82. Please cite the original work along with the present paper.

Apart from minor typographical detail, the present paper differs from the 2008 version primarily in that the original's Section 3 (Results) and Section 4 (Discussion) have been merged.

For this paper, the full author list including affiliations is as follows:

Stefan Heim[1,2,4], Julia Tschierse[3], Katrin Amunts[1,2,4], Marcus Wilms[4] , Simone Vossel[4], Klaus Willmes[5], Anna Grabowska[6], Elisabeth Bay[3,4], Marion Grande[3], Walter Huber[3]

[1]Dept. of Psychiatry and Psychotherapy, RTWH Aachen University, Germany
[2]Brain Imaging Center West, Jülich, Germany
[3]Dept. of Neurolinguistics, RWTH Aachen University, Germany
[4]Institute for Neurosciences and Biophysics, Research Centre Jülich, Germany
[5]Dept. of Neuropsychology, RWTH Aachen University, Germany
[6]Dept. of Psychophysiology, Nencki Institute of Experimental Biology, Warsaw, Poland

Speech perception and brain function: Experimental and clinical studies

Kenneth Hugdahl
Department of Biological and Medical Psychology, University of Bergen
Division of Psychiatry, Haukeland University Hospital, Bergen

Abstract

In this paper, I present data on speech perception and brain hemispheric function based on a dichotic listening paradigm with series of presentations of consonant-vowel syllables. The typical finding is a right ear advantage, i.e. more correct reports for the right compared with the left ear stimulus, indicating a left hemisphere phonology processing dominance. The right ear advantage has been validated through functional neuroimaging studies, both PET and fMRI which have shown increased activation in the left compared to the right posterior regions of the temporal lobe. Another ERP study has revealed that the neurons in the left temporal lobe react about 20 ms before the corresponding neurons in the right temporal lobe. The right ear advantage is, however, modified through systematic variation of voiced versus unvoiced syllable pairs, revealing that phonetic stimulus features affect hemisphere asymmetry. The voiced versus unvoiced stimulus feature has been used to study how children at risk for dyslexia fail to process dichotically presented consonant-vowel syllables.

1. Introduction

Despite all research devoted to the study of functional asymmetry, we still lack an understanding of the anatomical, or structural correlates of differences in function between the two hemispheres of the brain. A clue to such an understanding may be the uniqueness of human speech sound processing. The unique ability of the human brain to extract the phonetic code from an acoustic signal is perhaps the defining characteristics of being a human. Thus, language and speech processing may be the key to an understanding of hemispheric asymmetry. The upper posterior part of the peri-Sylvian region is a rare exception to the general principle of structural symmetry in the brain (Figure 1).

The planum temporale region is about 30–35% larger on the left compared to the right side (Heiervang et al., 2000; Steinmetz et al., 1989). An increase in grey matter volume in one hemisphere may indicate a functional benefit with regard to speed of information processing. This view is supported by observations that the neuronal cortical columns are more widely spaced on the left side. This would indicate greater connectivity per neuron, which in turn would increase processing speed. The axons in the left planum temporale area are also more heavily myelinated, a fact which would indicate greater impulse travelling speed on the left compared to the right side. Thus, the structural asymmetry seen for the planum temporale area clearly has functional significance. Considering that the planum temporale area functionally overlaps with Wernicke's area, it is a short step to infer that the structural asymmetry evolved as an anatomical support for rapid processing of speech sounds, and consequently for extracting the phonological code of the acoustic speech signal.

2. The dichotic listening paradigm

2.1. Dichotic listening for probing temporal lobe asymmetries

Our research group at the University of Bergen has used an auditory task, introduced into neuropsychology by Doreen Kimura in 1961 (see also Bryden, 1963, Hugdahl & Andersson, 1984), with dichotic presentations of digits or consonant-vowel (CV) syllables. Dichotic presentations mean that two different syllables are presented simultaneously, one in each ear. The subject remains, however, unaware of the fact that two different syllables are presented on each trial, and is simply instructed to respond with the syllable s/he perceives on each trial, emphasizing a single syllable answer. This may also have a methodological advantage since requiring the subject to report both syllables would introduce a working memory component into the speech perception situation. The reason for this is that one syllable has to remain maintained in working memory as the other syllable is reported. This may have the consequence that what is reported is not only the result of a lateralized speech processing module but also of a working memory module. It may be of interest to note in this context that the original dichotic listening (DL) studies by Kimura (1961a, 1961b) were actually working

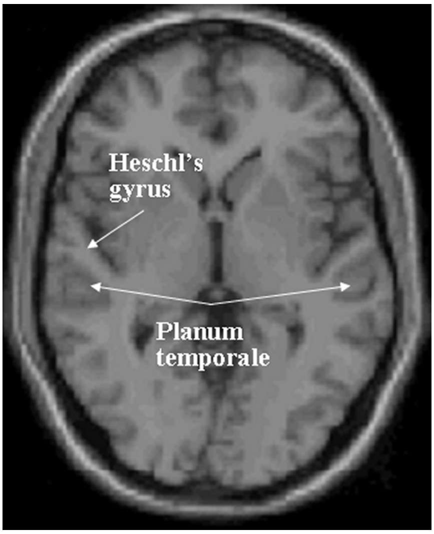


Figure 1. MR axial image through the upper part of the temporal lobes. Note the triangularly shaped grey Planum Temporale area on the left side.

memory studies rather than perceptual studies. Kimura used pairs of digits that were presented in blocks of six digits in a series, and the subject was then presented with four new digits and had to indicate if these digits were heard in the preceding series. This would require an intact working memory module so that the four first presented digits could be recalled and compared to the later presented digits. The CV syllables dichotic listening situation with requirements for single syllable answers constituted a methodological advance in this respect (e.g. Hugdahl & Andersson, 1984).

2.2. The stimulus material: Dichotic consonant-vowel syllable pairs

The stimuli used in our laboratory are typically paired presentations of the six stop-consonants /b/, /d/, /g/, /p/, /t/, /k/ together with the vowel /a/ to form dichotic CV syllable pairs of the type /ba/ – /ga/, /ta/ – /ka/ etc. The syllables are paired with each other for all possible combinations, thus yielding 36 dichotic pairs, including the homonymic pairs. The homonymic pairs are not included in the statistical analysis. The maximum score is 30 for each ear. Each CV-pair is recorded three times, with three different randomizations of the 36 basic pairs. Thus, the total number of trials is 108. The 108 trials are divided into three 36 trial-blocks, one trial-block for each instructional condition: non-forced attention (NF), forced right attention (FR), and forced left attention (FL). Each subject is given a standardized set of instructions prior to the test (see below). No significant differences for either the right or left ear scores have emerged when comparing different language sub-samples, most notably when comparing Norwegian and Swedish samples with Finnish samples, but also when comparing with German- and English-speaking samples. Figure 2 shows the principle behind the CV-syllable dichotic listening paradigm.

Figure 2. Schematic outline of the principles for dichotic listening, due to the preponderance of the contralateral auditory neural pathways favouring the processing of the right ear stimulus by the left hemisphere.

The syllables are typically read by a male voice with constant intonation and intensity. Mean duration is 350–400 ms (allowing for differences in voice-onset time length for unvoiced versus voiced CVs), and the inter-trial interval is on the average 4 s. The syllables are read through a microphone and digitized for later computer editing on a standard PC using state-of-the-art audio editing software (SWELL, Goldwave, CoolEdit, or comparable packages). In the original version of the task (Bergen, Norway), the syllables were recorded with a sampling rate of 44000 Hz and amplitude resolution of 16 bit. After digitization, each CV-pair was then displayed on the PC screen and synchronized for simultaneous onset at the first identifiable energy release in the consonant segment between the right and left channels. The stimuli are finally played to the subject using digital play-back equipment, connected to high-performance headphones, with an intensity between 70 and 75 dB. The exact editing procedure and equipment may, however, vary between laboratories due to ever more refined editing and presentation software packages.

