Lexical competition and predictive certainty in speech recognition

Modulations of pre-activation negativity amplitude by continuations, entropy and inhibition

Anna Hjortdal

Supervisor: Professor Mikael Roll
Co-supervisor: Dr Johan Frid
Centre for Language and Literature, Lund University
MA in Language and Linguistics, Phonetics
SPVR02 Language and Linguistics: Degree Project – Magister's (One Year) Thesis, 15 credits

January, 2022

Abstract

A growing body of evidence suggests that speech recognition is facilitated by rapid activation of possible lexical candidates and subsequent competition and selection. An event-related potential (ERP) component 136-204 ms after word onset, the pre-activation negativity (PrAN), correlates with lexical competition. Further, the increased negativity for low lexical competition correlates with increased activity in Broca's area. The effect has been interpreted as reflecting predictive certainty about the unfolding word and inhibition of irrelevant candidates. The present study investigated effects of lexical competition, predictive certainty and inhibition when only the first 2-3 speech sounds of words were available to listeners. The measures were calculated from a combined pronunciation lexicon and frequency list with a Python script. Correlations with ERP data from two experiments investigating Danish language were explored. In line with previous findings, word beginnings with few continuations (low lexical competition) showed more negative ERPs than word beginnings with many continuations. For one experiment, there was an inverse correlation between the number of possible continuations and PrAN amplitudes. This is consistent with parallel activation of multiple lexical candidates as predicted by a number of models of speech perception and spoken word recognition. According to the distributed model of speech perception, there is an inverse relationship between the number of pre-activated words and their semantic activation. The increased PrAN for low competition might reflect stronger semantic activation of one or a few candidates. Word beginnings with low entropy (more certainty) also showed more negative amplitudes. Within the framework of the predictive coding model of speech recognition, the findings suggest an interplay between pre-activated lexical candidates and updated phonetic expectations at lower-level, primary processing areas. There were no effects of inhibition of irrelevant candidates immediately after word onset. When more of the speech signal became available, prosodic cues during or immediately after the stressed vowel which inhibited many candidates yielded more negative, right frontal effects than speech sounds inhibiting few candidates. One interpretation is that inhibition only started later, as more of the word unfolded.

Keywords: Speech perception, spoken word recognition, lexical competition, ERPs, PrAN.

Acknowledgements

Working on this paper this paper has taught me so much and I am immensely thankful to the people who have helped me. First of all, I want to thank my supervisors Mikael Roll and Johan Frid. I am grateful to Mikael Roll for engaged and inspiring discussions about neurophonetics, entropy and prediction - and for continuous and thorough feedback on my work. I have developed a lot academically the past year and this is to a large extent because you keep challenging me to do a little better. I want to thank Johan Frid for sharing your Python code and for patiently helping me take my first steps in programming. I am also thankful for valuable guidance on corpus studies and normal distribution. On the topic of programming, I also want to thank Marcus Nyström, Henrik Garde and Diederick Niehorster from the Humanities Lab for giving an incredibly pedagogical Python course. The participants in the experiments also deserve my gratitude for taking their time to take part in the studies.

Also, I am thankful to the Lund Neurolinguistics group, Mikael Roll, Merle Horne, Sabine Gosselke Berthelsen, Mikael Novén, Renata Kochančikaitė, Claudia Sjöström and Tugba Lulaci, for valuable discussions and support. I would also like to thank my fellow phonetics students for being inspiring company and student councilors Åsa Wikström and Peter Marthinsson for helping me navigate in being a Dane in the Swedish education system.

Finally, I am thankful for support and interest from friends and family. I am especially thankful to Rasmus Walther Jensen for continuous encouragement and patience.

Table of contents

List of figures

List of tables

Abbreviations

ANOVA	Analysis of variance
BOLD	Blood oxygen level dependent
EEG	Electroencephalography
ERP	Event-related potentials
fMRI	Functional magnetic resonance imaging
gRMS	Global root mean squares
IPA	International phonetic alphabet
LAFS	Lexical access from spectrum
PoS	Part-of-speech
PrAN	Pre-activation negativity
SAMPA	Speech assessment methods phonetic alphabet
WIF	Word-initial fragment

1 Introduction

Understanding spoken language is a complex task. Listeners must derive speech sounds from a noisy acoustic input, conjoin them into words and decode the message behind (Goldinger et al., 1996). Yet, listeners are capable of doing all this with a racing speed of up to 15 phonemes per second (Perkell, 2006). Understandably, the question of how this achievement is possible has puzzled linguists. According to most models of spoken word recognition, lexical candidates are (pre-)activated based on the initial acoustic input. As the speech signal unfolds, candidates compete against each other and are continuously inhibited when they become inconsistent with the acoustic signal. Within this framework, 'activation' of candidates, 'competition' and 'selection' among these are central and distinct processes (Gaskell & Marslen-Wilson, 1997; Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland, 2013; McClelland & Elman, 1986; Norris, 1994; Norris & McQueen, 2008). Although several studies have investigated these processes (Allopenna et al., 1998; Goldinger et al., 1989; Luce & Pisoni, 1998; Magnuson et al., 2007; Zhuang et al., 2011; Zhuang et al., 2014), the temporal distribution of the corresponding neural activity is still poorly understood. The aim of this thesis is to investigate neural correlates of lexical competition, attempting to segregate effects of lexical competition, predictive certainty about the unfolding word, and inhibition of irrelevant candidates. Event-related potentials (ERPs) are used to capture the timing of the neural activity.

Behavioural and neurophysiological studies have shown that spoken word recognition is modulated by 'lexical competition', that is, the number of words competing for recognition. Stimulus words with few lexical competitors are recognised faster and more accurately than words with many competitors (Goldinger et al., 1989; Luce & Pisoni, 1998; Vitevitch, 2002; Vitevitch & Luce, 1999), indicating that when many candidates are activated, resolution takes longer. Eye-tracking studies indicate that spoken words are continuously mapped onto lexical candidates consistent with the unfolding speech signal (Allopenna et al., 1998; Magnuson et al., 2007), and neuroimaging studies suggest that lexical competition and selection to a large extent are resolved in Broca's area (Righi et al., 2010; Roll et al., 2017; Zhuang et al., 2011; Zhuang et al., 2014). In

ERP studies with speakers of Swedish, a neurophysiological correlate of lexical competition has been identified: the pre-activation negativity (PrAN). PRAN is a negative component 136-204 ms after word onset modulated by lexical competition (Roll, 2015; Roll et al., 2017; Söderström et al., 2016). The ERP amplitude is more negative for word beginnings with fewer possible continuations and high frequency and correlates with overall increased neural activity and blood oxygen level depend (BOLD) effect in Broca's area. The PrAN has been associated with selection among pre-activated lexical candidates. Further, it is understood as reflecting predictive certainty: the fewer activated candidates, the stronger confidence the listener can have in them (Roll et al., 2017).

In the present study, ERP data from two experiments with speakers of Danish was reanalysed, investigating neural correlates of lexical competition. It was hypothesised that, as in Roll et al. (2017), word beginnings with few continuations would yield more negative PrAN amplitudes over left central sites 136-204 ms after word onset. A new measure, 'entropy', was introduced and its correlation with PrAN amplitude was investigated. Entropy is a measure of uncertainty and is modulated by the number of different outcomes of an event and their respective probabilities (Shannon, 1948). Entropy was used as a measure of listeners' certainty about the unfolding word upon hearing the first two speech sounds. Estimates of continuations and entropy were calculated for the items in the experiment by merging a Danish pronunciation lexicon and a frequency list and looping through it with a Python script developed for the purpose.

Further, since the PrAN has been interpreted as reflecting selection among activated lexical candidates and inhibition of irrelevant ones, a new hypothesis was derived. Speech sounds which were inconsistent with *many* already activated lexical candidates were expected to yield increased activity, reflecting inhibition of a large number of candidates. This would be seen in more negative PrAN amplitudes. To investigate this hypothesis, the number of continuations inhibited were calculated at two different time points as the speech signal in one of the experiments unfolded. The two time points were: 1) lexical candidates activated by the first speech sound and inhibited by the second speech sound and 2) lexical candidates activated by the first two speech sounds and inhibited by 'stød' or 'non-stød' prosodic cues in or immediately after the vowel in the stressed syllable. Stød is a prosodic creaky voice cue specific to Danish whereas non-stød is its modular voice counterpart (Fischer-Jørgensen, 1989). Further, the decrease in entropy from one time point to the next was calculated. This was a measure of listeners' increased certainty of the lexical candidate. The neurophysiological correlates of these processes were explored. The results are discussed in the light of evidence from behavioural and neuroimaging studies and models of speech recognition.

The thesis is organised as follows. In the theoretical background, models of speech recognition are presented, including how lexical competition is quantified according to those models. After this section, behavioural and neurophysiological evidence of lexical competition is presented, and the new measure 'entropy' is motivated. Danish language is also briefly introduced, including phonetic and phonological features of interest for the present study. Then, the present study is described in more detail, including research questions, hypotheses and test implications. This section also includes presentation of the exploratory approach to investigating inhibition and change in certainty. A materials and methods section describes how ERPs were collected and how lexical competition was calculated for the items in the studies. After this, results of the experiments are presented and briefly discussed, and finally, there is a general discussion of the overall findings.

2 Theoretical background

2.1 Models of speech perception and spoken word recognition

Understanding spoken language requires picking up sound waves as they reach the ear drum, frequency analysis by receptor neurons in the spiral ganglion and transmission through the auditory nerve to primary auditory cortex (Denes & Pinson, 1993). Further, listeners must derive units such as phonemes from the speech signal, map them onto lexical representations and eventually extract meaning (Goldinger et al., 1996). According to Goldinger et al. (1996), while early models have been concerned with *either* speech perception (i.e. identifying phonemes in the speech signal) *or* spoken word recognition (i.e. mapping phonemes onto lexical representations), a number of models now tackle the whole process. Therefore, in the following, models of speech perception and spoken word recognition are discussed together. The terms speech perception and spoken word recognition are used as defined above whereas ‘speech recognition’ encompasses the entire process. According to most models of spoken word recognition, a number of possible lexical candidates are activated and subsequently compete against each other until one is ultimately selected. The present study is concerned with lexical competition. Therefore, if applicable, it is discussed how the different models quantify and resolve lexical competition.

2.1.1 The Cohort model

In Marslen-Wilson et al.’s ‘Cohort model’ (Marslen-Wilson & Tyler, 1980; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), word-initial speech sounds activate all words whose beginnings correspond to the initial acoustic input, corresponding to the first 1-2 phonemes or 150-200 ms of the speech signal (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978). For instance, the speech sound [k^h] in Danish, would activate candidates such as [ˈk^{h}ad̥] *kat* ’cat’, [ˈk^{h}asd̥] *kast* ‘throw’ and [ˈk^{h}asə] *kasse* ‘box’. Words activated by the same word-initial speech sound are in the same ‘cohort’. Activation is also modulated by word frequency with more frequent words being activated faster than less frequent ones (Marslen-Wilson, 1987). As more and more of the acoustic signal becomes available, candidates that are no longer consistent with the input are weeded out until eventually only one candidate remains. This is called the 'word recognition point'. Candidates may also be ruled out if they are inconsistent with semantic and syntactic constraints.

When first proposed, the Cohort model was in opposition to models in which words were activated serially such as Forster's Autonomous Search Theory (Forster, 1976 as referenced by Goldinger, Pisoni & Luce, 1996). According to this model, words are looped through serially, one at a time, rather than parallelly. The Cohort model has been criticised for being unable to 'recover' if the first phoneme for some reason was distorted and a candidate never activated in the first place. Such candidates would have no way of being recognised (McClelland & Elman, 1986).

A central concept is 'cohort competition'. Cohort competitors are words sharing the first two phonemes and competition is quantified by taking the frequency of the target word divided by the summed frequency of all cohort competitors multiplied by 100 (Zhuang et al., 2011). Competing candidates do not inhibit each other, meaning the activation of one candidate does not deactivate others (Marslen-Wilson & Welsh, 1978).

2.1.2 The Neighbourhood Activation model

While the initial speech sound has a special status in the Cohort model, the 'Neighbourhood Activation model' (Goldinger et al., 1989; Luce, 1986; Luce & Pisoni, 1998) emphasises similarity of the entire word. Neighbours need not be in the same cohort, that is, start with the same speech sound, but are words which can be converted to a stimulus word by substituting, adding or deleting one phoneme. For instance, *hat* [ˈhad̥] 'hat' and *klat* [ˈkʰlad̥] 'blob' are neighbours of the Danish word *kat* [ˈkʰad̥] 'cat', but [ˈkʰasə] *kasse* 'box' is not because it deviates from *kat* with more than one phoneme.

Lexical competition can be quantified as 'neighbourhood similarity' and is characterised by the density (i.e. the number of words in a neighbourhood) and frequency (i.e. neighbourhood frequencies relative to stimulus word frequency) (Goldinger et al., 1996; Luce, 1986). Another measure is 'onset density' which combines measures from the Neighbourhood Activation model and the Cohort model. Onset density refers to the proportion of neighbours sharing the initial phoneme, (i.e. are in the same cohort). Words with many neighbours sharing the same first phoneme have 'dense onsets' while words with few neighbours sharing the same first phoneme have 'sparse onsets' (Vitevitch, 2002). When listeners hear a word, all items in the similarity neighbourhood – those corresponding to words as well as non-words – are activated. As the speech signal progresses, the activation level of the stimulus word increases and those of neighbours decrease until one word reaches a threshold and is selected (Goldinger et al., 1989). Like in the Cohort model, competitors do not inhibit each other.

2.1.3 Lexical access from spectrum (LAFS)

Klatt (1979)'s lexical access from spectrum (LAFS) differs from the Cohort model and Neighbourhood Activation model in that it does not have a phonemic level. According to this model, spectral representations of the acoustic input are mapped directly onto lexical candidates without any intervening level. As a solution to the problem of coarticulatory effects, the model assumes that listeners have mental representations of all phonotactically permitted diphones. For English, this would sum up to a mental dictionary of around 2000 diphones. Speech processing and spoken word recognition implies finding the best path through a network of diphones. This is accomplished by analysing the acoustic input every 10 ms and comparing it to stored paths in the network in order to deduce the best hypothesis. Hypotheses not consistent with the acoustic input are weeded out. A lexical decision is reached when all but one hypothesis are deemed unlikely to increase in likelihood as the speech signal progresses.