3. The right ear advantage

The right ear advantage (REA) is a relatively strong empirical finding displayed in Figure 3 as a scatter plot of 651 adult subjects. The REA means that on the average, subjects report more correct syllables presented in the right compared to the left ear, also when controlling for hearing differences between the ears, or intensity differences for the two stimuli (Hugdahl et al., 2001; Tallus et al., 2007; Gadea et al., 2005). The REA is sensitive for age effects (Hugdahl et al., 2001), stimulus effects (Rimol et al., 2006),

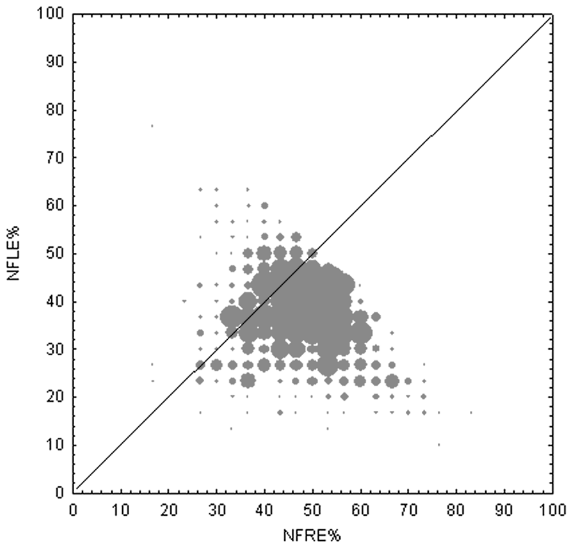


Figure 3. Scatter plot of % correct right (NFRE) and left (NFLE) ear scores for 651 adult subjects. The diagonal line represents the 45 degree symmetry line, with all dots "below" the line indicating subjects with a right ear advantage.

and handedness (Hugdahl & Andersson, 1984). It does typically not produce large sex effects, thought to modulate laterality differences, although when seen, it favours males (cf. Sommer et al., 2008). The REA is moreover reduced in certain clinical groups like schizophrenia (Løberg et al., 1999), dyslexia (Helland et al., 2008), brain lesioned patients (Benke et al., 1999), and patients with Parkinson's disease (Hugdahl & Wester, 2000). The REA is thought to be the result of two interrelated factors, left hemisphere dominance for processing of phonology and the preponderance of the contralateral auditory pathways (Kimura, 1967). Thus, the REA is a good, non-invasive, marker for left hemisphere speech processing dominance.

This brings up the question: what about the right hemisphere? From the anatomical model for dichotic listening performance and the REA (Kimura, 1967), two contrasting hypotheses could be advanced. On the one hand, it could be hypothesized that the right hemisphere does phonological processing like the left hemisphere, but to a lesser degree, that is, under normal circumstances the left hemisphere takes over the relevant phonological processing of the speech sounds (the so-called "direct access" hypothesis).

On the other hand, it could also be hypothesized that the right hemisphere does not perform any phonological processing at all, but that speech sounds have to be transferred across the corpus callosum to the left hemisphere in order to get processed ("callosal relay" hypothesis). A study by Pollmann et al. (2002) tried to resolve the controversy by comparing dichotic listening performance and the ear advantage in patients with vascular lesions in the caudal versus rostral sections of the corpus callosum. The results showed that patients with a lesion in the rostral 1/3 of the corpus callosum had a near perfect REA with no reports from the left ear stimulus. This proves that the right

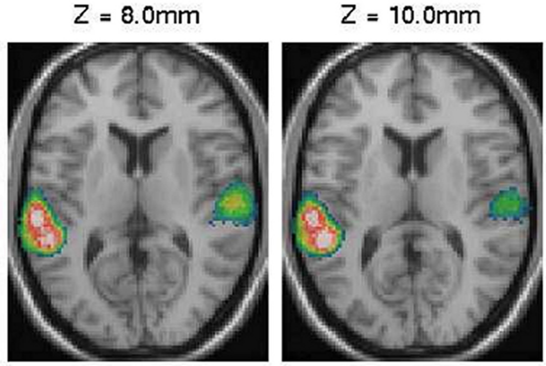


Figure 4. Brain activations in the left and right planum temporale area when subjects listen to consonant-vowel syllables. Note the marked left-sided asymmetry. The two panels show activation 8 and 10 mm above the AC–PC midline, respectively. From Hugdahl et al., 1999.

hemisphere under normal circumstances is not processing phonological characteristics of the stimuli, and that the speech stimuli are transferred across the callosum for processing (for a review, see Westerhausen & Hugdahl, 2008).

3.1. Neuronal correlates of the REA

The neuronal correlates of the behavioural REA were first investigated in a H_20 PET-study by Hugdahl et al. (1999), who used a target detection paradigm where the subjects had to press a button whenever they detected a pre-defined CV-syllable target stimulus (/ba/, /da/, or /ka/). The results are seen in Figure 4 and revealed a clear lateralized activation pattern covering the upper posterior part of the left peri-Sylvian region, also including the planum temporale area, in the NF instruction condition. The asymmetry was most marked in the range 8–20 mm above the AC–PC midline in the Talairach space. Interestingly, the asymmetry gradually became more and more left-sided the more posterior the activation. The PET data reported by Hugdahl et al. (1999) have been replicated also by fMRI by van den Noort et al. (2008), who used a verbal reports paradigm, like that which was used in the PET study while the subjects listened to dichotic presentations of consonant-vowel syllables. The novel aspect of the study of van den Noort et al. was that for the first time it could be shown that subjects could use a verbal report strategy when in an fMRI experiment without causing irresolvable head movement artefacts in the images. This was an important step forward in revealing the underlying neuronal structure of the REA since previous fMRI studies had used different recognition and target detection response strategies to avoid causing movement initiated artefacts. These strategies and procedures produce, however, another "artefact", namely the fact that in order to know which stimulus syllable to react to in a target detection paradigm, the subject must also keep the information in working memory throughout the experiment, which could interfere with activation to the syllables. The hemodynamic studies of the ear advantage in dichotic listening were recently validated with transcranial magnetic stimulation (TMS) in our laboratory where subjects first underwent the standard dichotic listening procedure without

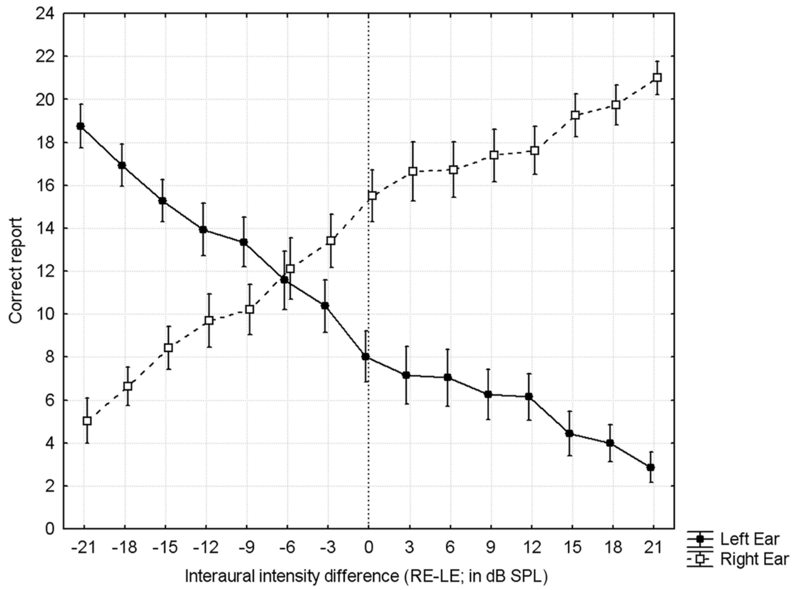


Figure 5. Correct reports for the right and left ear syllable during dichotic presentation when either the right or left ear channel is played more intensely. Each step on the x-axis represents a change of 3 dB favouring either the right (positive values) or left ear stimulus (negative values). From Hugdahl et al. (2008).

any TMS, then had 600 TMS pulses emitted over the left temporal lobe, and finally underwent the dichotic listening test once again immediately after the TMS stimulation. Preliminary results have revealed that the REA was reduced in 6 out of 7 subjects, in some cases rather dramatically so.

3.2. Quantifying the REA

We have recently been occupied with the question as to whether the strength of the REA can be quantified, i.e. whether it is possible to determine the psychophysical properties of the REA. We have approached this by manipulating the relative intensity of the right versus left ear syllable (see Hugdahl et al., 2008), while asking the question as to how much of an intensity difference in favour of the left ear stimulus can the REA withstand before it yields to a significant left ear advantage (LEA)? To achieve a gradual increase in intensity, we manipulated intensity in steps of 3 dB at a time, going from an increase in 3 dB to 21 dB, for either the right or left ear stimulus. The results are shown in Figure 5, and revealed that the REA can withstand an intensity difference between 6 and 9 dB before yielding to a significant LEA.

Figure 5 also shows that the change in the ear advantage as a consequence of manipulating the relative intensity for the right and left ear stimulus follows a linear relationship, with a gradual increase in the magnitude of the REA as the right ear stimulus intensity increases, and a similar linear increase in the magnitude of the LEA, after the crossing point at –6 dB, as the intensity of the left ear stimulus increases. This shows that the REA can be quantified and that it is a rather robust cognitive phenomenon. In a follow-up study (Westerhausen et al., submitted) which also included instructions to attend to either the left or right ear stimulus, it was shown that the REA could withstand an intensity difference between the ears of 12 dB, thus showing that the effect of intensity is dynamic and partly dependent on the exact paradigm used. The use of interaural intensity differences is a way of manipulating stimulus parameters that affect the REA in a bottom-up manner. The use of instructions to attend to either the left or right ear stimulus is similarly a way of manipulating the REA in a top-down, or cognitive, manner. Hugdahl and colleagues showed already in 1986 (Hugdahl & Andersson, 1986) that the REA could be affected, and even switched to a left ear advantage (LEA) when the subjects were instructed to pay attention to and report from only the left ear stimulus, while it was significantly enhanced when instructed to pay attention to the right ear stimulus. It is therefore clear that the REA in dichotic listening, although a basic speech perception phenomenon, can be affected by both bottom-up sensory and top-down cognitive factors. It may be interesting in this context to note that the extent to which speech perception is influenced by e.g. attentional factors have been severely understudied in linguistics and phonetics, although the reason humans possess such a strong attentional cognitive module may be to filter out irrelevant from relevant speech signal inputs whenever there is more than one source of input at the same time.

4. Voice Onset Time and at-risk for dyslexia

Rimol et al. (2006) asked the question whether the right ear advantage would be sensitive to manipulation of phonetic characteristics of the dichotic stimuli, and whether it thus would be possible to manipulate other bottom-up factors than intensity differences between the left and right ear stimulus. The three stop consonants /b/, /d/, and /g/ are characterized by short voice onset time (VOT), i.e. the time between the initiation of the consonant to the onset of the vowel (also called voiced consonants), while the stop consonants /p/, /t/, and /k/ are characterized by long VOT (also called unvoiced consonants). By systematically combining syllables with short and long VOT, respectively, Rimol et al. (2006) found that presenting a long VOT syllable in the right ear and a short VOT syllable simultaneously in the left ear produced a stronger right ear advantage than the opposite combination, with the homonymic combinations showing an intermediate right ear advantage.

Thus, by systematic variation of the voicing parameter, it is possible to affect the right ear advantage in dichotic listening, with the consequence that lateralization of speech perception is not a unitary phenomenon but dependent on dynamic factors like voicing and VOT. The principle of VOT manipulation is seen in Figure 6. An extension

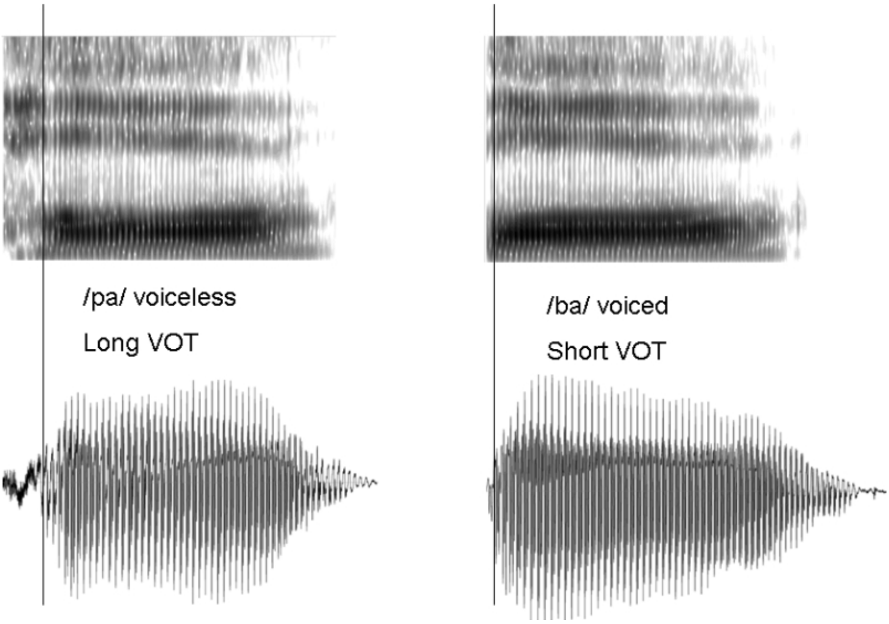


Figure 6. Spectrograms (upper) and oscillograms (lower), respectively, showing the onset of the vowel (vertical line) for a non-voiced and voiced CV-syllable.

of the discovery by Rimol et al. (2006) is that the VOT manipulation can be considered a test of phonological awareness in clinical groups. If an individual is not responding to the short-long manipulation, they could be operationally said to lack phonological awareness, only responding to the lateralization aspect of dichotic listening. This was used in a recent study in our laboratory on children at risk for developing dyslexia that were tested longitudinally from the age of 5 to 8 years. Dyslexia is traditionally defined as failure of phonological decoding (Høien & Lundberg, 2000), meaning that individuals with dyslexia have problems finding the corresponding phoneme structures of the orthography when reading. If so, it could be predicted that individuals with dyslexia would fail to show changes in the right ear advantage in dichotic listening as a consequence of the VOT manipulation. We have taken this prediction one step further by studying children as young as 5 years that have been identified as being at risk for dyslexia. Since phonological awareness develops along with the ability to decode the phonological structure when learning to read, it could be predicted that failure to develop an adequate phonological awareness would show up as a failure to chain the right ear advantage as a result of manipulating the VOT parameter.

We studied this in a group of at-risk children compared with a matched control group. At-risk was identified from a questionnaire given to nursery school teachers and parents containing questions related to "birth complications", "motor development", "speech development", "special education needs", and "dyslexia in the family"

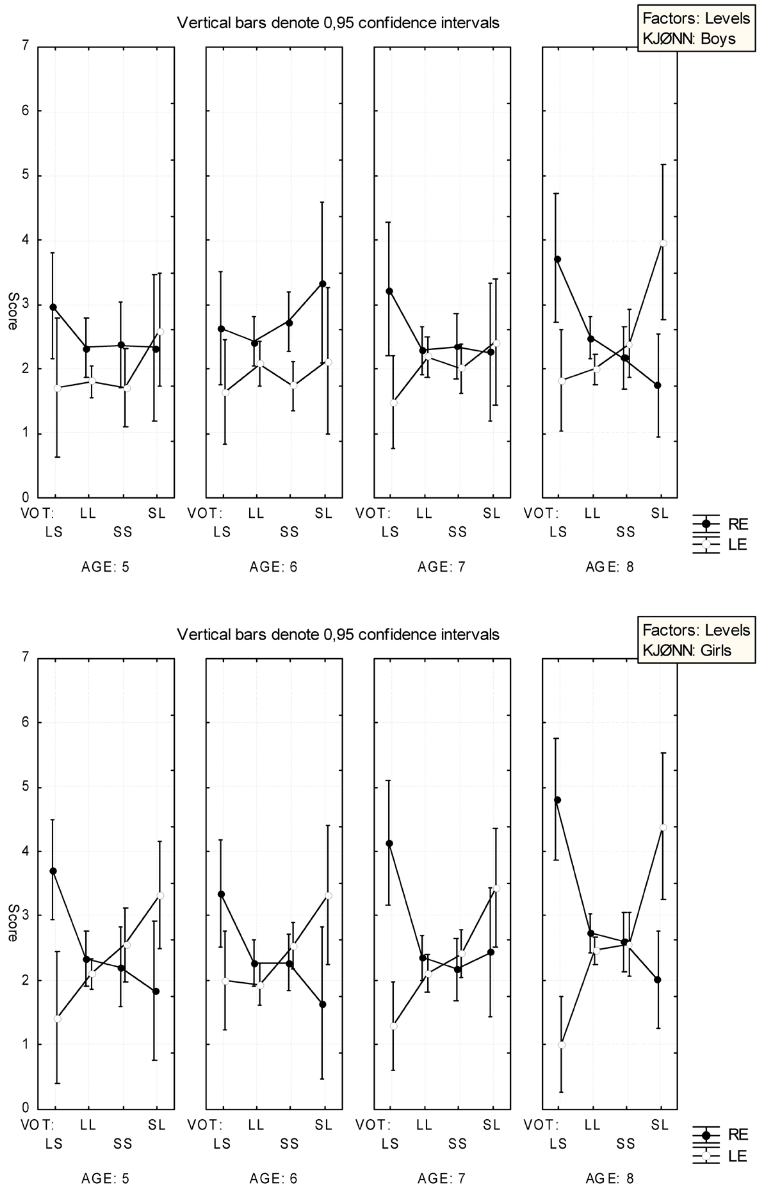


Figure 7. Correct report for boys (upper) and girls (lower) for the right versus left ear stimulus as a function of combining all four possible combinations of VOT (LS: long VOT syllable in the right ear and short VOT syllable in the left ear; LL: long VOT syllable in both ears; SS: short VOT syllable in both ears; SL: short VOT syllable in the right ear and long VOT syllable in the left ear).