2.1.4 Fuzzy logical model of perception

The fuzzy logical model of perception has three stages: 'Feature evaluation', 'prototype matching' and 'pattern classification' (Massaro & Cohen, 1991; Oden & Massaro, 1978). First, each acoustic cue in the acoustic string is perceived independently and features are evaluated to determine whether a certain feature is present. These features are continuous or "fuzzy" rather than binary. The speech signal is subsequently compared to prototypes and identified based on how well it matches different prototypes. As in the LAFs model, the stored prototypes are diphones rather than phonemes because realisation of consonants is heavily dependent on vowel context. In spoken word recognition, the fuzzy logical model of perception integrates information from the bottom-up speech signal as well as linguistic context, but the fuzzy logical model of perception is bottom-up in the sense that feedback from higher levels such as the lexical level cannot affect perception on lower levels (Massaro & Cohen, 1991).

2.1.5 Connectionist models

Connectionist approaches to speech perception and spoken word recognition are based on artificial neural networks which mimic the brain. 'Units' are analogues of neurons or groups of neurons, typically organized in three hierarchical layers. 'Connections' represent synapses, through which units can exchange information (Stufflebeam, 2006). The most famous connectionist models are probably TRACE (McClelland & Elman, 1986; McClelland et al., 2014) and the Shortlist models (Norris, 1994; Norris & McQueen, 2008). Both models resemble the Cohort model in that phonemes and words are first activated based on bottom-up information from the initial speech signal. Candidates are subsequently inhibited – but, contrary to the Cohort model, not completely eliminated – when they become inconsistent with the speech signal. The model can therefore 'recover' as the signal progresses if, for some reason, the initial speech sound was distorted (McClelland, 2013). The most noticeable difference between the Shortlist and TRACE is that while TRACE has 'bidirectional' connections between levels, Shortlist is 'unidirectional'. This means that in TRACE, activation on higher levels, for example the word level, can activate lower levels, for instance the phonemic level. Norris et al. (2016) term this 'activation feedback'. In the Shortlist models, on the other hand, activation is purely feedforward from lower to higher levels (Norris et al., 2016).

TRACE (McClelland & Elman, 1986) has processing units called 'nodes' on three levels: A feature level, a phonemic level and a word level. The feature level has continuous – rather than binary – speech sound dimensions like the fuzzy logical model of perception (Massaro & Cohen, 1991; Oden & Massaro, 1978) such as 'consonantal', 'vocalic' and 'diffuseness'. The phoneme and word levels have units for all phonemes and words in a language (McClelland & Elman, 1986). The nodes are connected through excitatory and inhibitory connections. Mutually consistent nodes on adjacent levels have excitatory connections while nodes on the same level have inhibitory connections. This is termed ‘lateral inhibition’. As another neuroanatomical analogue, the nodes ‘fire’ when a certain threshold is reached and information passed on through excitatory connections (McClelland et al., 2014). Shortlist A (Norris, 1994) resembles TRACE in many ways. A difference is that only a small subset of lexical competitors, a so-called ‘shortlist’, of words is activated rather than all words consistent with the initial speech sound. Also, as mentioned above, activation on higher cannot affect lower levels.

Since first proposed in the 1980’s and 1990’s, both TRACE and Shortlist have been updated and are now both based on Bayesian principles (McClelland et al., 2014; Norris & McQueen, 2008). In Bayesian models of speech perception and spoken word recognition, central concepts are ‘prior’ and ‘posterior’ probabilities. The ‘posterior probability’ is the likelihood that some hypothesis is true. Its likelihood depends on 1) given evidence, 2) the ‘prior probability’, which is the probability of the hypothesis before the evidence is presented, and 3) the probability of the evidence if the hypothesis were true (McClelland, 2013). For speech recognition, this can be exemplified with the likelihood of hearing the Danish word *hatten* ’the hat’ if the sentence context is *Jeg købte…* ‘I bought…’ and the speech input, at an early stage, is [ha]. A prior probability could correspond to the general frequency of the word *hatten* ‘the hat’ in Danish (in the present study, *hatten* had a frequency of 976) and grammatical and semantic constraints such as the fact that ‘hat’ is a noun and is buyable. The evidence would be the acoustic input [ha]. The probability of the evidence if the hypothesis were true is the likelihood of hearing [ha] if the word is indeed *hatten* ‘the hat’ (relatively high, since *hatten* ‘the hat’ does start with those speech sounds). The posterior probability would correspond to the probability that *hatten* ‘the hat’ is heard, given the prior probability (context, word frequency) and the probability of hearing the evidence [ha], if the word is *hatten* ‘the hat’. Thus, in Shortlist A’s successor Shortlist B (Norris & McQueen, 2008), the term ‘activation’ is replaced by ‘likelihood’ and ‘probability’ of hearing a certain word. The model postulates that over time, listeners learn what is the most likely lexical candidate behind a specific speech signal. There is an interplay between frequency and perceptual effects: if the perceptual evidence is poor, frequency is more important, but as the perceptual input improves, frequency effects are diminished. However, the main principles remain the same as in Shortlist A: A shortlist of lexical candidates is taken into account based on bottom-up speech input and the idea of activation feedback is still refuted – contrary to TRACE's successor 'the Interactive Activation Hypothesis' (McClelland et al., 2014).

The so-called 'Ganong effect' is often drafted as evidence of activation feedback. Ganong (1980) asked listeners to label word-initial speech sounds which were ambiguous on a voicing spectrum between /g/ and /k/ and found that listeners tended to make phonemic categorisations which made real words. For instance, the same speech sound would be labelled /g/ when followed by /ift/ and /k/ when followed by /is/, indicating that feedback from the lexical level affects perception on lower levels (McClelland & Elman, 1986). In a similar way, ambiguous words are more likely to be labelled as the high-frequency word (Connine et al., 1993). According to Norris et al. (2016), Bayesian models such as Shortlist B provide another explanation for context and frequency effects: Real and high-frequency words have higher prior probabilities than non-words and low frequency words and are therefore more likely to be recognised. Norris and McQueen (2008) argue that activation feedback could mislead perception and induce hallucinations by boosting activation of phonemes which in turn boost activation of the word – and so on. McClelland et al. (2014) acknowledge this possibility, but say that this is exactly what to expect. 'Hallucinations' are even useful in recovering distorted phonemes (McClelland, 2013; McClelland et al., 2014).

A third connectionist model is Gaskell and Marslen-Wilson (1997)'s 'distributed model of speech perception'. As in the Cohort model, several candidates are activated based on bottom-up information from the speech input. Similar to the LAFS model (Klatt, 1979), the acoustic input is mapped directly onto representations without any intervening phonemic level. Setting the model apart from other connectionist models, representations are 'distributed' and represent both phonology and semantics of pre-activated words. If more than one lexical candidate is consistent with the speech signal, the semantic activation is a blend of the activated candidates. Partial semantic representations are available for all activated words, but if a large cohort is activated, the representations are degraded. At any time point after initial activation, the activation of different lexical candidates depends on the number of words activated and their relative frequencies.

2.1.6 A predictive coding model of speech recognition

Inspired by a birdsong model, Yildiz et al. (2013) proposed a Bayesian model of speech recognition based on ‘predictive coding’. According to theories of predictive coding, the brain has an ‘internal generative model’ of the surrounding world (Friston, 2005; Friston, 2010). Based on this internal model, impressions such as auditory and visual sensations are mapped onto objects and events that most likely caused them. ‘Recognition’ is equal to inferring a cause from a sensory input – for instance recognising a spoken word based on some acoustic input. Further, predictions about sensations about to be encountered in the immediate future are constantly generated. The goal is to minimise ‘prediction error’ which is surprise when predictions and the actual sensory input do not match (Friston, 2005; Friston, 2010).

Yildiz et al. (2013)’s generative model is ‘hierarchical’, mirroring the hierarchical organisation of the human (and songbird) auditory system. A hierarchical organisation means that lower-level, less complex representations such as features are encoded closer to primary auditory cortex whereas more complex, high-level representations such as phonemes, words and phrases are encoded along a forward-going spatial complexity gradient (DeWitt & Rauschecker, 2012). In Yildiz et al. (2013)’s speech recognition model, incoming speech is first mapped onto neural activity on a ‘cochlear level’. The cochlea is a spiral-shaped cavity in the inner ear. In the cochlea, incoming sound waves set the basilar membrane into vibration, setting frequency-sensitive hair cells into motion which in turn sends pulses to connected fibres in the auditory nerve (Denes & Pinson, 1993). The output of Yildiz et al. (2013)’s cochlear level is a cochleagram which maps the neuronal activity based on frequency-specific firing rates in the auditory nerve. Activity is fed forward to a two-level hierarchical model. On the first level, neural ensembles (i.e. a group of neurons) represent spectral features of the cochleagram. Like the TRACE model (McClelland & Elman, 1986; McClelland et al., 2014), the predictive coding model of speech recognition has bidirectional connections. Predictions are sent from the second level to the first level, encoding second-level expectations about activity at the first level. These expectations are compared to the actual activity and predictions errors, reflecting differences between predictions and reality, are sent from the first to the second level. Thus, recognition can happen incrementally, while speech is unfolding. The authors suggest that the second level could correspond to the inferior frontal gyrus, pars opercularis in Broca's area.

2.2 Behavioural studies of lexical competition

Words with high lexical competition yield longer response times and lower response accuracy than words with low lexical competition (Goldinger et al., 1989; Vitevitch, 2002; Vitevitch & Luce, 1999; Zhuang et al., 2011; Zhuang et al., 2014). As mentioned above, 'lexical competition' is quantified in different ways depending on the model of spoken word recognition. Words from sparse and low-frequency neighbourhoods, as employed in the Neighbourhood Activation model, are recognised more quickly and accurately than words from dense and high-frequency neighbourhoods, presumably because it takes longer to resolve competition between different activated patterns (Goldinger et al., 1989; Luce, 1986). Similarly, words with high cohort competition yield longer response times (Zhuang et al., 2011; Zhuang et al., 2014) and words with sparse onsets are named and recognised faster than words with dense onsets (Vitevitch, 2002). Eye-tracking studies have revealed continuous effects of lexical competition (Allopenna et al., 1998; Dahan et al., 2001b; Magnuson et al., 2007). Eye-tracking studies can reveal fine-grained effects of continuously unfolding speech which are not captured by response time measures given at one point in time. Lexical competition has typically been investigated in 'visual world paradigms' (Allopenna et al., 1998) in which participants look at images on a computer screen and listen to stimuli instructing them to perform tasks such as 'move the beaker'. Typically, the images depict a target word (e.g. *beaker*) as well as distractors, one or more of which are lexical competitors (e.g. *beetle*). The proportion of fixations to the images on the screen are interpreted as reflecting lexical activation over time. Allopenna et al. (1998) reported effects of both cohort and 'rhyme' activation. The latter refers to words rhyming with the target word. Early in the word recognition process, there were more fixations to target words and cohort competitors like *beaker* and *beetle* compared to unrelated and rhyme distractors. As more of the acoustic signal became available, fixations to cohort competitors dropped while fixations to rhyme competitors, like *speaker*, rose. The activations for cohort competitors could support most models, but the rhyme effect was taken as support for connectionist models such as TRACE and Shortlist (Allopenna et al., 1998). Further, they contradict predictions from the Cohort model which cannot explain how candidates can re-enter competition when they have been ruled out by initial speech sounds.

In addition to lexical competition affecting response times, a bias towards high-frequency words has also been observed (Connine et al., 1993; Warren & Marslen-Wilson, 1987). In a gating study, Warren and Marslen-Wilson (1987) found that high frequency increases the probability that a word is produced as a response. In a visual word eye-tracking study, Dahan et al. (2001a) found that participants were more likely to look to cohort competitors with high frequency, e.g. *bed* than to low-frequency competitors, e.g *bell.* Even when none of the distractors were cohort competitors, there were more looks to words with higher frequencies. In a visual word eye-tracking study, Magnuson et al. (2007) uncovered effects of frequency, neighbourhood density and 'frequency-weighted cohort density', but the effects varied over time. Frequency weighted cohort density is the sum of the log-transformed frequencies of all words in the cohort, including the target word (Magnuson et al., 2007). Fixation proportions for high-frequency words and low cohort density were higher already at the beginning of the recognition process. Effects of neighbourhood density emerged only later (Magnuson et al., 2007). Dahan et al. (2001b) found that even subphonemic cues could guide lexical activation. Stimulus words were cross-spliced so that some had onsets from a cohort competitor word while other had onsets from a non-word. For instance, the word *net* occurred with [nɛ] from the word *neck* and *nep*, respectively, as well as in the original condition. In an eye-tracking study, listeners fixated more slowly on target pictures when onsets came from competitor words than non-words. The fastest fixations were for the original words. The authors interpreted the findings as evidence for 'lateral inhibition' as postulated by the connectionist models. Lateral inhibition means that activation of one word not only depends on whether it is itself compatible with the speech signal, but is also modulated by activation of competitor words. Finally, grammatical contexts have been found to constrain lexical competition (Strand et al., 2014). In a response time study, participants listened to words in constrained and unconstrained grammatical contexts. Stimulus words were responded to faster in the constrained contexts and the effect was greater for words with low competition from words in the same grammatical class.

Together, the findings indicate that 1) the acoustic signal is continuously mapped onto lexical candidates, 2) there are early effects of cohort competition and frequency 3) neighbourhood effects emerge later, 4) there is lateral inhibition between activated candidates and 5) grammatical constraints influence lexical competition.

2.3 Neural correlates of lexical competition

In neuroimaging studies, lexical competition correlates with activation in left inferior frontal gyrus (Broca's area) (Righi et al., 2010; Zhuang et al., 2011; Zhuang et al., 2014). As captured by many of the models of speech perception and spoken word recognition, there are both activation and selection aspects to lexical competition and these have been found to yield effects in somewhat different brain areas. In a combined fMRI and eye-tracking study, Righi et al. (2010) investigated the neural systems employed in competition and selection between two competitors. Participants were presented with four pictures, heard a stimulus word and were asked to look at the relevant picture. In half the trials, the target, for example *beaker*, was accompanied by one onset competitor, e.g. *beetle* and two distractors, for instance *cat* and *train*. In the other half of trials, there was only a target and three distractors. When a target words and a distractor competed for recognition, increased activation was measured in left inferior frontal gyrus, pars opercularis, indicating that the area is sensitive to lexical competition and selection driven by phonological factors. Further, activation in bilateral supramarginal gyrus was reported. Zhuang et al. (2014) isolated effects of cohort competition and lexical selection. In an fMRI study, participants listened to words and non-words and performed a lexical decision task. Some non-words had late 'nonword points', that is, the point at which a sequence becomes inconsistent with a real word. These non-words varied in initial cohort size and 'drop-out rate'. Drop-out rate refers to the number of words in the initial cohort dropping out before the non-word point. The measures were used to investigate effects of competition and selection, respectively. While words with high cohort sizes yielded increased activation in ventral inferior frontal gyrus (pars orbitalis and orbitofrontal cortex), words with higher drop-out rates yielded increased activation in dorsal inferior frontal gyrus (pars opercularis, pars triangularis and insula). 'Dorsal' means more towards the top of brain and 'inferior' refers to the part of the brain closer to the neck. There was activation in both left and right hemispheres although the effect appeared to be stronger in the left hemisphere, which is more associated with language processing. Zhuang et al. (2014) interpreted the findings as evidence that competition between initially activated candidates and selection among these are separate functions due to their separate neural substrates.