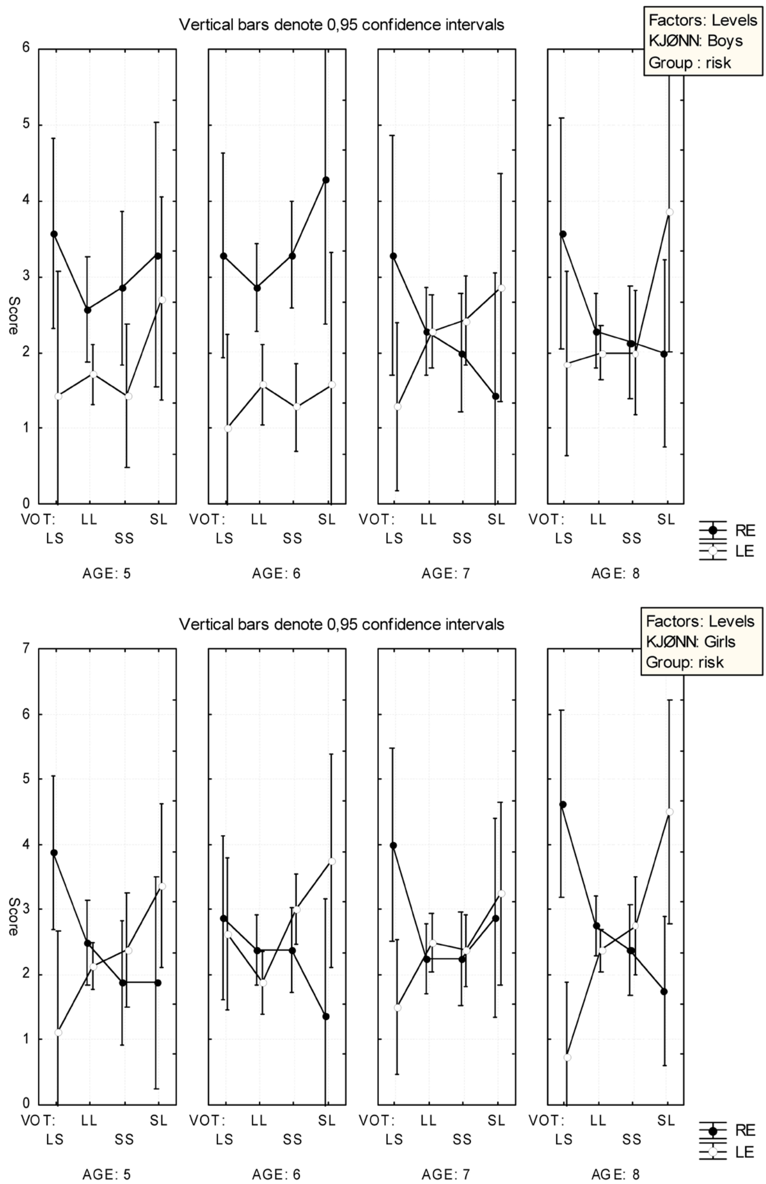


Figure 8. Correct report for at-risk boys (upper) and at-risk girls (lower) for the right versus left ear stimulus as a function of combining all four possible combinations of VOT (see Figure 7 for explanations).

(for further details, see Helland et al., 2006). The children were identified and screened from nursery schools in the three counties in western Norway and followed for four years with tests including, among others, dichotic listening (they were also scanned with fMRI at the age of 6 and 8). In addition to the at-risk and control groups, we also looked at how the ability to be aware of the VOT manipulation develops with age from 5 to 8 years, since this is the time period when most children would go from a pre-literate to a literate stage in their normal development. We could also compare boys and girls, with the hypothesis that girls would be sensitive to the VOT manipulation earlier than boys, from the known fact that girls mature faster than boys when it comes to reading. Figures 7–8 show the results, with Figure 7 showing the normal development from the age of 5 to 8, split for boys and girls, and Figure 8 showing the further split into at-risk and control groups. Figure 7 clearly shows how the ability to respond to the VOT manipulation increases with increasing age, with a qualitative shift from age 6 to age 7 years, which would correspond to the time when most children acquire the ability to decode phonology, i.e. to read. Thus, the VOT manipulation in dichotic listening seems to be a valid indicator of phonological awareness, correlating with reading ability. It is also obvious from a look at Figure 7 than the boys lag behind the girls in this respect, with a less marked "X" structure of the response-curves, which again would correspond to the lag that boys show when they acquire reading ability. Figure 8 further shows that the at-risk children lag behind the control children in phonological awareness, especially the boys at-risk for dyslexia. Thus, the results show that the VOT manipulation in dichotic listening may be a novel way of operationally defining phonological awareness, in an objective and experimental manner, that could have clinical value.

The results shown in Figures 7 and 8 have also been correlated with tests for actual reading ability when the same children were 8 years of age (Helland et al., 2006), i.e. dividing the children according to their ability to read at the age of 8 and then comparing their performance on the dichotic listening test already at the age of 5 again showed that failure in changing the right ear advantage as a function of VOT is related to reading ability and development of reading. It could also be mentioned in this context that fMRI brain scanning of the same children at the age of 6 (Specht et al., in press) revealed significant differences between the at-risk and control groups, particularly in fronto-temporal areas when having to decode simple words. In more detail, the word stimuli were either regularly spelled using early-acquired rules ("alphabetic") or more complex ("orthographic"). fMRI responses differed between the groups with activation in the alphabetic and orthographic conditions in the left angular gyrus correlated with individual at-risk index scores, and activation in inferior occipito-temporal regions further indicated differential activation for the two groups related to orthographic processing. These results correspond with findings in adult dyslexics when processing written words. It thus appears that sensitivity to the cortical differentiation of reading networks is established prior to formal literacy training.

5. Summary and conclusions

In this article I have reviewed research done in our laboratory at the University of Bergen on speech perception and brain function, with a focus on the use of a dichotic listening procedure to reveal hemispheric differences for perception and cognitive processing of simple syllable stimuli. The dichotic listening situation provides a unique opportunity to study the interaction of bottom-up, or stimulus-driven, laterality effects, and how these can be modulated through instructions to attend to and report only the right or left ear syllable of the dichotic pair. We have moreover looked at the effects on frequency of correct reports of systematically pairing voiced and unvoiced syllables, with e.g. a voiced syllable in the left ear and simultaneously an unvoiced syllable in the right ear. Results show that systematic variation of voiced versus unvoiced syllable pairs affect the frequency of correct right ear reports in the subjects. We have finally used the voice-onset-time manipulation as a way of studying children at risk for developing dyslexia from the age of 5 to 8 years. Results show that children at-risk for dyslexia, and particularly boys, show a delay of up to two years compared to controls when they are responding to the voicing manipulation in the dichotic listening situation. We interpret this to mean that they are delayed with regard to their control peers in developing phonological awareness, which is necessary for the acquisition of reading skills.

References

Benke, T., Bartha, L., Delazer, M. & Hugdahl, K. (1999). Effects of left-hemispheric lesions on dichotic listening performance. *Brain and Language*, *69*, 372–374.

Bryden, M. P. (1963). Ear preference in auditory perception. *Journal of Experimental Psychology*, *65*, 103–105.