In ERP studies, reduced lexical competition yields more negative amplitudes around 200 ms after word onset (Hunter, 2013; Roll et al., 2017; Söderström et al., 2016). Roll et al. (2017) and Söderström et al. (2016) isolated a negative-going potential, termed pre-activation-negativity (PrAN), 136-204 ms after word onset and 136-280 ms after F_0 onset, respectively. The PrAN amplitude is more negative for word-beginnings with few lexical continuations, that is, few cohort competitors. Early, 136-204 ms after word onset, Roll et al. (2017) also saw effects of the combined frequency of these competitors, but no such effects were reported by Söderström et al. (2016). In a similar way, Hunter (2013) found that high neighbourhood density was associated with a positivity approximately 250 ms after word onset. This could be understood as negativity associated with few lexical competitors as well since ERP effects are (almost always) a comparison between two different conditions. Roll et al. (2017), Roll et al. (2015) and Söderström et al. (2016) interpreted the PrAN as a negativity for low lexical competition because the condition with few lexical competitors showed increased 'global root mean squares' (gRMS) peak 184-204 ms after word onset (Roll et al., 2017). gRMS reflects major changes in neural activity calculated over all electrodes. The finding is highly relevant because it suggests that accounts of spoken word recognition must explain the increased activation for *few* competitors as compared to many. The increased neural activity for few competitors at the gRMS peak correlated with blood oxygen level dependent (BOLD) activation mainly in inferior frontal gyrus, pars opercularis in Broca's area and left angular gyrus (Roll et al., 2017). The region has, as mentioned above, been found to be involved in lexical competition and, more specifically, selection between different candidates. Roll et al. (2017) suggest that the PrAN reflects the lexical selection stage of the Cohort model (Marslen-Wilson & Tyler, 1980; Marslen-Wilson, 1987), that is, the stage where lexical competitors are inhibited when they become incompatible with the incoming speech signal. A regression was found for a left central electrode, C3, 136-204 ms after word onset, where PrAN increased as the number of possible continuations decreased and their summed frequencies increased. Further, PrAN has been interpreted as an index of predictive certainty: the fewer possible candidates, the more the listener can commit to pre-activation of those candidates left (Roll et al., 2017). More negative PrAN amplitudes have also been reported for Swedish accent 1 as compared to accent 2 (Roll, 2015; Roll et al., 2010; Roll et al., 2017; Söderström et al., 2017a). Word accents are tones realised on word stems. They are distinctive but also, in many cases, morphologically conditioned (Riad, 2012). The finding is in line with accent 1 being associated with, on average, 11 times fewer continuations than accent 2 (Söderström et al., 2016). In an fMRI study, early gRMS peaks correlated with activity in left Heschl's gyrus in primary auditory cortex and adjacent superior temporal gyrus and inferior frontal gyrus.

In sum, lexical competition is associated with increased activation in Broca's area (left inferior frontal gyrus). Competition has been associated with activity in ventral parts while selection between candidates has been associated with activity in dorsal parts. In ERP studies, a left central negativity, correlating with increased activity in Broca's area, has been reported for *low* as compared to high lexical competition.

2.4 Entropy

Introduced in information theory in 1948 by Claude Shannon, entropy has in recent years also found use as a measure in language processing (Frank et al., 2015; Klimovich-Gray et al., 2019; Willems et al., 2016). Entropy is a measure of 'uncertainty' or 'choice' of the outcome of an event and is higher for more uncertainty (Shannon, 1948). Entropy can be calculated with the equation below where *p(x)* is the probability of an outcome and *log* is the base-2 logarithm. Thus, entropy can be calculated by taking the negative sum of the probability of all outcomes multiplied by the logarithm[1] of the probability of all outcomes.

[1] According to Shannon (1948), logarithmic measures are used because they are convenient mathematically and because many outcomes vary linearly with the logarithm of the outcomes. Since the logarithmic base is 2, the entropy is measured in bits, i.e. the number of bits required to store the outcome of the event.

$$Entropy = -\sum_{i=1}^{n} P(x_i) logP\ (x_i)$$

Entropy depends on the number of different outcomes as well as their probabilities. If probabilities of outcomes are equal, entropy also increases. Entropy also increases with the number of possible outcomes if they are equally probable. The highest entropy is found for events where *all* outcomes are possible and these events are equally likely (Shannon, 1948).

2.4.1 Entropy measures in language processing

Entropy measures have been used in studies of language processing concerned with predictive coding. As described in 2.1.6, according to theories of predictive processing, the brain tries to predict upcoming sensory information to avoid surprise and prediction error. Entropy has been used as a measure of the certainty of upcoming speech (Choi et al., 2021; Frank et al., 2015; Klimovich-Gray et al., 2019; Willems et al., 2016). Klimovich-Gray et al. (2019) investigated predictions generated by modifiers about following nouns in a combined magnetoencephalographic and EEG study using source localization. While EEG and ERP generally have excellent temporal resolution but poor spatial resolution, source localization can be used to estimate the likely sources of brain activity, combining the best of two worlds. Entropy measures were used to quantify certainty about upcoming nouns. Probability scores were obtained in a gating study where participants heard either 1) only the modifier or 2) the modifier and the first 50 ms of the noun. They were asked to guess what they thought the noun was and how confident they were about the guess. 'Entropy' in this study was calculated based on the different noun competitors given before noun onset as well as the confidence scores. If most participants gave one or two continuations with high certainty, entropy was low. If many continuations were provided, entropy was higher. In addition, 'entropy change' was calculated. This was the difference between entropy measured when only the modifier was available and that measured when the first 50 ms of the noun had been heard. The authors reported effects of entropy in left inferior frontal gyrus pars triangularis from 70 ms before to 165 ms after noun onset. There were effects of entropy change in left Heschl's gyrus in primary auditory cortex 140-180 ms after noun onset. The latter finding was interpreted as support for contextual constraints and phonological cues regulating activity in primary sensory areas such as primary auditory cortex, but that continuous sensory input is also necessary.

The findings were concerned with entropy from one word to the next whereas in the present study, entropy measured listeners' certainty about unfolding words. Entropy encompasses measures which have been found to be relevant in lexical competition such as the number of competitors and their frequencies (Goldinger et al., 1989; Magnuson et al., 2007; Roll et al., 2017; Zhuang et al., 2011). These measures are also employed in most measures of lexical competition such as neighbourhood density (Goldinger et al., 1989) and cohort competition (Zhuang et al., 2011). Entropy is related to posterior probabilities as described in Bayesian models such as Shortlist B (Norris & McQueen, 2008). Posterior probabilities are the likelihood of a word given prior probabilities and the likelihood of a speech sound if a specific word is heard. Entropy takes into account all posterior probabilities and is higher (less certainty) if many words have similar probabilities and lower (more certainty) if few words have high probabilities.

2.5 Danish

Danish is spoken by roughly 5.6 million speakers, most of them in Denmark (Lewis et al., 2015). It is an East Scandinavian language, belonging to the North Germanic branch of the Indo-European language tree (Pereltsvaig, 2017). Danish is known for its exceptionally rich vowel system with 16 contrastive vowels. In addition comes a two-way distinction in vowel length (Basbøll, 2005). Further, Danish has a suprasegmental creaky voice feature 'stød'. Only words with 'stødbasis' can have stød. Stødbasis requires either a long vowel or a short vowel followed by a sonorant consonant. Words which do not live up to these sonority constraints lack stødbasis (Fischer-Jørgensen, 1989). Subtle, phonetic differences are present in the speech signal already from word onset, but the 'stød proper' (Fischer-Jørgensen, 1989) or the 'phonological locus' (Basbøll, 2014) typically begins midway through a long vowel or in the sonorant consonant following a short vowel. Stød manifests itself as irregular vocal fold vibrations and a considerable intensity fall (Fischer-Jørgensen, 1989). Stød is distinctive and there are numerous minimal pairs with stød and non-stød (Grønnum & Basbøll, 2001). This is relevant for the present study because in experiment 1, prosodic information about stød and non-stød became available to listeners 100-200 ms into the stimulus words. When this happened, a number of lexical candidates immediately became inconsistent with the speech signal. Some examples of words with stød and non-stød in the experiment are [ˈhɛlˀd̥n̥] *helten* ‘the hero’ and [ˈsg̊æːb̥ə] *skabe* ‘closets’. Stød is marked with ‘ˀ’.

2.6 Present study

The present study used event-related potentials to investigate neural correlates of lexical competition and predictive certainty. The correlation between lexical competition and the ERP component pre-activation negativity (PrAN) was investigated. Further, a new measure, entropy, was employed as a measure of predictive certainty. The research question was:

R1) What are the neural correlates of lexical competition and predictive certainty?

The hypotheses were that PrAN amplitudes would correlate with lexical competition and certainty about the unfolding word. Negativity was expected to increase with less competition and more certainty, as reported by Roll et al. (2017).

H1) Low lexical competition is reflected in more negative PrAN amplitudes.

H2) High predictive certainty is reflected in more negative PrAN amplitudes.

Lexical competition is difficult to quantify because listeners have different mental lexica, that is, mental dictionaries of words in a language, including their phonological representations and semantics. As Roll et al. (submitted, p. 5) put it, “the most relevant lexical competitors are based on the mental dictionary of the listener”. Since getting access to the different mental dictionaries of the participants of the study was not possible, estimated lexical competition for all item words in the experiments was calculated from a Danish pronunciation lexicon and a frequency list. Lexical competition was operationalised as ‘possible continuations’ which refers to the number of words consistent with a word-initial fragment (WIF) of a word, that is, the first two phonemes of an item word. For instance, in the present study, the Danish WIF [hɛ] from the item word [ˈhɛld̥ə] *helte* ‘heroes’ had 1169 continuations. The measure ‘continuations’ was chosen because ERPs were analysed immediately after word onset when only the first 2-3 speech sounds had been heard.

Cohort competition is a more relevant measure at earlier stages while neighbourhood and rhyme effects become important later (Allopenna et al., 1998; Magnuson et al., 2007).

Predictive certainty was operationalised as entropy which was an estimate of listeners' certainty about the unfolding word upon hearing specific speech sounds. Entropy was calculated from a word-initial fragment's number of continuations and their respective frequencies. This measure was chosen because frequencies of cohort competitors have been found to influence lexical competition (Goldinger et al., 1989; Roll et al., 2017) and because PrAN has been understood as a measure of predictive certainty (Roll, 2015; Roll et al., 2015; Söderström et al., 2017a). As mentioned above, the PrAN for low lexical competition has been interpreted as reflecting lexical selection and inhibition of irrelevant candidates as more of the acoustic signal becomes available to listeners. It is, in other words, an index of predictive certainty, because the fewer possible outcomes there are, the more the listener can commit to one of them (Roll et al., 2017). Including both measures, continuations and entropy, might enable isolation of those effects. Based on H1 and H2, four test implications were derived. While hypotheses are abstract predictions, test implications are measurable effects expected to occur if the hypotheses were true (Hempel, 1966).

T1a) WIFs with few continuations yield more negative PrAN amplitudes 136-204 ms after word onset than WIFs with many continuations.

T1b) As the number of continuations decreases, negative PrAN amplitudes 136-204 ms after word onset increase.

T2a) WIFs with low entropy show more negative PrAN amplitudes 136-204 ms after word onset than WIFs with high entropy.

T2b) As entropy decreases, negative PrAN amplitudes 136-204 ms after word onset increase.

Data from two previously conducted ERP studies was reanalysed. The neural effects were measured 136-204 ms after word onset when only the first 2-3 speech sounds of a word had occurred in the speech signal.

Further, a new hypothesis was derived from Roll et al. (2017). The authors hypothesised that the negativity for WIFs with few continuations reflected the selection stage of the Cohort model. This would make the prediction that WIFs *inhibiting* more continuations activated based on initial acoustic input would yield more negative PrAN amplitudes. The hypotheses were:

H3) Speech sounds inhibiting many continuations yield more negative PrAN amplitudes.

H4) Speech sounds leading to large entropy decrease lead to more negative PrAN amplitudes. Using data from experiment 1, effects of 'inhibition' and 'entropy decrease' were investigated. Inhibition was defined as the number of continuations no longer consistent with the speech signal at a point of time. It was calculated by subtracting the number of words in the cohort at later stage from the number of words activated at a previous stage. Zhuang et al. (2014) used the term 'dropout rate', but this was calculated a bit differently and it indicates that competitors really do drop out as predicted by the Cohort model rather than just being inhibited. Models of speech recognition diverge on the matter. Entropy decrease was similar to the measure entropy change employed by Klimovich-Gray et al. (2019). In the present study, it was termed entropy decrease because there was always a decrease in entropy as words unfolded. Entropy decrease was calculated by subtracting the entropy for the second speech sound from the entropy of the first speech sound(s).

The effects of inhibition and entropy decrease were investigated at two time points in the speech signal. The first time point was the same as in the abovementioned analysis, 136-204 ms after *word onset*. Here, the difference between second speech sounds inhibiting many candidates first activated by the first speech sound was investigated. This can be exemplified with the word [hɛlˀd̥n̩] *helten* 'the hero' from experiment 1. The WIF [hɛ] had low inhibition because few competitors activated upon hearing [h] were inhibited upon hearing the [ɛ]. The second time point was 136-204 ms after the onset of stød or non-stød prosodic cues, on average 166 ms, *SD* = 47, after word onset. Here, it was investigated what happened when competitors activated by the first two speech sounds were no longer consistent with the speech signal. The late WIF [hɛlˀ] had a high inhibition as many candidates consistent with [hɛ] were inhibited upon hearing the stød realized in the sonorant consonant, [lˀ]. On the contrary, *helten*'s non-stød counterpart [hɛld̥ə] *helte* 'heroes' had low inhibition because few words were inhibited upon hearing non-stød in the sonorant consonant [l]. Test implications were derived from hypotheses 2 and 3.

T3a) WIFs inhibiting many continuations yield more negative PrAN amplitudes 136-204 ms after word onset than WIFs inhibiting few continuations.

T3b) As the number of inhibited continuations increases, negative PrAN amplitudes 136-204 ms after word onset increase.

T4a) WIFs with large entropy decrease show more negative PrAN amplitudes 136-204 ms after word onset than WIFs with small entropy decrease.