Gadea, M., Gomez, C., Gonzalez-Bono, E., Espert, R. & Salvador, A. (2005). Increased cortisol and decreased right ear advantage (REA) in dichotic listening following a negative mood induction. *Psychoneuroendocrinology*, *30*, 129–138.

Heiervang, E., Hugdahl, K., Stevenson, J., Smievoll, A. I., Ersland, L., Lund, A., Lundervold, A. & Steinmetz, H. (2000). Planum temporale, planum parietale and dichotic listening in dyslexia. *Neuropsychologia*, *38*, 1704–1713.

Helland, T., Ofte, S. H. & Hugdahl, K. (2006). "Speak up!" A longitudinal study of children at-risk of developing language, reading, writing, and mathematics impairment. In A. E. Asbjørnsen (Ed.), *Proceedings from Nordic Network in Logopedics*. Bergen: University of Bergen.

Helland, T., Asbjørnsen, A., Hushovd, A. E. & Hugdahl, K. (2008). Dichotic listening and school performance in dyslexia. *Dyslexia*, *14*, 42–53.

Høien, T. & Lundberg, I. (2000). *Dyslexia: From theory to intervention.* Dordrecht, NL: Kluwer Academic Publishers.

Hugdahl, K. & Andersson, B. (1984). A dichotic listening study of differences in cerebral organization in dextral and sinistral subjects. *Cortex*, *20*, 135–141.

Hugdahl, K. & Andersson, L. (1986). The "forced-attention paradigm" in dichotic listening to CV-syllables: A comparison between adults and children. *Cortex*, *22*, 417–432.

Hugdahl, K. & Wester, K. (2000). Neurocognitive correlates of stereotactic thalamotomy and thalamic stimulation in Parkinson patients. *Brain and Cognition*, *42*, 231–252.

Hugdahl, K., Brønnick, K., Kyllingsbæk, S., Law, I., Gade, A. & Paulson, O. B. (1999). Brain activation during dichotic presentations of consonant-vowel and musical instruments stimuli: A 15O-PET study. *Neuropsychologia*, *37*, 431–440.

Hugdahl, K., Carlsson, G. & Eichele, T. (2001). Age effects in dichotic listening to consonant-vowel syllables: Interactions with attention. *Developmental Neuropsychology*, *20*, 449–457.

Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S. & Hämäläinen, H. (2008). The effect of stimulus intensity on the right ear advantage in dichotic listening. *Neuroscience Letters*, *431*, 90–94.

Kimura, D. (1961a). Some effects of temporal-lobe damage on auditory perception. *Canadian Journal of Psychology*, *15*, 156–165.

Kimura, D. (1961b). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology*, *15*, 166–171.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex*, *3*, 163–168.

Løberg, E. M., Hugdahl, K. & Green, M. F. (1999). Hemispheric asymmetry in schizophrenia: A "dual deficits" model. *Biological Psychiatry*, *45*, 76–81.

Pollmann, S., Maertens, M., von Cramon, D. Y., Lepsien, J. & Hugdahl, K. (2002). Dichotic listening in patients with splenial and nonsplenial callosal lesions. *Neuropsychology*, *16*, 56–64.

Rimol, L. M., Specht, K. & Hugdahl, K. (2006). Controlling for individual differences in fMRI brain activation to tones, syllables and words. *Neuroimage*, *30*, 554–562.

Sommer, I.E, Aleman, A, Somers, M, Boks, M & Kahn, S. (2008). Sex differences in handedness, asymmetry of the Planum Temporale and functional language lateralization. *Brain Research, 1206*, 76–88.

Specht, K., Hugdahl, K, Ofte, S, Nygård, M, Bjørnerud, A, Plante, E. & Helland, T. (in press). Brain activation reveals at-risk for dyslexia in 6-year old children. *Scandinavian Journal of Psychology.*

Steinmetz, H., Rademacher, J., Huang, Y., Hefter, H., Zilles, K., Thron, A. & Freund, H. J. (1989). Cerebral asymmetry: MR planimetry of the human planum temporale. *Journal of Computer Assisted Tomography*, *13*, 996–1005.

Tallus, J., Hugdahl, K., Alho, K., Medvedev, S. & Hämäläinen, H. (2007). Interaural intensity difference and ear advantage in listening to dichotic consonant-vowel syllable pairs. *Brain Research*, *1185*, 195–200.

Van den Noort, M., Specht, K., Rimol, L. M., Ersland, L. & Hugdahl, K. (2008). A new verbal reports fMRI dichotic listening paradigm for studies of hemispheric asymmetry. *Neuroimage*, *40*, 902–911.

Westerhausen, R. & Hugdahl, K. (2008). The corpus callosum in dichotic listening studies on hemispheric asymmetry: A review of clinical and experimental evidence. *Neuroscience & Biobehavioral Reviews*, *PMID*, 18499255 [PubMed – in process].

Westerhausen, R., Grüner, R., Specht, K. & Hugdahl, K. (submitted). Functional relevance of individual differences in temporal lobe callosal pathways: A DTI tractography study.

Linguistic analysis of spontaneous language in aphasia

Elisabeth Meffert, Katja Hussmann, Marion Grande
Department of Neurology, University Hospital Aachen

Abstract

In the field of aphasia, the importance of the analysis of spontaneous language in research as well as in clinical practice is widely accepted. However, clinicians often refrain from conducting detailed analyses of spontaneous language in daily routine, as most methods are very complex and time-consuming. Additionally, as to methodology, many questions remain open from the few studies conducted on spontaneous language so far. This article systemises different methods of elicitation and analysis of spontaneous language and summarises some of the general problems of spontaneous language analysis. Furthermore, the quantitative linguistic analysis of spontaneous language by means of basic parameters is described. Basic parameters are essential units of language like word categories or syntactic completeness and can be identified in both impaired and unimpaired adult and child language. So far, they have not been used exclusively as metrics for the analysis of spontaneous language, as most previous analyses focus on linguistic deviations. Finally, it will be discussed to what extent this method facilitates the use of spontaneous language analysis in clinical routine.

1. Introduction

When asked about the relevance of the analysis of spontaneous language in speech and language therapy, most therapists would probably agree upon the significant contribution of such analyses to the assessment and treatment of patients with communicative impairment. Yet, despite its undoubted importance, the analysis of spontaneous language is not always part of clinical practice. This may be due to several reasons. To mention some, the transcription and analysis of spontaneous language is very time-consuming. Second, many analyses are very complex and require a high level of linguistic knowledge. Third, there is not much research on this topic, resulting in many open questions concerning for instance the psychometric properties of the analysis of spontaneous language, the lack of normative data or the consequences of diagnosis for therapy (see also Prins & Bastiaanse, 2004).

This paper will give a short overview about terminology, relevance, and different methods of elicitation and analysis of spontaneous language. Furthermore, analysis of spontaneous language by means of so-called basic parameters will be described from the example of the computer-assisted instrument ASPA (see below). Finally, it will be discussed to what extent the findings from the use of basic parameters facilitate the use of analyses of spontaneous language in clinical routine.

1.1. Terminology

In the context of speech and language therapy, there are different terms for the concept of continuous speech-language production, e.g. spontaneous speech, spontaneous language, continuous speech, discourse. To some extent, there is a linkage between the aim of the analysis and the term used. As an example, the term “continuous speech” is mainly applied when articulatory or voicing parameters are analysed, whereas “discourse” refers to discourse analysis, and the terms “spontaneous speech” or “spontaneous language” relate to analyses using linguistic parameters. However, none of the terms is clearly defined. In the following, as this paper focuses on linguistic analysis, the term spontaneous language will be used.

All terms have in common that the person assessed is asked to produce continuous speech or language, and the end product is analysed as to the metrics of interest. However, there is no standard method for the elicitation of continuous speech-language production. For the analysis of articulatory or voicing parameters, it may be sufficient to make a person read aloud, whereas for a linguistic analysis or conversation analysis, the situation in which the language probe is recorded should be as natural as possible.