T4b) As entropy decrease surges, negative PrAN amplitudes 136-204 ms after word onset increase.

3 Method and materials

3.1 Event-related potentials

Electroencephalographic (EEG) data collected from two previous experiments was reanalysed. EEG data is electrical activity measured above the scalp. The activity stems from postsynaptic potentials occurring when neurons communicate with each other (Luck, 2014). In the present study, participants wore a cap with 32 electrodes distributed over the head according to the 10-20 system (Klem et al., 1999), each measuring voltage fluctuations at their respective positions. Event-related potentials are electric potentials associated with specific ‘events’. In the present study, one such event could be the onset of a stimulus word. ERP components are event-related potentials elicited by specific cognitive processes. Of interest for the present study is the pre-activation negativity (PrAN) described in 2.3. EEG has excellent temporal resolution and is therefore particularly well-suited for investigating early effects of lexical competition. The spatial resolution is quite poor, however, because activity measured at the scalp could, at least in theory, be the result of a virtually infinite number of brain source combinations.

In the first experiment (Hjortdal, 2021), participants listened to 40 Danish nouns with stødbasis, each occurring in singular and plural conditions. In the second experiment, the same participants listened to another 40 Danish nouns without stødbasis, also in singular and plural conditions. While Hjortdal (2021) was concerned with ERPs associated with stød or non-stød prosodic cues cuing grammatical suffixes, the present study investigated ERPs associated with lexical competition.

3.1.1 Participants

Sixteen native speakers of Danish, mean age 27.6 ± 4.9 years, participated in the study. All were right-handed. All participants spoke English and, in some cases, other languages, in addition to Danish. The Swedish Ethical Review Authority (https://etikprovningsmyndigheten.se/, approval number 2020-03035) approved the study. Participants gave informed consent and were offered remuneration for participation. All participants took part in both experiments.

3.1.2 Stimuli

The stimulus words in experiment 1 were 40 nouns with stødbasis. In experiment 2, the stimulus words were 40 nouns without stødbasis. Each item occurred in definite singular and indefinite plural conditions. Both definite singular and indefinite plural are marked as a suffix on Danish nouns. Definite singular was marked [n̩]/[ð̩] *en/et* and indefinite plural [ə] *-e.* Since consonants can be syllabic in Danish (marked with '̩' as in [n̩]), both conditions were disyllabic. Table 1 shows some stimulus words. All stimulus words can be found in the appendix A.

Table 1. Each stimulus words occurred in different singular and plural conditions in which they were disyllabic.

Experiment 1 (stødbasis)		**Experiment 2 (non-stødbasis)**	
Helt 'hero'	['hɛlˀd̥n̩] *helten* ['hɛld̥ə] *helte*	*Dusk* 'tuft'	['d̥usg̊n̩] *dusken* ['d̥usg̊ə] *duske*
Skab 'closet'	['sg̊æˀb̥ð̩] *skabet* ['sg̊æ:b̥ə] *skabe*	*Flok* 'flock'	['flʌg̊n̩] *flokken* ['flʌg̊ə] *flokke*
Væg 'wall'	['vɛˀg̊n̩] *væggen* ['vɛ:g̊ə] *vægge*	*Krop* 'body'	['kʰʁʌb̥n̩] *kroppen* ['kʰʁʌb̥ə] *kroppe*

Stimulus words were incorporated into carrier sentences with the structure *Ruth fandt* ***stimulus word*** *på pladsen/på græsset/på loftet* 'Ruth found **stimulus word** at the place/on the grass/on the loft'. Thus, all stimulus words were items and the sentence context was the same. The stimulus sentences were recorded by a female speaker of Standard Copenhagen Danish in an anechoic chamber. Recordings were made with the recording and editing software Audacity® (https://audacityteam.org). Each stimulus sentence was recorded twice, once with the stimulus noun in singular and once in plural. To avoid pronunciation being systematically influenced by intonation differences, half the stimulus sentences were read with the singular condition before the plural. For the other half, the plural was read before the singular. All carrier sentences were read as answers to context questions to avoid focus on stimulus nouns because focus has a lengthening effect in Standard Copenhagen Danish (Grønnum & Basbøll, 2001).

For the purpose of the original experiment, words were cross-spliced in Praat (Boersma & Weenink, 2020) to create conditions with suffixes validly and invalidly cued by word stems. This meant that some stimulus words occurred with suffixes which had been cross-spliced onto them from another condition, (e.g. singular word stems with plural suffixes). Since suffix onset was not until, on average, 282 ms, *SD* = 46 ms, after word onset in experiment 1 and 264 ms, *SD* = 66 ms, in experiment 2, this could not have affected the word onset ERPs 136-204 investigated in the present study. For the effects investigated 136-204 ms after stød/non-stød onset, the window overlapped with suffix onset. However, the suffix-splicing happened after vowel and stød/non-stød onset and therefore did not affect the contrast investigated here. Further, all items occurred in the same number of valid and invalid conditions so these effects should not affect comparison between stimulus words differing with respect to lexical competition. A possibility would have been to exclude spliced items, but since there is a lot of 'noise' in ERP studies, stemming from skin potentials and random fluctuations (Kretzschmar & Alday, forthcoming; Luck, 2014), it is crucial to maximise the number of trials. Stimulus nouns were spliced back into the carrier sentences.

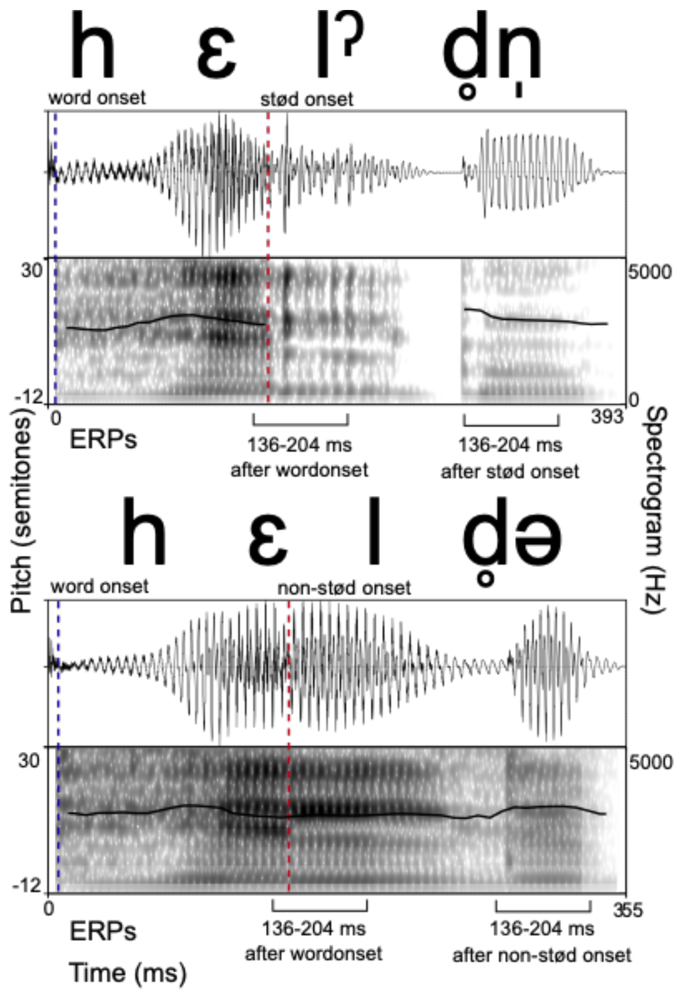

Figure 1. The item *helt* 'hero' in its singular (top) and plural (bottom) conditions. The time windows 136-204 ms after word onset and 136-204 ms after stød/non-stød onset are marked.

3.1.3 Procedure

Both experiments were conducted in the experimental software programme Psychopy and run on a stationary PC (Peirce et al., 2019). Experiment 1 was always run before experiment 2. Participants sat in a chair in front of the computer screen, looking at a fixation cross and wearing 32-channel Braincap-MR from EasyCap. They listened to stimulus sentences via headphones. Stimulus sentences were presented in randomised order with stimulus onset asynchrony jittered between 4 and 6 seconds to prevent artificial boosting of time-locking of ERPs to the stimuli (Luck, 2014). In experiments 1 and 2, each stimulus item was presented 8 and 4 times, respectively, in different conditions, to each participant. Altogether, there were 480 trials, 320 in experiment 1 and 160 in experiment 2. Participants were instructed to respond as fast as possible to whether the stimulus word represented one or more things by pressing a button on the keyboard. Since the behavioural task was related to identification of morphological suffixes rather than lexical competition, these responses were not analysed in the present study.

There were 'time-locking points' for word onset and stød/non-stød onset. Time-locking points mark specific events (e.g. word onset) and link stimuli and enable extracting and averaging ERPs associated with specific events. Word onset was operationalised as the onset of the first speech sound in a stimulus word. Stød onset was operationalised as the point where vibrations started getting irregular and non-stød onset was the corresponding time after F_0 onset in words without stødbasis. See Figure 1.

3.1.4 EEG recording and pre-processing

EEG recordings were made at a sampling rate of 500 Hz using a BRAINAMP MR PLUS Amplifier and Brainvision recorder (BrainProducts). Impedances were kept below 5kΩ. Pre-processing was done in Eeglab (Delorme & Makeig, 2004). A centro-frontal electrode (FCz) was used as online reference and re-referenced offline to mastoid average (electrodes TP9 and TP10). EEG was low-pass offline filtered at 30 Hz and high-pass filtered online at 0.05 Hz. Eye movements, including blinks and saccades, were compensated for using independent component analysis (ICA) (Jung et al., 2000).

Item and single-trial epochs ending 800 ms after word onset/stød/non-stød onset and with a 200 ms pre-stimulus baseline were extracted for the relevant electrodes. Epochs with voltage exceeding ±100 V after ICA were discarded. There were 5,120 trials in the stødbasis experiment of which 47 (0.92 %) were discarded and 2,560 trials in the non-stødbasis experiment of which 34 (1.3 %) were discarded.

Microvolt (μV) item averages for the two (singular and plural) conditions of each test item were calculated in Eeglab (Delorme & Makeig, 2004). Further, single-trial microvolt (μV) averages were calculated. Thus, rather than averaging EEG data over conditions, one trial per condition, per item, per participant was extracted. Missing data point were coded as N/A's.

3.2 Lexical competition

To quantify the lexical competition of the word-initial fragments (WIFs) of the items in the experiment, the number of continuations and the entropy for each WIF was calculated using Unix and Python scripts. The scripts were based on scripts developed for Swedish for the same purpose (Roll et al., 2017; Söderström et al., 2016), but adapted for Danish phonetics by the author. Additional code was written to extract the new measure entropy. Scripts can be found in appendix B.

3.2.1 Pronunciation lexicon

Pronunciations were obtained from the Danish NST lexicon, a freely available full-form pronunciation lexicon for Danish (Andersen, 2011). Developed by Nordisk Språkteknologi Holding, the lexicon was made available to the public by the Norwegian National Library after 2003 when Nordisk Språkteknologi went bankrupt. In the Danish NST, all words in the lexicon have been manually transcribed with the Speech Assessment Methods Phonetic Alphabet (SAMPA). The SAMPA resembles the International Phonetic alphabet (IPA), but has been adapted to be computer-readable (Wells, 1997). In the lexicon, there is a distinction between 26 vowels and 19 consonants. Stød, stress, secondary stress and word boundaries are also marked in the lexicon (Bjerg Nielsen, 2001). There are 237,873 items in the lexicon, 236,307 of which have been part-of-speech (PoS)-tagged for word class by Center for Sprogteknologi in Copenhagen. Nouns constitute 52 % of the lexicon, proper nouns 24 %, adjectives 12 %, verbs 10 %, adverbs 1.9 % and other grammatical categories 0.4 % (Andersen, 2011). In the present study, proper nouns were excluded. The NST also has lexica for Norwegian and Swedish. The Swedish lexicon was used for the Söderström et al. (2016) and Roll et al. (2017) studies. The Danish lexicon has fewer word forms than the Swedish lexicon, but it is still very extensive and well documented. Contrary to the Swedish and Norwegian lexica, all items in the Danish lexicon have been manually checked. A weakness is that it has not been updated since 2003 and some neologisms are therefore absent.

3.2.2 Frequency list

A frequency list was generated based on KorpusDK[2] (https://korpus.dsl.dk). KorpusDK comprises three corpora, Korpus 90, Korpus 2000 and Korpus 2010, with Danish texts from diaries, blogs, newspapers, fiction etc. from around 1990, 2000 and 2010, respectively. In total, the corpus contains more than 100 million words and is ePOS tagged. EPOS is an extended version of the Part-of-speech (PoS) tagset which includes inflectional information (Asmussen, 2015). In the ePOS tagset, there is a distinction between verbs, adjectives, numerals, nouns, pronouns, adverbs, interjections, prepositions, conjunction, lexical element, inflectional ending and 'unique', including inter alia infinitive markers, and 'residual', which includes tagging errors or foreign words.

KorpusDK is a written language corpus. This is of course not ideal when the object of investigation is spoken language. A PoS-tagged spoken language corpus, Danpass (Grønnum, 2009), was considered as an alternative, but it only contains 73,227 running words – as compared to 100 million in KorpusDK – which is too little for the present purposes. First, word frequencies calculated based on smaller corpora are less accurate. Second, since many WIFs from the experiments would have few and low-frequency continuations, fine-grained differences would be lost. Henrichsen (2002) compared Danish written and spoken language corpora. The most notable

[2] Language resource compiled by Jørg Asmussen, Society for Danish Language and Literature, DSL.

difference was that common words are even more common in speech than in writing. The fifty most common types make up 60.4 % of the spoken corpus and 40.1 % of the written corpus. This would indicate that items generally have fewer continuations and lower entropy in speech than in writing, but this is not expected to vary systematically between items and was deemed acceptable. However, it should be kept in mind when interpreting the results. Another option considered for obtaining word frequencies was the PAROLE corpus[3] whose Swedish sister was used to obtain word frequencies for the Swedish script (Söderström et al., 2016). The PAROLE project aims to compile large text corpora for all EU languages. However, the Danish PAROLE only includes 250.000 morphosyntactically tagged words which was also deemed too few to obtain the fine-grained competition measures desired.

All words, including ePOS word-class tags, were extracted from KorpusDK using a Python script and frequencies for each item were calculated with a Unix script run in the terminal. Word class was taken into account to avoid collapsing frequencies across word classes. Thus, words with the same orthography but belonging to different grammatical categories – and possibly with different pronunciations - such as the noun [‘tæːlʌ] *taler* ‘speaker’ and the verb [‘tæˀlʌ] *taler* ‘speaks’ were distinguished. A frequency list with 1,485,541 unique items was generated.

3.2.3 Combining pronunciation and frequency data

To be able to access both a word’s pronunciation and its frequency, the NST lexicon and the frequency list were merged. First, orthography, PoS tags and pronunciation data were extracted from the NST lexicon with a Unix script (Frid, 2015). In Python scripts written by the author, the pronunciation lexicon was pre-processed and merged with the frequency list. Since the pronunciation lexicon and KorpusDK were tagged with two different word class tagging styles, PoS and ePOS, respectively, the PoS-tags in the NST were replaced with ePOS word class tags. Further, all orthographies were lowercased. Thirty-five stimulus words used in the experiments but not in the NST lexicon were added manually. Then, the pronunciation lexicon was merged

[3] *PAROLE-DK and ePAROLE*. Compiled by Ole Norling-Christensen, Britt-Katrin Keson, Jørg Asmussen m.fl., Society for Danish Language and Literature, DSL.

with the frequency list, creating a file with frequency, orthography, word class tag and pronunciation. In total, 140,283 items, including information about pronunciation and word class, remained. The number of items in the merged list was substantially lower than the number of items in the pronunciation lexicon and frequency list, respectively, since only items present in both were included in the merged list.

3.2.4 Calculating lexical competition and entropy for word-initial fragments

A Python script was developed for calculating continuations and entropy. The script was based on a script developed for Swedish (Söderström et al., 2016) but adapted for Danish phonetics. Word items in the lexicon were marked for stød (with a '?'), but its counterparts, 'non-stød' and 'non-stødbasis', were not marked in the lexicon. Non-stød refers to words with stødbasis (a long vowel or a short vowel followed by a sonorant consonant) but no stød. Non-stødbasis refers to words without stødbasis which can never have stød due to sonority constraints. To tag word items for non-stød and non-stødbasis, respectively, the script searched them for stressed syllables which did not have stød (marked '?') but with either 1) a long vowel or 2) a short vowel followed by a sonorant consonant. Those words had non-stød. Sonorant consonants in Danish are [m], [n], [ŋ], [l], [w], [j] and [ð] (Basbøll, 2005). Further, in Danish, /r/ has been vocalised and is realised [ɐ̯], except in onset position. Words with a short vowel followed by /r/ thus have acquired stødbasis and are in an ongoing process of acquiring stød (Grønnum, 2005; Høeg, 2020) and [ɐ̯] was therefore also tagged as a sonorant consonant. Further, [ɐ̯] can fuse with a preceding vowel (Basbøll, 2005) as in [jɒːd̥] *hjort* 'deer', but this is already captured because the vowel is long. All remaining words were categorised as not having 'non-stødbasis', meaning they could not have stød.

To calculate a WIF's number of continuations and entropy, the Python script looped through the combined pronunciation and frequency list, counting all words with the WIF in question in word-initial position and – for entropy – their respective frequencies. Figure 2 shows some of the continuations of the WIF [hɛ] (sampa [hE]), including details about, for example, frequency and stød.

To calculate a WIF's entropy, the probability, (*p*-value) for each continuation was calculated by dividing its frequency by the frequency of all words in the cohort. The entropy itself was calculated by taking the negative sum of all p-values multiplied by the binary logarithm of all p-values. If, at a certain point, there is evidence of, for instance, two words with similar frequencies and thus prior probabilities, both would have lower posterior probabilities and produce higher entropy – until more of the speech signal is available and the inconsistent candidate drops out.

```
Enter a word-initial fragment: h_E
hensyn h_E-_n N INITIAL 15694 STRESS NONSTODBASIS
hænder h_E-_n_? N INITIAL 6742 STRESS STOD
hænderne h_E-_n_? N INITIAL 5215 STRESS STOD
helle h_E- N INITIAL 4974 STRESS NONSTODBASIS
henblik h_E-_n N INITIAL 3638 STRESS NONSTODBASIS
herre h_E+ N INITIAL 3587 STRESS NONSTOD
heste h_E- N INITIAL 2923 STRESS NONSTODBASIS
henvisning h_E-_n N INITIAL 2749 STRESS NONSTODBASIS
herren h_E+ N INITIAL 2529 STRESS NONSTOD
helvede h_E-_l N INITIAL 2342 STRESS NONSTODBASIS
hemmelighed h_E- N INITIAL 2331 STRESS NONSTODBASIS
hæren h_E+_? N INITIAL 2037 STRESS STOD
henvendelse h_E-_n N INITIAL 1954 STRESS NONSTODBASIS
henvendelser h_E-_n N INITIAL 1823 STRESS NONSTODBASIS
hensigten h_E-_n N INITIAL 1774 STRESS NONSTODBASIS
helbred h_E-_l N INITIAL 1651 STRESS NONSTODBASIS
hesten h_E- N INITIAL 1639 STRESS NONSTODBASIS
hensigt h_E-_n N INITIAL 1631 STRESS NONSTODBASIS
herrer h_E+ N INITIAL 1608 STRESS NONSTOD
henhold h_E-_n N INITIAL 1575 STRESS NONSTODBASIS
hensynet h_E-_n N INITIAL 1465 STRESS NONSTODBASIS
helte h_E-_l N INITIAL 1169 STRESS NONSTODBASIS
hærværk h_E-_6- N INITIAL 1041 STRESS NONSTODBASIS
herrens h_E+ N INITIAL 945 STRESS NONSTOD
hændelser h_E- N INITIAL 918 STRESS NONSTODBASIS
henseende h_E-_n N INITIAL 889 UNSTRESSED NONSTODBASIS
hektar h_E-_g N INITIAL 869 UNSTRESSED NONSTODBASIS
hælene h_E+ N INITIAL 798 STRESS NONSTOD
hærens h_E+_? N INITIAL 786 STRESS STOD
hestene h_E- N INITIAL 755 STRESS NONSTODBASIS
hemmeligheder h_E- N INITIAL 749 STRESS NONSTODBASIS
helikopter h_E- N INITIAL 718 UNSTRESSED NONSTODBASIS
hensigter h_E-_n N INITIAL 637 STRESS NONSTODBASIS
helten h_E-_l_? N INITIAL 591 STRESS STOD
hensyntagen h_E-_n N INITIAL 557 STRESS NONSTODBASIS
helikoptere h_E- N INITIAL 556 UNSTRESSED NONSTODBASIS
hertug h_E-_6- N INITIAL 530 STRESS NONSTODBASIS
herkomst h_E-_6- N INITIAL 495 STRESS NONSTODBASIS
hænde h_E- N INITIAL 468 STRESS NONSTODBASIS
henvisninger h_E-_n N INITIAL 461 STRESS NONSTODBASIS
```

Figure 2. The script looped through the combined pronunciation and frequency lexicon for each WIF, calculating the number of continuations and entropy.

For the present study, some constraints were applied: Only polysyllabic nouns were included as this was the category used in the ERP experiment because grammatical category has been shown to limit search space in word recognition (Strand et al., 2014). Also, words with frequencies lower than 2 were excluded because in a frequency list calculated based on a 100-million-word corpus, these words would be very rare. The same constraints were applied in Roll et al. (2017), facilitating comparison. Table 2 shows some word-initial fragments from both experiments as well as their number of continuations and entropies. For experiment 2, one item's two realisations, [ɔsd̥n̩] *osten* 'the cheese' and [ɔsd̥ə] *oste* 'cheeses', were removed because the number of continuations and entropy could not be calculated appropriately because there was only one speech sound in the first syllable, [ɔ]. A list with all items, their WIFs at different time points as well as their pronunciations and frequencies can be found in appendix C. The number of continuations was log-transformed for further analysis to approximate a normal distribution, because they were not normally distributed. To decide whether to log-transform entropies, it was investigated whether entropies were normally distributed. Entropies for all phonotactically legal WIFs were calculated and a Jarque-Bera test was carried out in Excel. A Jarque-Bera test examines whether the skewness and kurtosis of the data is consistent with a normal distribution, investigating the null-hypothesis that data *is* normally distributed. A p-value of 0.707 confirmed the null-hypothesis, indicating that entropy data was normally distributed. Therefore, entropies were not log-transformed.

Table 2. For each word-initial fragment (WIF) (i.e. the first two speech sounds of a word), the number of continuations and entropy was calculated. See appendix C for a full list of WIFs, continuations and entropies.

Experiment 1			**Experiment 2**		
WIF	**Continuations**	**Entropy**	**WIF**	**Continuations**	**Entropy**
[vɛː]	79 (few)	3.80 (low)	[du]	47 (few)	3.97 (low)
[sv]	393 (few)	6.71 (high)	[tˢʌ]	337 (few)	6.27 (high)
[kʰɑ]	556 (many)	5.88 (low)	[kʰʁ]	881 (many)	5.13 (low)
[hɛ]	474 (many)	6.05 (high)	[sd̥]	2,535 (many)	8.07 (high)

Using a median split, WIFs were divided into few/low and many/high groups for number of continuations and entropy, respectively. Group means and ranges are reported in Table 3.

Table 3. Word-initial fragments were divided into groups based on whether they had few or many continuations and high and low entropy.

Experiment 1	**Group**	**Average**	**Range**
Continuations	Few	227	38-407
	Many	1,373	438-2535
Entropy	Low	5.17	3.59-5.92
	High	7.35	6.05-8.58
Experiment 2	**Group**	**Average**	**Range**
Continuations	Few	289	47-474
	Many	1,200	479-2535
Entropy	Low	5.34	3.97-6.08
	High	7.16	6.13-8.58

The number of WIFs inhibited by the second speech sound and by stød/non-stød onset was calculated. Inhibition at the first time point was the number of continuations activated by the first speech sound and inhibited by the second speech sound. It was calculated by subtracting the number of continuations at the second time point from the number of continuations at the first time point. Entropy decrease was the decrease in entropy from the first to the second speech sound. It was calculated by subtracting lexical entropy for speech sounds from lexical entropy for one speech sound. Inhibition and entropy decrease were calculated again as more of the speech signal became available to listeners at stød or non-stød onset, that is, on average 166 ms, *SD* = 47 ms, after word onset. Inhibition measures were log-transformed for further analysis. See appendix C for a list of word beginnings, including inhibition and entropy decrease. Using a median split, WIFs were divided into low and high groups for inhibition and entropy decrease, respectively.

3.3 Statistical analysis

3.3.1 Analysis of variance

To investigate whether there were significant differences between the few/low and many/high groups, one-way analyses of variance (ANOVAs) were carried out. The dependent variable was PrAN amplitude and the independent variables (tested in separate models) were continuations, entropy, inhibition and entropy decrease. Each independent variable had two levels: few/low and many/high (see 3.2.4). ANOVAs can test whether there is difference between groups while simultaneously taking variance into account. However, it is not informative about the nature of the effect, that is, *which* group yields more negative PrAN amplitudes. ANOVAs have the null-hypothesis that there is no difference between groups and the p-value tells us the probability that the null-hypothesis is true. The confidence level is 100 minus the significance level. Thus, if the p-value is 0.05, the confidence level is 0.95 or 95 % (Rasinger, 2013).

3.3.2 Regression analysis

Regression analyses were employed to test whether item average PrAN amplitudes varied continuously with the independent variables. From the correlation between two variables, a regression equation can be calculated on the form $y = \mathrm{a} + \mathrm{b}x$. Here, a is the intercept, which is the point where the regression line crosses the y-axis and b is the slope. The slope indicates how much y, for example, PrAN amplitude, increases per x unit increase, for example entropy. R^2 reflects how much of the variability in variable y can be explained by variable x. Thus, an R^2 value of 0.40 can explain 40 % of the variability whereas the remaining 60 % is due to other factors. The p-value again reflects the confidence level (Rasinger, 2013). The one-way ANOVA and regression analyses were applied in Roll et al. (2017) and Söderström et al. (2016). These analyses thus also facilitated comparison with these studies.

3.3.3 Linear mixed-effects models

Further, the single-trial ERP averages were tested in linear mixed-effects regression models. Linear mixed-effects models have been argued to be superior to by-subject or by-item (as the ones above) analyses because they simultaneously model variance associated with subjects (i.e. participants) and items (Barr et al., 2013; Kretzschmar & Alday, forthcoming). By-subject analyses have been criticised for failing to take into account the variance associated with the different items in a study, making the assumption that participants react the same to all items. It has been argued that in such studies no inferences can be drawn beyond the actual test items used in the particular study. On the other hand, by-item analyses, as the ones employed in the present study, do not take subject variance into account (Barr et al., 2013; Kretzschmar & Alday, forthcoming). In linear mixed-effects models, the problem is tackled by including fixed effects as well as random effects. Random effects are typically added for 'grouping variables' such as subject and item. For each grouping variable, random intercepts and slopes are added. They reflect the variances in fixed effects, e.g. variance between subjects. Random intercepts are how the average of a dependent variable varies with the grouping variable (e.g. does one participant overall show more negative PrAN amplitudes?) while random slopes reflect how the effect of an independent variable varies with the random effects (e.g. does a participant react more strongly to e.g. low entropy?) (Kretzschmar & Alday, forthcoming). By-item and by-subject random intercepts and random slopes for the fixed effects investigated were added.

For the present study, the effects of continuations, entropy, inhibition rate and entropy were investigated in separate models. In the models, nuisance variables were added to account for confounds stemming from other factors varying between stimuli (Sassenhagen & Alday, 2016). They were 'word grammatical number' (if the word was singular or plural), 'sentence grammatical number' (if the carrier sentence had been recorded with a word in singular or plural) and – for inhibition rate and entropy change sparked by stød or non-stød cues – 'stød level' (whether the word had stød or non-stød). All factors were balanced within items, meaning items occurred in the same number of, for example, singular and plural conditions. The only exception was sentence grammatical number which was balanced between items and could therefore have affected ERPs already before word onset.

It should be noted that all items were preceded by the exact same words *Ruth fandt* 'Ruth found' but phonetic cues during these words varied between items. Listeners have been found to be sensitive to even subtle phonetic cues (Archibald & Joanisse, 2011; Dahan et al., 2001b; Salverda et al., 2014). Controlling for such effects is therefore important. The nuisance variables each had two levels and were deviation-coded. This means that the dependent variable mean for each level of a variable is compared to the grand mean rather than one factorial level (Singmann & Kellen, 2019). If models did not converge, random correlations were removed (Barr et al., 2013). Effects of nuisance variables are not reported in the thesis, but can be found in the output tables in appendix D.