1.2. Elicitation of spontaneous language

Yet, this naturalness is difficult to achieve as soon as an examiner has to create such a situation in order to obtain a language sample, be it in clinical routine or in research. The following elicitation methods are commonly used: picture description, story retelling, role plays, interviews, and naturalistic dialogues. Prins and Bastiaanse (2004) suggest to classify these methods as spontaneous (interview, dialogue) or semi-spontaneous (picture description, story retelling, role-plays) in order to signal that the results of the analysis of spontaneous language may vary when the language samples have been elicited by different methods. Another way of classifying elicitation methods is to subdivide in monologic and dialogic methods (e.g. Springer, 2004). Indeed, when comparing the language samples gained from different elicitation methods, some authors found systematic variability related to the elicitation method applied. Springer (2004) showed in a study with 8 agrammatic participants that both the proportion of complete and of finite sentences were significantly lower in dialogic situations than in monologic. The same pattern was found in the control group. Sproedefeld (2008) elicited spontaneous language of 10 agrammatic participants in four different situations (interview, story retelling, picture description, role play). The speech samples differed significantly in syntactic completeness and in the mean length of utterances (e.g. more complete sentences in picture description than in the interview).

To sum up, there is evidence that different elicitation methods may influence linguistic performance. Consequently, examiners should choose their methods carefully with respect to the aim of the analysis conducted. Whereas more structured situations like picture description may evoke maximal syntactic performance, role play or interviews may lead to a more natural situation reflecting communicative and pragmatic strategies and deficits. Furthermore, pre- and post-testing should always be conducted with the same elicitation method.

1.3. Clinical analysis of spontaneous language

Even though the terminology in the field of spontaneous language is not well-defined, the relevance of the analysis of spontaneous language in speech and language therapy is without controversy, and such analyses are conducted not only in aphasia, but also in child language and in dementia. Spontaneous language analysis is used in the domains of assessment, treatment, and evaluation of treatment, contributing to the following points:

- On the one hand, the analysis of spontaneous language allows for assessing different linguistic levels separately (phonetics, phonology, semantics, lexicon, morphology, syntax, pragmatics). On the other hand, in spontaneous language, language production can be assessed as a whole, with preserved and affected functions interacting. As an example, incomplete sentences can result from syntactic problems; however, they can also result from word finding difficulties. Which cause applies to a specific patient cannot be discovered from assessment on single word or sentence level but only arises from spontaneous language analysis.
- In aphasia, the analysis of spontaneous language contributes significantly to the classification of syndromes. This applies to the Aachen Aphasia Test (AAT, Huber et al., 1983), in which the accuracy of classification from the analysis of spontaneous language is nearly equivalent to the classification from the subtests. Furthermore, in the acute phase of aphasia when systematic testing is not rea-

sonable (e.g. Biniek, 1993), a first classification in fluent and non-fluent aphasia by analysing the spontaneous language is possible in most cases.

- In residual aphasia, spontaneous language is, due to its complexity, a sensitive indicator of remaining difficulties which would not appear any more in structured testing (e.g. Jaecks, 2006).
- Especially when elicited by natural conversation or interviews, spontaneous language is a highly overlearned task which can even be fulfilled when emotional or cognitive conditions do not allow for conducting structured testing (e.g. in shy children or persons with dementia).
- Spontaneous language reflects everyday life conditions and therefore allows for evaluating language deficits on the levels of functioning and activity (World Health Organisation, 2001). Furthermore, spontaneous language is a key to participation and therefore indicates potentials and limits in communication. In clinical settings, these potentials and limits can be best estimated in situations when spontaneous language is applied.
- In spontaneous language, not only deficits can be seen but also preserved functions and strategies applied in order to compensate for the language deficits. In order to derive treatment strategies, such information is just as important as detecting specific linguistic deficits.
- From the analysis of spontaneous language, treatment strategies aiming at enabling a patient for dealing with everyday life can be derived.
- Independent from the contents of speech and language therapy, the main aim is to transfer the items or functions which have been practised to everyday life communication. Whether this has been achieved can be best assessed by means of spontaneous language analysis, as spontaneous language is as close as possible to everyday life situations and therefore is suitable for the evaluation of treatment.

2. Methods of analysis

Depending on the aim of the analysis, there are different methods of analysing spontaneous language.

First, the assessment may focus on different levels. Linguistics can be assessed in order to define deficits in phonetics, phonology, semantics, lexicon, morphology, or syntax or for measuring change in one of these levels during time course (e.g. Holland et al., 1985). In order to assess functional communication and pragmatics, therapists may focus on communicative and pragmatic aspects of spontaneous language (e.g. Amsterdam-Nijmegen Everyday Language Test (ANELT), Blomert & Buslach, 1994; Functional Assessment of Communication Skills for Adults (ASHA-FACS), Frattali et al., 1995). Conversation analysis may be applied in order to assess the interaction between dialogue partners and their involvement in success and failure of communication, for instance with the aim of instructing a patient and his or her family members in supportive communication strategies (e.g. Bongartz, 1998; Bauer & Kulke, 2004).

Furthermore, analysis of spontaneous language may be conducted in a qualitative or a quantitative manner. Qualitative analyses are mainly descriptive and frequently use rating scales in order to estimate the parameter value (e.g. analysis of spontaneous language in the Boston Diagnostic Aphasia Examination (BDAE), Goodglass & Kaplan, 1972/1983; Goodglass et al., 2000; analysis of spontaneous language in the Aachen Aphasia Test (AAT), Huber et al., 1983). Such rating scales are very useful in daily clinical routine as they are quick and easy to administer. However, as such scales comprise only few points, they only allow for a rough estimation. With the goal of simplifying the rating system, some rating scales describe more than one linguistic level (e.g. in the AAT, the scale "syntactic structure" comprises syntactic aspects like sentence length and sentence complexity but also morphological aspects like the correct use of grammatical morphemes and function words). It may appear that in some patients exhibiting symptoms on both levels these aspects are represented by two distinct parameter values, which may complicate the rating procedure. Consequently, for many rating scales, interrater-reliability is not ideal, which again limits the scales' sensitivity to change.

For these reasons, other authors apply quantitative methods, in which certain phenomena are exactly counted. From these, parameters are derived and calculated (e.g. Vermeulen et al., 1989; Bastiaanse et al., 1996; Berndt et al., 2000). In order to improve the objectivity of these quantitative methods, the authors try do provide exact definitions for the phenomena of interest. Quantitative analyses are more sensitive to change than qualitative analyses. However, the expenditure of time needed for counting and calculating quantitative parameters hamper the application of these methods in daily clinical routine. This problem can be partially solved by using computer-assisted methods in which some processes are automated and thus speeded up. Automation further improves objectivity and interrater-reliability. Most computer-assisted analyses have been developed for English child language (e.g. LARSP, Crystal et al., 1976; SALT, Miller & Chapman, 1983; CLAN and CHILDES, McWhinney, 1995). There are some adaptations for aphasic language (Holland et al., 1985) and for German child language (COPROF, Clahsen & Hansen, 1991).

Most analyses use linguistic deviations as metrics for the analyses, e.g. aphasic symptoms in the field of aphasia. Yet, there are some problems going along with these measures. First, deviations are not easy to capture and to classify; for instance, in some cases, it may be difficult to decide whether a word is semantically adequate or not (semantic paraphasia). These problems affect interrater-reliability. Second, differentiation between impaired and unimpaired speakers is not always evident as normative data are lacking and, in addition, cannot be meaningfully derived from analyses using linguistic deviations. Instead, linguistic parameters as found in unimpaired speakers, so-called basic parameters, can be applied. Basic parameters are essential units of language and can be identified in every conceivable verbal output (cf. Biniek et al., 1991). Thus, they can be assessed in both impaired and unimpaired adult and child language. Basic parameters are defined on different levels (e.g. lexical and syntactic) and differ with respect to the degree of elaborateness (e.g. open/closed class in contrast to a more refined analysis of word classes or even of different types of verbs) depending on the particular research interest (Hussmann et al., 2006; Grande et al., 2008). Examples of basic parameters are the type-token-ratio (TTR), the mean length of utterances (MLU), or the percentage of open class words. Such parameters are widely used in different methods of analysis of spontaneous language, but rarely exclusively, as most analyses use both linguistic deviations and basic parameters as metrics.

By the use of basic parameters, the derivation of normative data of spontaneous language becomes possible, as they do not classify a phenomenon as deviant or not but allow for a neutral way of capturing linguistic characteristics. Consequently, by means of basic parameters, any language sample can be described. Furthermore, they are simple enough to cover basic lexical and syntactic characteristics in all severity degrees of aphasia. As basic parameters are easier to identify and to classify than linguistic deviations, interrater-reliability and objectivity values are expected to be better than in symptom-based methods.