To sum up, ANOVAs and linear regressions were chosen for comparison with Roll et al. (2017) and Söderström et al. (2016) whereas linear mixed-effects regressions were carried out to better control for nuisance variables, items and participants. Thus, the linear mixed-effects regressions were more conservative. All analyses were carried out in the R software (R Core R Core Team, 2017). For the linear mixed-effects model, the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages were used. The two experiments were analysed separately. For the replication parts, ERP effects were tested in a priori spatiotemporal window based on a previous study, that is, at electrode C3 136-204 ms after word onset (Roll et al., 2017). A spatiotemporal window refers to where above the head – and in which time window – the ERPs are extracted. A priori means that the window is defined in advance, before seeing the data. For the more exploratory analyses, effects were tested separately at six different centro-frontal electrodes (FC1, FC2, Fz, C3, Cz and C4) 136-204 ms after word onset and stød/non-stød onset, respectively.

4 Results and discussion

4.1 Experiment 1

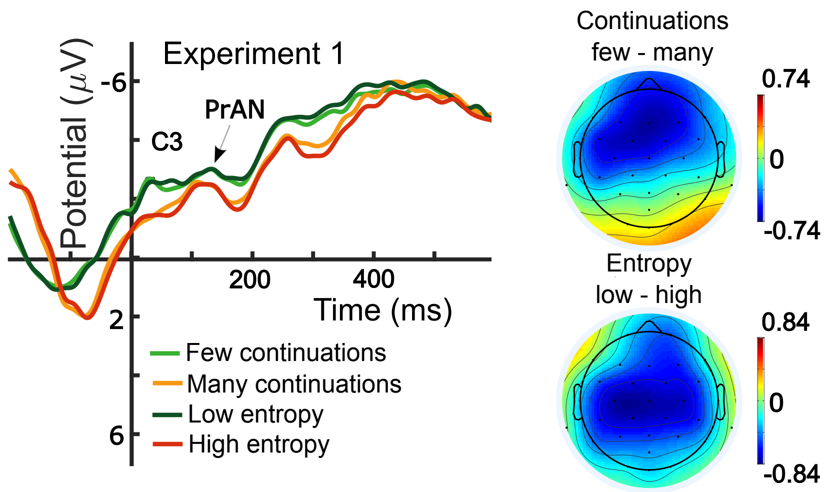

Figure 3. Word beginnings with few lexical continuations and low entropy yielded more negative amplitudes. The plots to the left are for the left central electrode, C3. Negativity is plotted upwards. The plots to the right show the voltage differences 136-204 ms after word onset.

Overall, word-initial fragments (WIFs) with low lexical competition resulted in more negative PrAN amplitudes than WIFs with high lexical competition in the a priori spatiotemporal window. An ANOVA showed that there was a significant difference between WIFs with few and many competitors, $F(1,78) = 5.75$, $p = 0.019$ over the left-central a priori electrode, C3, 136-204 ms after word onset. Further, a significant regression function was found, $F(1,78) = 13.28$, $p < 0.001$, $r = 0.38$, $r^2 = 0.15$, showing that the PrAN amplitude varied as a function of the number of continuations with fewer continuations yielding more negative amplitudes. See Figure 4.

1) $PrAN = 0.7(log\ continuations) - 4.94$

The plots in Figure 3 indicate that there could have been a difference in the signal already before word onset (y-axis intercept) which may have been caused by subtle, phonetic cues during the preceding sentence context (*Ruth fandt* 'Ruth found') which could not be controlled completely in the present study. However, when the effects were tested in a linear mixed-effects regression model, which took such context effects into account, the effect was still significant, $\beta = 0.91$, $SE = 0.31$, $df = 36.97$, $t = 2.96$, $p = 0.005$. As the difference plot in Figure 3 shows, the effect was strongest at left frontal sites.

Similarly, WIFs with low entropy yielded more negative PrAN amplitudes than WIFs with high entropy in the a priori spatiotemporal window. An ANOVA showed a significant difference between the low and high groups in the a priori spatiotemporal window, $F(1,78) = 11.52$, $p = 0.001$ and a significant regression function was found, $F(1,78) = 12.7$, $p < 0.001$, $r = 0.37$, $r^2 = 0.14$. See Figure 4.

2) $PrAN = 0.31(entropy) - 4.35$

A linear mixed-effects model confirmed that the PrAN amplitude varied as a function of entropy, $\beta = 0.31$, $SE = 0.10$, $df = 34.08$, $t = 3.00$, $p = 0.005$. As can be seen from the difference plot in Figure 3, the effect was stronger over central sites and somewhat lateralised.

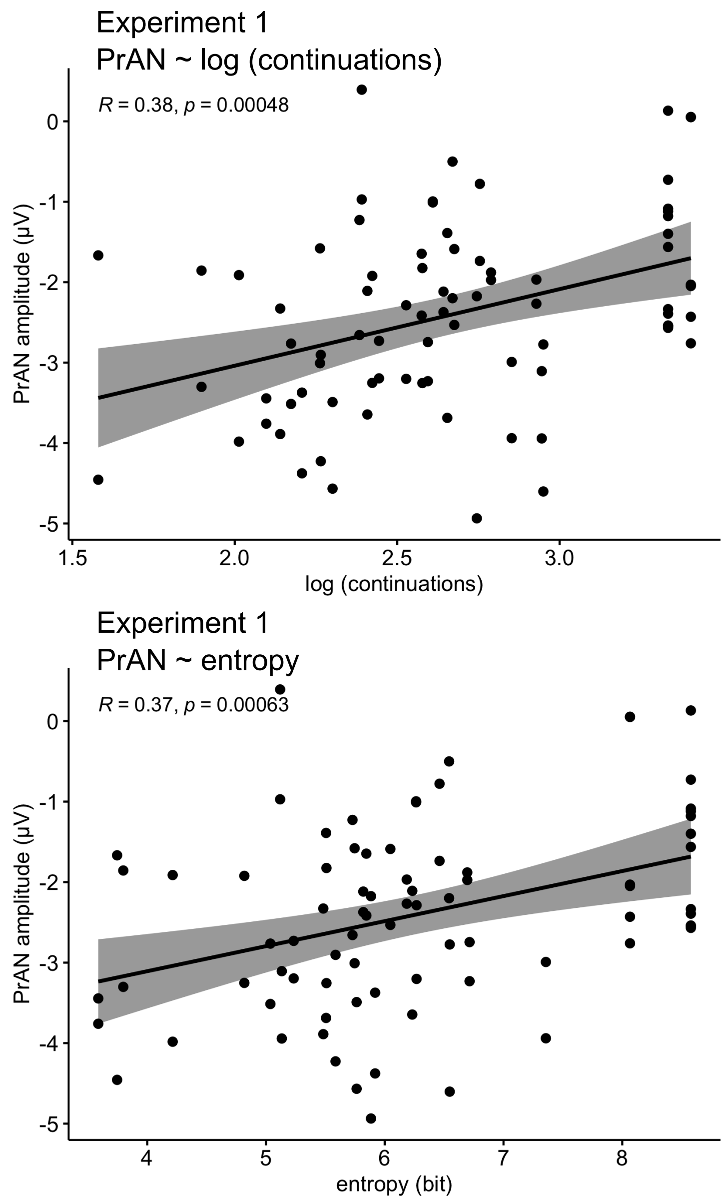

Figure 4. Regression equations were found for continuations as well as entropy over a left central electrode, C3.

Test implications 1a and 1b, that WIFs with few continuations would yield more negative PrAN amplitudes than WIFs with many continuations and that PrAN amplitudes would vary continuously with the number of continuations, were confirmed. The experiment replicated the findings of Roll et al. (2017) in that negative PrAN amplitude increased as the number of continuations decreased. Further, test implications 2a and 2b, that amplitudes would vary with entropy, were also confirmed.

4.2 Experiment 2

WIFs with few continuations and low entropy yielded more negative PrAN amplitudes than WIFs with many continuations and high entropy in the a priori spatiotemporal time window. An ANOVA showed that the difference for number of continuations was significant over C3 136-204 ms after word onset, $F(1,76) = 5.82$, $p = 0.018$, but the regression function was not. Similarly, for entropy, there was a significant difference between WIFs with low and high entropy, $F(1,76) = 5.09$, $p = 0.027$, but the regression was not significant. As shown in Figure 5, the effects were somewhat more central than those observed in experiment 1. Therefore, effects were also tested over a central electrode, Cz, where there were also significant differences between both continuations (few/many), $F(1,78) = 5.90$, $p = 0.018$ and entropy (low/high), $F(1,78) = 5.14$, $p = 0.026$, but no significant regressions. Neither were there any significant effects when the data was analysed in linear mixed-effects models.

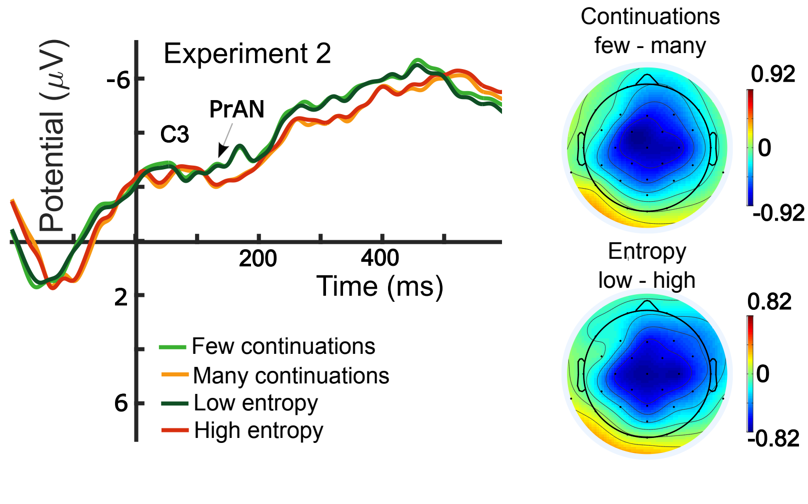

Figure 5. Word-initial fragments (WIFs) with few continuations or low entropy yielded more negative PrAN amplitudes than WIFs with many continuations or high entropy.

Thus, for experiment 2, test implications 1a and 2a were confirmed but test implications 1b and 2b were not. Some conditions differed between experiment 1 and 2. Experiment 2 was always run after experiment 1 with the same participants and there might have been a training effect. Further, there was less variance in the WIFs' number of continuations in experiment 2. While the few and many group means for experiment 1 were 227 and 1,373, respectively, they were 289 and 1,201 in experiment 2. Finally, there were only half as many trials in experiment 2 compared to experiment 1, meaning the sample size was smaller, which might explain why the continuous effect did not reach statistical significance.

4.3 Exploratory approach: experiment 1

Using data from experiment 1, a new hypothesis was investigated. It was hypothesized that WIFs which inhibited many continuations or had high entropy decrease would yield more negative PrAN amplitudes.

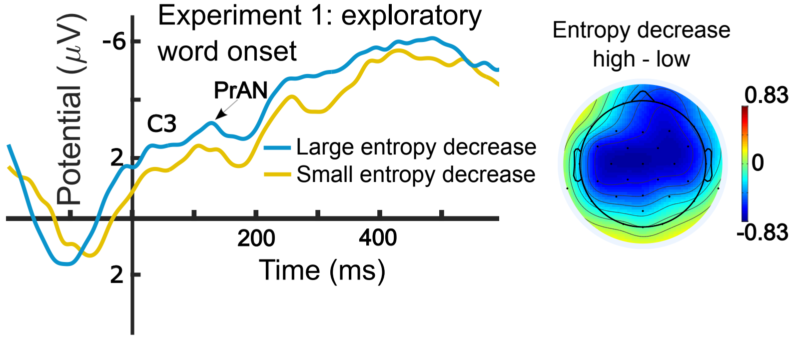

Figure 6. Second speech sounds which led to larger entropy decrease compared to the initial speech sound yielded more negative PrAN amplitudes.

PrAN amplitudes measured at C3 136-204 ms after word onset were modulated by entropy decrease, that is, how many continuations activated by the first speech sound were inconsistent with the second speech sound. Amplitudes were more negative for WIFs with high entropy decrease. An ANOVA showed that there was a significant difference between high/low groups, $F(1,78) = 10.93$, $p = 0.001$. A significant regression function was found, $F(1,78) = 13.71$, $p < 0.001$, $r = 0.39$, $r^2 = 0.15$.

3) $\text{PrAN} = -0.43(\textit{entropy}) - 9.8$

The effect was confirmed by a linear mixed-effects regression, $\beta = -0.42$, $SE = 0.13$, $df = 37.07$, $t = -3.13$, $p = 0.003$.

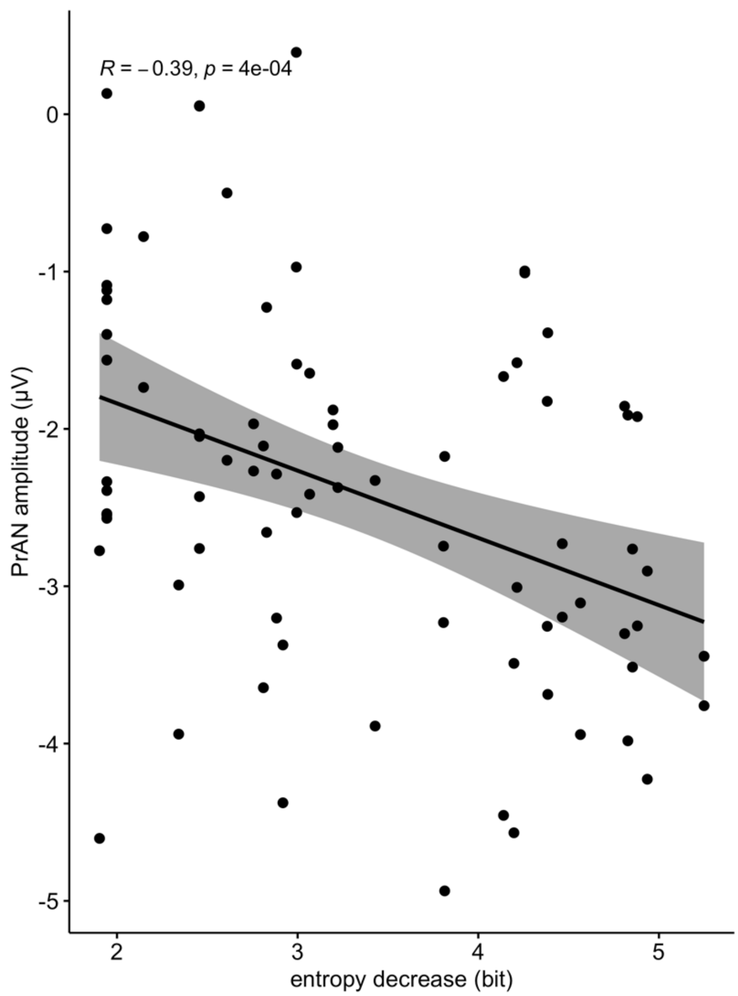

Figure 7. WIFS with larger entropy decrease from the first to the second speech sounds showed increasingly negative PrAN amplitudes.

Thus, test implications 4a and 4b were confirmed because PrAN amplitudes increased as entropy decrease increased. There were no significant effects of inhibition and test implications 3a and 3b were thus disconfirmed for effects at word onset. It could be that the number of possible candidates activated by the first speech sound was not a relevant measure because *all* first speech sounds had many possible continuations and *all* second speech sounds inhibited many candidates. Possibly, a measure taking into account the *proportion* of inhibited words would have been more adequate. Models of speech recognition diverge on the number of candidates actually activated by the first speech sound.

As the speech signal in experiment 1 progressed, prosodic information about whether a word had stød or non-stød became available to listeners. There were no effects of entropy decrease from the first two speech sounds to stød/non-stød cues. However, words in which stød and non-stød prosodic cues inhibited many continuations yielded more negative PrAN amplitudes over a frontal, right site compared to words in which those cues only inhibited few continuations. See Figure 8. The negative deflection reached its maximum over a right, frontal electrode, FC2, where an ANOVA showed a difference between WIFs with high and low inhibition rates 136-204 ms after stød/non-stød onset, $F(1,78) = 9.73$, $p = 0.003$. The same distribution has previously been reported for stød as compared to non-stød in the original analysis of the data (Hjortdal, 2021). This effect was seen in a somewhat later time window, 260-430 ms after stød/non-stød onset, and is visible in Figure 7 as negative deflections for the stød conditions peaking around 350 ms after word onset. However, the present findings point to both stød and non-stød ruling out irrelevant candidates, but that the effect lasts longer for stød. This might be because stød is a stronger cue: Upon hearing irregular vocal fold vibration and intensity fall, candidates without stød can immediately be ruled out. Non-stød, on the other hand, is the absence of that same cue and listeners

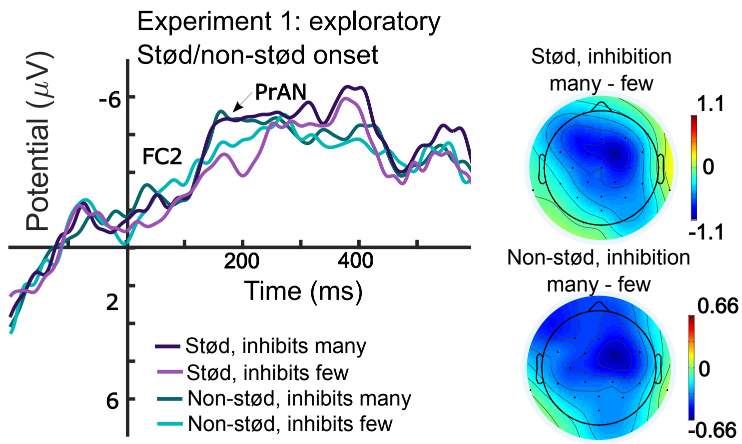

Figure 8. Items with stød and non-stød prosodic cues which inhibited many continuations showed more negative amplitudes 136-204 ms after stød/non-stød onset.

might be more uncertain about whether a stød cue is yet to come. A negative regression was found, $F(1,78) = 4.47$, $p = 0.038$, $r = -0.24$, $r^2 = 0.05$.

4) PrAN = -0.52*(log inhibited continuations)* – 1.51

However, when tested in a linear mixed-effects model, the effect was not significant, $\beta = -0.61$, $SE = 0.31$, $df = 40.79$, $t = -1.95$, $p = 0.058$. Thus, for effects at stød/non-stød onset, test implication 3a, that WIFs inhibiting more continuations would yield more negative PrAN amplitudes, was confirmed, but 3b was disconfirmed. It could be that nuisance parameters, such as sentence context, included in this model explained more of the variance in the data. The effect would require replication in an experiment with better control of such effects.

5 General discussion

The present study combined data from a frequency-weighted pronunciation lexicon with two ERP experiments, investigating the neural correlates of lexical competition and predictive certainty. ERPs 136-204 ms after word onset for spoken words with low and high lexical competition, respectively were compared. Further, ERP effects of entropy decrease and inhibition were investigated as the speech signal unfolded and more information became available to listeners.

5.1 Replication studies

Experiment 1 showed that the PrAN amplitude varied as a function of Danish word-beginnings' number of continuations 136-204 ms after word onset. As the number of continuations decreased, the negative PrAN amplitude increased. This is in line with Roll et al. (2017). In a study with speakers of Swedish, the authors reported PrAN amplitudes varying with the number of continuations and the total frequency of those continuations in the same spatiotemporal window.

A new measure, entropy, was investigated. Entropy can be understood as an index of how certain listeners are about the word they are hearing. In the present study, entropy appeared to be a relevant measure in experiment 1 where there was a significant difference in ERPs between low and high groups and a linear regression was found. The entropy measure is compatible with how posterior probabilities are calculated in Shortlist B (Norris & McQueen, 2008). Here, posterior probabilities of lexical candidates depend on their prior probabilities (e.g. frequency) and the given evidence (the first speech sounds of the signal). Low entropy would be consistent with one or a few lexical candidates with high posterior probabilities at a given time point whereas high entropy would be consistent with many candidates with low posterior probabilities at a given time.

For word beginnings in experiment 2, there were significant differences between word groups with few as compared to many continuations and low as compared to high entropy, but no significant regressions were found. This might be due to reduced power owing to a smaller number of trials in experiment 2 than in experiment 1. Further, the effect was more central than what was observed for experiment 1 as well as Roll et al. (2017). This should be investigated further.

An important finding of Roll et al. (2017) was that the negativity for *few* lexical competitors correlated with increased overall neural activity and increased BOLD in Broca's area. This negativity, reflecting increased activity, which was replicated in the present study, is somewhat difficult to reconcile with fewer lexical competitors which one would instinctively assume to yield less activation. In Shortlist (Norris, 1994; Norris & McQueen, 2008) and TRACE (McClelland & Elman, 1986; McClelland et al., 2014), there is lateral inhibition between lexical candidates which could explain why high lexical competition does not show increased activity, but these models do not explain why low lexical competition would show increased activity. Roll et al. (2017) proposed that the effect reflected listeners' certainty about and commitment to the few candidates consistent with the signal. The distributed model of speech perception (Gaskell & Marslen-Wilson, 1997) would make somewhat similar predictions. According to this model, initial speech sounds do not just activate phonological forms but also semantic features of lexical candidates. The degree of semantic activation depends on the number of words are activated by a word beginning. If many candidates are activated at the same time, semantic activation is weak or even non-existing. Speculating, the stronger effect of few lexical continuations might reflect stronger semantic activation of few lexical continuations.

5.2 Effects of inhibition and entropy decrease

Another interpretation of the negativity for few continuations was that it might reflect lexical selection and inhibition of irrelevant lexical candidates (Roll et al., 2017). If a WIF has few continuations, many candidates would need to be inhibited. Effects of inhibition were investigated taking a more exploratory approach. At word onset, there were no effects of the number of lexical candidates activated by the first speech sound and inhibited by the second. Later, however, as prosodic stød and non-stød cues became available to listeners, there was an effect of inhibition. However, while an ANOVA showed a significant difference between continuations which inhibited many as compared to few continuations, the regression was not significant when tested in a linear mixed-effects model.

It may be that the number of *possible* candidates activated by the first speech sound and inhibited by the second is not a relevant measure at this early time point because the first speech sound had the potential to activate so many words that they were not all activated. This has to be investigated further, but the interpretation would be in line with the Shortlist models (Norris, 1994; Norris & McQueen, 2008), which predict that only a shortlist of the most relevant candidates is activated and compete, or lateral inhibition between activated candidates as in TRACE (McClelland & Elman, 1986; McClelland et al., 2014). On the contrary, according to the Cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978) and Neighbourhood Activation model (Goldinger et al., 1989; Luce, 1986), all possible candidates are activated, but in these models, activation is modulated by a word's frequency. Therefore, frequency-weighted measures, such as entropy, might be more relevant at this early point because it takes into account that not all continuations are activated equally strongly. The findings of Roll et al. (2017) and Söderström et al. (2016) indicate that frequency was important early in the recognition process but lost importance later. While Roll et al. (2017) reported effects of the frequency of possible continuations 136-204 ms after word onset, Söderström et al. (2016) only found effects of the number of continuations in a later time window, 136-280 ms after F_0 onset, occurring 265-409 ms after word onset. Effects of the combined frequency of the continuations were also investigated in the latter study but no effect was reported. This is also in line with how speech recognition is modelled in Shortlist B in which frequency priors are important when the perceptual evidence is sparse, for instance early in a word. The effect decreases as better perceptual input is available, such as later in a word (Norris & McQueen, 2008). This would indicate that frequencies are important very early in the competition process, whereas their role seem to diminish as more of the word is available and bottom-up cues might take over. In any case, with the results of the present study, it is difficult to say whether the effect *did* reflect inhibition of irrelevant candidates, but that not all possible continuations were in fact activated – or whether there was no inhibition at this early stage.

Effects of entropy decrease (i.e. the increase in predictive certainty attained from the emergence of specific speech sounds) were also investigated. Second speech sounds which led to larger entropy decrease yielded more negative PrAN amplitudes than second speech sounds yielding small entropy decrease. There were no effects of entropy decrease at stød/non-stød onset. This finding supports that the PrAN effect is modulated by predictive certainty, as proposed by Roll et al. (2017). The findings could indicate that predictive certainty plays a more prominent role early on when many competitors are in play but becomes less important as fewer lexical candidates remain and bottom-up perceptual cues take over.

5.3 Neural sources

While continuations showed a frontal effect, the effects of entropy and entropy decrease were more central. ERP topographies are deceiving because activity in one part of the brain can, at least in theory, manifest itself at a part of the skull over another brain region, depending on the orientation and combination of the underlying brain sources (Luck, 2014). Any speculations should therefore be taken with a grain of salt. However, the frontal effects of continuations are in line with studies reporting activity in left or bilateral inferior frontal gyrus associated with lexical competition (Righi et al., 2010; Roll et al., 2017; Roll et al., 2015; Zhuang et al., 2011; Zhuang et al., 2014), although the ERP effect is not informative about whether it stems from ventral parts, associated with lexical competition or dorsal parts, associated with selection. The more central effect of entropy might have sources in Heschl's gyrus in primary auditory cortex and adjacent posterior superior temporal gyrus. Klimovich-Gray et al. (2019) reported an effect of entropy change in left Heschl's gyrus starting 140 ms after word onset and Roll et al. (2015) reported effects in Heschl's gyrus, superior temporal gyrus and inferior frontal gyrus, correlating with gRMS peak 136 ms after word onset. In both studies, the effects were left-lateralised whereas the ERP effects of the present study seem to be relatively central.

Within the framework of Yildiz et al. (2013)'s predictive coding account of word recognition, the negative effect of entropy decrease might be understood as reflecting updated predictions from the second level (proposed to be Broca's area) to the first level (corresponding to primary processing areas, including Heschl's gyrus). Klimovich-Gray et al. (2019) interpreted the effect in Heschl's gyrus as reflecting interaction between low level phonological cues and high-level constraints, in line with predictive processing accounts. For the present study, a possible effect of entropy decrease (i.e. increased certainty about the lexical candidate) in primary auditory areas might indicate that activations of lexical candidates at higher areas can modulate activity at lower levels such as primary auditory processing areas, possibly pre-activating relevant features or phonemes. If this interpretation is correct, it would be consistent with bidirectional connections, that is, activity on higher levels modulating activity on lower levels, as postulated by the predictive coding model of speech recognition (Yildiz et al., 2013), TRACE (McClelland & Elman, 1986; McClelland et al., 2014) and the distributed model of speech perception (Gaskell & Marslen-Wilson, 1997). It would be less consistent with the Shortlist models (Norris, 1994; Norris & McQueen, 2008) and the fuzzy logical model of perception (Massaro & Cohen, 1991; Oden & Massaro, 1978) because according to those models, connections are unidirectional and purely feedforward. This should be followed up in a combined fMRI and ERP study to further investigate the spatial distribution of the effect of entropy decrease because, as mentioned earlier, ERP topographies are deceiving.

Overall, the findings are in line with eye-tracking studies indicating that competitors are activated in parallel by the acoustic input and continuously mapped onto lexical candidates which are inhibited once they become inconsistent with the speech signal (Allopenna et al., 1998; Dahan et al., 2001a; Dahan et al., 2001b; Magnuson et al., 2007). This was seen by the ERPs varying with the number of competitors compatible with the speech signal 136-204 ms after word onset. The findings are also compatible with models of speech recognition predicting that competitors are activated based on bottom-up information in the speech signal based on initial speech sounds. Subsequently, competitors drop out, are inhibited or have decreasing probabilities, depending on the model, as more of the speech signal becomes available. Such models are the Cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), Shortlist (Norris, 1994; Norris & McQueen, 2008), TRACE (McClelland & Elman, 1986; McClelland et al., 2014) and the distributed model of speech perception (Gaskell & Marslen-Wilson, 1997). It would seem less compatible with the LAFs model (Klatt, 1979) and the fuzzy logical model of perception (Massaro

& Cohen, 1991; Oden & Massaro, 1978) where units are diphones rather than phonemes, because effects of entropy decrease were observed for candidates activated by the first speech sound already upon hearing the second speech sound. There was no effect of the number of continuations inhibited by the second phoneme. This could be because not all candidates were activated by the first speech sound, as predicted by e.g. the Cohort model, but that only a subset or shortlist of candidates was activated as predicted by the Shortlist models (Norris, 1994; Norris & McQueen, 2008). It might also be that candidates were not inhibited yet at this early stage.

6 Conclusion

Word beginnings with few as compared to many continuations showed more negative PrAN amplitudes, replicating the findings of (Roll et al., 2017). Further, effects of entropy (i.e. predictive certainty) were isolated. In one experiment, PrAN amplitudes varied continuously with the number of possible continuations and entropy, increasing when lexical competition and predictive certainty increased. In the other experiment, no such continuous effects were found.

There was no effect of the number of candidates activated by the first speech sound and inhibited by the second, possibly because the number of *possible* candidates inhibited was not a relevant measure at this early time. Effects were present later, for candidates inhibited by stød and non-stød prosodic cues. Further, there was an effect of entropy decrease 136-204 ms after word onset over central sites. Effects of entropy and entropy change 136-204 ms after word onset might reflect updated expectations on lower-level areas modulated by pre-activated lexical candidates on higher levels. This would be in line with the predictive coding model of speech recognition (Yildiz et al., 2013) and activation feedback as in TRACE and the distributed model (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; McClelland et al., 2014).