2.1. Need of normative data

In order to decide whether a person's language sample is deviant or not, examiners depend on the existence of normative data indicating a normal range and cut-off values. Although most studies dealing with the analysis of spontaneous language report results of an experimental group and a control group, there are hardly any generally accepted normative data from samples of representative size and composition, which additionally would be needed separately for every national language. This lack is partially due to the fact that most of the analyses are symptom-based, thus not allowing for the derivation of normative data. Instead, as discussed above, by the use of basic parameters, the derivation of normative data becomes possible and, at the same time, is necessary, as not all parameters are expected to reach 100% in the normal-speaking population. For instance, it is not at all evident whether 70% of complete sentences are within the normal range or not.

However, it is widely discussed to what extent normative data of spontaneous language are useful at all, as performance may vary interindividually with many factors, e.g. age, education, sex, method of elicitation, examiner, situation of assessment, health, etc. Indeed, several authors have shown systematic variability of spontaneous language depending from age, education, and sex (Ardila & Rosselli, 1996; Le Dorze & Bédard, 1998). Other factors may also cause systematic variability but have not been examined so far. Consequently, parameters of spontaneous language should always be tested for their sensitivity to the most prominent factors (age, education) in control speakers. If necessary, age- and education-dependent normative values should be developed.

Another question to be answered with the definition of normative data is to what extent the chosen parameters allow for a differentiation between impaired and unimpaired speakers and which parameters are most prominent in this regard. As for many parameters a considerable variance is presumed both within the control speakers and in the aphasic population, one would expect a notable overlap of the two ranges. Consequently, when normative data are established for specific parameters, their discriminative power should be tested as well.

2.2. ASPA: Quantitative linguistic analysis using basic parameters

The computer-assisted instrument ASPA (Aachener Sprachanalyse; Huber et al., 2005) designed for the analysis of German spontaneous language allows for a detailed quantitative analysis using basic parameters on the lexical and the syntactic level. Barthel et al. (2006) have proven ASPA to be a valid method with high intra- and interrater reliability. The use of this method requires transcription of the participant's spontaneous language, which can be done within the program or by importing an already existing text file. The program analyses the transcript based on several parameters referring to the word and sentence level. The elementary categories are words, interjections and neologisms, open and closed class words, complete, incomplete and elliptic clause-like units (CLUs), as well as simple and complex CLUs. Each category is indicated both in absolute numbers and in its proportion relative to other categories (e.g. proportion of words in relation to interjections and neologisms). Each category can refer to the whole sample or exclusively to complete, incomplete, or elliptic CLUs (e.g. the proportion of words in complete CLUs). In addition, mean length of utterances (MLU) is indicated (and can also refer to the whole sample or to a special type of CLU), and type–token ratio is calculated separately for open and closed class words.

The studies using ASPA conducted so far have mainly focused on the following five basic parameters (for a detailed description, see Grande et al., 2008):

1. Percentage words (W), a measure of the lexical content of a speech sample.
2. Percentage open class words (OCW), a measure of the semantic content of a speech sample.
3. Syntactic completeness (COMPL), i.e. the percentage of syntactically complete CLUs.
4. Syntactic complexity (CPX), i.e. the percentage of CLUs in compound sentences.
5. Mean length of utterances in words (MLU).

The computer-assisted mode eases transcription and analysis significantly. A transcript of 30 CLUs, a transcript length shown to be sufficient for ASPA, requires about 15–40 minutes of transcription and editing from a trained examiner, depending on the fluency of the patient. Control speakers are transcribed even more quickly. Analysis is conducted automatically, with the results being displayed in tables and graphs. Transcription and analysis can thus be conducted within an acceptable amount of time (Hussmann et al., 2006).

ASPA has been used for the analysis of spontaneous language in the field of aphasia, impaired child language and unimpaired language of children and adults, with various research aims to be followed. Some results will be summarised in the following.

Hussmann et al. (2006) examined the spontaneous language of 14 participants with fluent aphasia by means of ASPA before and after 7 weeks of intensive language training. The aim was to investigate whether basic parameters were more sensitive to change in spontaneous language than the rating scales from the Aachen Aphasia Test (AAT, Huber et al., 1983). Critical differences were calculated for each parameter in order to detect significant change in basic parameters. Seven participants exhibited change in at least one basic parameter, whereas only three participants showed change on the AAT spontaneous language rating scales (all of which also showed change in at least one basic parameter). Four participants did not show any change in spontaneous language. Altogether, these findings support the view that quantitative analyses by means of basic parameter reflect even small changes in spontaneous language in the course of aphasia whereas such small changes remain undetected by conventional rating scales.

Similar results were found in a bigger sample (Grande et al., 2008) of 28 aphasic participants of which 14 had non-fluent and 14 fluent aphasia. The procedure was the same as in the study of Hussmann et al. (2006). Of the participants, 20 showed change in spontaneous language, 16 of which showed change in at least one basic parameter but not on the AAT rating scales; 4 participants showed change both in basic parameters and on the rating scales; and 9 participants showed no change in spontaneous language. Consequently, there is further evidence that basic parameters are more sensitive to small changes in spontaneous language than conventional rating scales, not only in fluent but also in non-fluent aphasia. However, the authors discussed the question of how to classify change as improvement or deterioration. For basic parameters, such a classification is less clear-cut than for symptom-based analyses. Grande et al. (2008) defined improvement and deterioration for each of the applied parameters separately. For some (e.g. the parameter syntactic complexity), the data from control speakers were necessary in order to define whether the parameter values of aphasic participants were going beyond the normal range, which would not represent improvement in the respective parameter although there could be significant change.

Bay (2007) analysed the language samples of 60 adults without language impairment (17–79 years, 8–21 years of education, 30 female) in three age groups and two educational groups by means of ASPA and derived normative values for six basic parameters in German language. Even though she found some minor influences of age and education, most parameters were robust against these factors, so that there was no need to further differentiate between age groups or educational groups. Discriminative power of the applied parameters was investigated by comparing the sample of control speakers to the 28 aphasic participants from the study of Grande et al. (2008). Differentiation between aphasics and controls was excellent (88% of correctly reassigned cases), with syntactic completeness being the most important parameter for discriminating between the two groups. However, some single cases with fluent aphasia were reassigned to the group of controls. This may be due to the fact that the chosen basic parameters did not clearly reflect the semantic and phonologic problems which are prominent in fluent aphasia. Consequently, basic parameters allow for differentiating between language samples of aphasic and control speakers. For fluent aphasia, additional qualitative analyses may improve the discriminative power.

Hussmann et al. (2008) report further evidence for the clinical applicability of ASPA for aphasia diagnosis on the one hand and for monitoring the course of aphasia on the other hand. Thus, quantitative analysis of spontaneous language by means of basic parameters supplements the diverse diagnostic methods in the field of aphasia.

3. Summary and conclusion

To sum up, there are a number of different methods to elicit and analyse spontaneous language. Although the analysis of spontaneous language is considered to be a method of high relevance in the field of aphasia, it is not always used in clinical practice. Some of the reasons have been discussed in the paragraphs above. The computer-assisted instrument ASPA allowing for a quantitative analysis of German spontaneous language by means of basic parameters offers partial solutions to some of the problems. First, the expenditure of time needed for transcription and analysis is markedly reduced, which eases the use of spontaneous language analysis in daily clinical routine. Second, validity and interrater-reliability of the analysis are high, thus providing a valuable instrument for the examination of language. Third, basic parameters have shown to be sensitive enough to reflect even small changes in spontaneous language, which is important for monitoring the course of aphasia and for showing treatment effects. Finally, the use of basic parameters makes it possible to derive normative data, which is necessary for aphasia diagnosis and the differentiation between impaired and unimpaired speakers. Normative values also allow for classifying changes in spontaneous language as improvement and deterioration.

Still, there remain a lot of open questions, as listed in Prins and Bastiaanse (2004). For aphasia therapy, one of the most important ones will be to learn about the relationship between diagnosis and treatment, in order to promote improvement in spontaneous language. In the words of Prins and Bastiaanse (2004, p. 1088):

> All this shows that there is a lot of work to do in the important, but relatively unexplored, field of spontaneous speech analysis. We can only hope that the still existing walls between the word level (formal testing) and text level (for discourse analysis) on the one hand, and between theoretical research and clinical practice on the other, will be broken down by interdisciplinary (linguistics, psychology, speech, and language therapy) collaboration.

References

Ardila, A. & Rosselli, M. (1989). Neuropsychological characteristics of normal aging. *Developmental Neuropsychology 5(4),* 307–320.

Bay, E. (2007). *Basisparameter der Spontansprache: eine Normierungsstudie für die Aphasiediagnostik.* Diplomarbeit, RWTH Aachen.

Barthel, G., Djundja, D., Meinzer, M., Rockstroh, B. & Eulitz, C. (2006). Aachener Sprachanalyse (ASPA): Evaluation bei Patienten mit chronischer Aphasie. *Sprache-Stimme-Gehör 30,* 103–110.

Bastiaanse, R., Edwards, S. & Kiss, K. (1996). Fluent aphasia in three languages: Aspects of spontaneous speech. *Aphasiology 10*, 561–575.

Bauer, A. & Kulke, F. (2004). Language exercises for dinner: Aspects of aphasia management in familiy settings. *Aphasiology 18* (12)*,* 1135–1160.

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E. & Schwartz, M. (2000). *QPA – Quantitative Production Analysis.* Hove: Psychology Press.

Biniek, R., Huber, W., Willmes, K., Glindemann, R., Brand, H., Fiedler, M. & Annen, C. (1991). Ein Test zur Erfassung von Sprach- und Sprechstörungen in der Akutphase nach Schlaganfällen. *Nervenarzt, 62,* 108–115.

Blomert, L. & Buslach, D. C. (1994). Funktionelle Aphasiediagnostik mit dem Amsterdam-Nijmegen-Everyday-Language-Test (ANELT). *Forum Logopädie 2/1994*, 3–6.

Bongartz, R. (1998). *Kommunikationstherapie mit Aphasikern und Angehörigen.* Stuttgart: Thieme.

Clahsen, H. & Hansen, D. (1991). *COPROF – Ein linguistisches Untersuchungsverfahren für die sprachdiagnostische Praxis.* Köln, Eigenverlag.

Crystal, D., Fletcher, P. & Garman, M. (1976). *The grammatical analysis of language disability*. London: Edward Arnold.

Frattali, C. M., Thompson, C. K., Holland, A. L., Wohl, C. B. & Ferketic, M. M. (1995). *Functional assessment of communication skills for adults (ASHA FACS).* Rockville MD: American Speech-Language-Hearing Association.

Grande, M., Hussmann, K., Bay, E., Christoph, S., Piefke, M, Willmes, K. & Huber, W. (2008). Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia. *International Journal of Language and Communication Disorders, 43*(4), 408–426.

Goodglass, H. & Kaplan, E. (1972). *The assessment of aphasia and related disorders.* Philadelphia: Lea & Febinger.

Goodglass, H. & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea & Febinger.

Goodglass, H., Kaplan, E. & Barresi, B. (2000). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia: Lippincott, Williams & Wilkins.

Holland, A. L., Miller, J., Reinmuth, O. M., Bartlett, C., Fromm, D., Pashek, G., Stein D. & Swindell, C. (1985). Rapid recovery from aphasia: A detailed language analysis. *Brain and Language 24,* 156–173.

Huber, W., Poeck, K., Weniger, D. & Willmes, K. (1983). *Aachener Aphasie Test.* Göttingen: Hogrefe.

Huber, W., Grande, M. & Springer, L. (2005). *Aachener Sprachanalyse. Handanweisung.* Vertrieb: Delta Systems, Aachen.

Hussmann, K., Grande, M., Bay, E., Christoph, S., Piefke, M., Willmes, K. & Huber, W. (2006). Aachener Sprachanalyse (ASPA). Computergestützte Analyse von Spontansprache anhand linguistischer Basisparameter. *Sprache-Stimme-Gehör, 30,* 95–102.

Hussmann, K., Grande, M., Bay, E., Christoph, S., Willmes, K. & Huber, W. (2008). Quantitative Analyse anhand linguistischer Basisparameter zur Status- und Verlaufsdiagnostik aphasischer Spontansprache. *Bulletin Aphasie und verwandte Gebiete 23*(2), 23–36.

World Health Organisation. (2001). *International Classification of Functioning, Disability and Health*, http://www.who.int/classifications/icf/en/

Jaecks, P. (2006). *Restaphasie. Eine empirische Untersuchung von linguistischer Symptomatik, Gesprächsverhalten, Differentialdiagnose und Ursache minimal aphasischer Störungen nach Schlaganfall.* Dissertation, Universität Bielefeld.

Le Dorze, G., Bédard, C. (1998). Effects of age and education on the lexico-semantic content of connected speech in adults. *Journal of Communication Disorders 31*, 53–71.

MacWhinney, B. (1995). *The CHILDES-Project: Tools for Analyzing Talk.* Hillsdale, NJ: Lawrence Erlbaum.

Miller, J., Chapman R (1983). *Systematic analysis of language transcripts (SALT): A computer program designed to analyze free speech samples.* Language Analysis Laboratory, Waisman Center, University of Wisconsin-Madison.

Prins, R. & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology, 18*(12), 1075–1091.

Sproedefeld, A., Hussmann, K., Grande, M., Bay, E., Christoph, S., Willmes, K., Huber, W. (2008). Der Einfluss der Erhebungsmethode auf die spontansprachlichen Leistungen bei Aphasie. *Bulletin Aphasie und verwandte Gebiete 23*(2), 7–21.

Springer, L. (2004). *Medienspezifische Sprachperformanz. Eine empirische Studie zum Einfluss von Medialitätsmodus und Interaktivität bei Agrammatikern und Sprachgesunden.* Idstein: Schulz-Kirchner Verlag.

Vermeulen, J., Bastiaanse, R. & Van Wageningen B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language 36,* 252–274.

Appendix A:
List of speakers and poster presenters

Speakers

Mireille Besson	CNRS, Marseille
Ina Bornkessel-Schlesewsky	Max Planck Institute, Leipzig
Stefan Heim	Research Center Jülich, Jülich
Kenneth Hugdahl	University of Bergen
Francisco Lacerda	Stockholm University
Kerry Ledoux	Johns Hopkins University, Baltimore
Martin Meyer	University Hospital of Zurich
Inger Moen	University of Oslo
Leticia Pablos	University of Reading
Ann Pannekamp	Max Planck Institute, Leipzig
Mikael Roll	Lund University
Matthias Schlesewsky	Philipps University Marburg
Yury Shtyrov	Medical Research Council, Cambridge
Ulrike Toepel	Centre Hospitalier Universitaire Vaudois, Lausanne
Janne von Koss Torkildsen	University of Bergen

Poster presenters

Giorgos Argyropoulos	University of Edinburgh
Zoltán Bánréti	Hungarian Academy of Sciences
Elisabeth Meffert [Bay], Marion Grande, Katja Hussmann, Swetlana Christoph, Klaus Willmes, Martina Piefke, Walter Huber	University Hospital Aachen
Sarah Bihler, Dorothee Saur, Stefanie Abel, Dorothee Kümmerer, Walter Huber, Jürgen Dittmann, Cornelius Weiller	University Hospital Freiburg, University of Freiburg, University Hospital Aachen
Angèle Brunellière, Ulrich Hans Frauenfelder	University of Geneva
Francesca Citron, Brendan Weekes, Evelyn Ferstl	University of Sussex
Tina Ibertsson, Lena Ásker-Arnason, Kristina Hansson, Birgitta Sahlén	Lund University
Armina Janyan, Ivo Popivanov, Elena Andonova	New Bulgarian University, University of Bremen
Hamutal Kreiner, Sibylle Mohr, Klaus Kessler, Simon Garrod	University of Glasgow
Francisco Lacerda, Ulla Sundberg, Iris-Corinna Schwarz, Ulla Bjursäter, Ellen Marklund, Eeva Klintfors, Göran Söderlund, Lisa Gustavsson	Stockholm University
Frida Mårtensson, Merle Horne, Mikael Roll, Pia Apt	Lund University, Malmö University Hospital
Kathrin Pusch, Jessica Rosenberg, Rainer Dietrich	Humboldt University
Bengt Sigurd, Bengt Nilsson	Lund University
Sverker Sikström and Petter Kallioinen	Lund University
Jaana Simola, Kenneth Holmqvist, Magnus Lindgren	Lund University