The findings are in line with eye-tracking studies reporting continuous effects of lexical competition (Allopenna et al., 1998; Magnuson et al., 2007). An inverse correlation between the number of activated candidates and PrAN amplitude is compatible with the connectionist models (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; McClelland et al., 2014; Norris & McQueen, 2008; Norris et al., 2000). These models predict that a number of lexical candidates are activated based on bottom-up speech input and continuously inhibited when they become inconsistent with the speech signal. Altogether, the distributed model of speech perception (Gaskell & Marslen-Wilson, 1997) best explains the data. Like TRACE, the model predicts that lexical effects can affect perception at lower levels which is in line with a possible interplay between activated lexical candidates and updated lower-level phonetic expectations. Further, the model predicts weaker semantic activation for words in larger cohorts. Therefore, the increased negativity associated with increased neural activity could reflect stronger semantic activation.

7 Outlook for future research

Effects of lexical competition and entropy should be further investigated in a study with words without stødbasis but varying more in number of continuations. Such an experiment could further explore whether the lack of a significant regression can be explained by too little variance in the number of continuations or whether other factors play in. Effects of inhibition should be further investigated to examine whether there was no inhibition already 136-204 ms after word onset – or if the number of *possible* candidates inhibited was simply not a relevant measure 136-204 ms after word onset, because too many candidates were in the cohort. Late effects of inhibition should also be investigated. In the present study, there was an effect of inhibition at stød/non-stød onset, but the regression was not significant when tested in a linear mixed-effects model. Future studies should better control for effects of sentence context to investigate late effects of inhibition.

8 References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419-439. https://doi.org/10.1006/jmla.1997.2558

Andersen, G. (2011). *Leksikalsk database for dansk*. Technical report, Nasjonalbiblioteket. https://www.nb.no/sprakbanken/en/resource-catalogue/oai-nb-no-sbr-26/

Archibald, L. M., & Joanisse, M. F. (2011). Electrophysiological responses to coarticulatory and word level miscues. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1275-1291. https://doi.org/10.1037/a0023506

Asmussen, J. (2015). *Design of the ePOS tagger*. Det Danske Sprog- og Litteraturselskab. https://korpus.dsl.dk/documentation/index.html

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Basbøll, H. (2005). *The Phonology of Danish*. Oxford University Press.

Basbøll, H. (2014). Danish stød as evidence for grammaticalization of suffixal positions in word structure. *Acta Linguistica Hafniensia*, *46*(2), 137-158. https://doi.org/10.1080/03740463.2014.989710

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bjerg Nielsen, K. (2001). *Transskriptionskonventioner*. Techical report, Nasjonalbiblioteket. https://www.nb.no/sprakbanken/en/resource-catalogue/oai-nb-no-sbr-26/

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer (Version 6.1.16)*. https://praat.org

Choi, H. S., Marslen-Wilson, W. D., Lyu, B., Randall, B., & Tyler, L. K. (2021). Decoding the real-time neurobiological properties of incremental semantic interpretation. *Cerebral Cortex*, *31*(1), 233-247. https://doi.org/10.1093/cercor/bhaa222

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 81-94. https://doi.org/10.1037//0278-7393.19.1.81

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317-367. https://doi.org/10.1006/cogp.2001.0750

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5/6), 507-534. https://doi.org/10.1080/01690960143000074

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Denes, P. D., & Pinson, E. (1993). *The speech chain*. W. H. Freeman and Company.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academic Sciences of the United States of America*, *109*(8), E505-514. https://doi.org/10.1073/pnas.1113427109

Fischer-Jørgensen, E. (1989). *A phonetic study of the stød in standard Danish*. University of Turku Phonetics.

Forster, K. I. (1976). Acessing the mental lexicon. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language processing* (pp. 257-287). North Holland.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1-11. https://doi.org/10.1016/j.bandl.2014.10.006

Frid, J. (2015). Frequency weighting of the NST lexicon. *A byteful of air*. https://person2.sol.lu.se/JohanFrid/webapps/pmwiki/index.php/Blog/20150918

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B Biological Sciences*, *360*, 815-836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. https://doi.org/10.1038/nrn2787

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110-125. https://doi.org/10.1037//0096-1523.6.1.110

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5-6), 613-656. https://doi.org/10.1080/016909697386646

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501-518. https://doi.org/10.1016/0749-596x(89)90009-0

Goldinger, S. D., Pisoni, D. B., & Luce, P. A. (1996). Speech perception and spoken word recognition: Research and theory. In N. Lass (Ed.), *Principles of Experimental Phonetics* (pp. 277-327). Moseby.

Grønnum, N. (2005). *Fonetik og Fonologi* (3rd ed.). Akademisk Forlag.

Grønnum, N. (2009). A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication*, *51*(7), 594-603. https://doi.org/10.1016/j.specom.2008.11.002

Grønnum, N., & Basbøll, H. (2001). Consonant length, stød and morae in Standard Danish. *Phonetica*, *58*, 230-253. https://doi.org/10.1159/000046177

Hempel, C. (1966). *Philosophy of Natural Science*. Prentice Hall.

Henrichsen, P. J. (2002). Some frequency based differences between spoken and written Danish. *Gothenburg Papers in Theoretical Linguistics*, *88*.

Hjortdal, A. (2021). *Danish stød as a cue to upcoming suffixes: an ERP and response time study* [Bachelor's thesis, Lund University].

Høeg, A. (2020). *Nye stød i københavnsk rigsmål - en undersøgelse af forekomst og variation* [Master's thesis, Uppsala University].

Hunter, C. R. (2013). Early effects of neighborhood density and phonotactic probability of spoken words on event-related potentials. *Brain and Language*, *127*(3), 463-474. https://doi.org/10.1016/j.bandl.2013.09.006

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163-178. https://doi.org/10.1017/S0048577200980259

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279-312. https://doi.org/https://doi.org/10.1016/S0095-4470(19)31059-9

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement*, *52*, 3-6.

Klimovich-Gray, A., Tyler, L. K., Randall, B., Kocagoncu, E., Devereux, B., & Marslen-Wilson, W. D. (2019). Balancing prediction and sensory input in speech comprehension: The spatiotemporal dynamics of word recognition in context. *Journal of Neuroscience*, *39*(3), 519-527. https://doi.org/10.1523/JNEUROSCI.3573-17.2018

Kretzschmar, F., & Alday, P. M. (forthcoming). Principles of statistical analyses: Old and new tools. In M. Grimaldi, Y. Shtyrov, & E. Brattico (Eds.), *Language Electrified. Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language*. Springer.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss

Lewis, M. P., Simons, G. F., & Fennig, C. D. (2015). *Ethnologue: Languages of the World, Eighteenth edition* http://www.ethnologue.com

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon* [Doctoral dissertation, Indiana University]. ERIC. Bloomington.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear*, *19*(1), 1-36. https://doi.org/10.1097/00003446-199802000-00001

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). The MIT Press.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133-156. https://doi.org/10.1080/03640210709336987

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1-71. https://doi.org/10.1016/0010-0277(80)90015-3

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102. https://doi.org/10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29-63. https://doi.org/https://doi.org/10.1016/0010-0285(78)90018-X

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, *23*(4), 558-614. https://doi.org/10.1016/0010-0285(91)90006-a

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Frontiers in Psychology*, *4*, 503. https://doi.org/10.3389/fpsyg.2013.00503

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38*(6), 1139-1189. https://doi.org/10.1111/cogs.12146

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189-234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357-395. https://doi.org/10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences 23*(3), 299-325; discussion 325-270. https://doi.org/10.1017/s0140525x00003241

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4-18. https://doi.org/10.1080/23273798.2015.1081703

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172-191. https://doi.org/10.1037/0033-295X.85.3.172

Peirce, J., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behaviour Research Methods*. https://doi.org/10.3758/s13428-018-01193-y

Pereltsvaig, A. (2017). *Languages of the world - an introduction*. Cambridge University Press.

Perkell, J. S. (2006). Articulatory processes. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences*. Wiley-Blackwell.

R Core Team. (2017). *R: A language and environment for statistical computing*. In R Foundation for Statistical Computing. http://www.R-project.org/

Rasinger, S. M. (2013). *Quantitative Research in Linguistics* (2nd ed.). Bloomsbury.

Riad, T. (2012). Culminativity, stress and tone accent in Central Swedish. *Lingua*, *122*(13), 1352-1379. https://doi.org/10.1016/j.lingua.2012.07.001

Righi, G., Blumstein, S. E., Mertus, J., & Worden, M. S. (2010). Neural systems underlying lexical competition: An eye tracking and fMRI study. *Journal of Cognitive Neuroscience*, *22*(2), 213-224. https://doi.org/10.1162/jocn.2009.21200

Roll, M. (2015). A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nordic Journal of Linguistics*, *38*(2), 149-162. https://doi.org/10.1017/S0332586515000189

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology – ERPs of Swedish word processing. *Brain Research*, *1330*, 114-123. https://doi.org/10.1016/j.brainres.2010.03.020

Roll, M., Söderström, P., Frid, J., Mannfolk, P., & Horne, M. (2017). Forehearing words: Pre-activation of word endings at word onset. *Neuroscience Letters*, *658*, 57-61. https://doi.org/10.1016/j.neulet.2017.08.030

Roll, M., Söderström, P., & Horne, M. (submitted). Pre-activation negativity (PrAN): A neural index of predictive strength of phonological cues.

Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., & Horne, M. (2015). Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain and Language*, *150*, 14-21. https://doi.org/10.1016/j.bandl.2015.07.009

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145-163. https://doi.org/10.1016/j.jml.2013.11.002

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, *162*, 42-45. https://doi.org/10.1016/j.bandl.2016.08.001

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379-423, 623-656. https://doi.org/https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4-31). Psychology Press.

Söderström, P., Horne, M., Frid, J., & Roll, M. (2016). Pre-activation negativity (PrAN) in brain potentials to unfolding words. *Frontiers in Human Neuroscience*, *10*, 1-11. https://doi.org/10.3389/fnhum.2016.00512

Söderström, P., Horne, M., & Roll, M. (2017a). Stem tones pre-activate suffixes in the brain. *Journal of Psycholinguistic Research*, *46*, 271-280. https://doi.org/10.1007/s10936-016-9434-2

Strand, J., Simenstad, A., Cooperman, A., & Rowe, J. (2014). Grammatical context constrains lexical competition in spoken word recognition. *Memory & Cognition*, *42*(4), 676-687. https://doi.org/10.3758/s13421-013-0378-6

Stufflebeam, R. (2006). *Connectionism*. The Mind Project. https://mind.ilstu.edu/curriculum/connectionism_intro/

Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 270-278. https://doi.org/10.1037/0096-1523.28.2.270

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Memory and Language*, *40*, 374-408. https://doi.org/https://doi.org/10.1006/jmla.1998.2618

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, *41*(3), 262-275. https://doi.org/10.3758/bf03208224

Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506-2516. https://doi.org/10.1093/cercor/bhv075

Yildiz, I. B., von Kriegstein, K., & Kiebel, S. J. (2013). From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLoS Computational Biology*, *9*(9), 1-6. https://doi.org/10.1371/journal.pcbi.1003219

Zhuang, J., Randall, B., Stamatakis, E. A., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The interaction of lexical semantics and cohort competition in spoken word recognition: An fMRI study. *Journal of Cognitive Neuroscience*, *23*(12), 3778-3790. https://doi.org/10.1162/jocn_a_00046

Zhuang, J., Tyler, L. K., Randall, B., Stamatakis, E. A., & Marslen-Wilson, W. D. (2014). Optimally efficient neural systems for processing spoken language. *Cerebral Cortex*, *24*(4), 908-918. https://doi.org/10.1093/cercor/bhs366

Appendix A

Experiment 1 (stødbasis)		Experiment 2 (non-stødbasis)	
Alk ‘razorbill’	[ˈalˀg̊n̩] *algen* ‘the razorbill’ [ˈalg̊ə] *alge* ‘razorbills’	*Biks* ‘corner shop’	[ˈb̥eg̊sn̩] *biksen* ‘the corner shop’ [ˈb̥eg̊sə] *bikes* ‘corner shops’
Bold ‘ball’	[ˈb̊ʌlˀd̥n̩] *bolden* ‘the ball [ˈb̊ʌld̥ə] *bolde ‘balls’*	*Blok* ‘pad, block’	[ˈb̥lʌg̊ŋ̍] *blokken* ‘the pad, the block’ [ˈb̥lʌg̊ə] *blokke* ‘the pads, the blocks’
Bænk ‘bench’	[ˈb̊ɛŋˀg̊ŋ̍] *bænken* ‘the bench’ [ˈb̊ɛŋg̊ə] *bænke* ‘the benches’	*Boks* ‘box’	[ˈb̥ʌg̊sn̩] *boksen* ‘the box’ [ˈb̥ʌg̊sə] *bokse* ‘the boxes’
Damp ‘steam’	[ˈd̥ɑmˀb̥n̩] *dampen* ‘the steam’ [ˈd̥ɑmb̥ə] *dampe* ‘steams’	*Buk* ‘billy goat’	[ˈb̥ɔg̊ŋ̍] *bukken* ‘the billy goat’ [ˈb̥ɔg̊ə] *bukke* ‘billy goats’
Falk ‘falcon’	[ˈfalˀg̊ŋ̍] *falken* ‘the falcon’ [ˈfalg̊ə] *falke* ‘falcons’	*Busk* ‘bush’	[ˈb̥usg̊ŋ̍] *busken* ‘the bush’ [ˈb̥usg̊ə] *buske* ‘bushes’
Fjært ‘fart’	[ˈfjæɐ̯ˀd̥n̩] *fjærten* ‘the fart’ [ˈfjæɐ̯d̥ə] *fjærte* ‘farts’	*Bæk* ’brook’	[ˈb̥ɛg̊ŋ̍] *bækken* ‘the brook’ [ˈb̥ɛg̊ə] *bække* ‘brooks’
Flab ‘lout, mouth’	[ˈflæb̥n̩] *flaben* ‘the lout, the mouth’ [ˈflæb̥ə] *flabe* ‘louts, mouths’	*Disk* ‘disk’	[ˈd̥esg̊ŋ̍] *disken* ‘the disk’ [ˈd̥esg̊ə] *diske* ’disks’
Font ‘font’	[ˈfʌnˀd̥n̩] *fonten* ‘the font’ [ˈfʌnd̥ə] *fonte* ‘fonts’	*Drik* ‘drink’	[ˈd̥ʁɛg̊ŋ̍] *drikken* ‘the drink’ [ˈd̥ʁɛg̊ə] *drikke* ’drinks’
Greb ‘pitchfork’	[ˈg̊ʁɛˀb̥n̩] *greben* ‘the pitchfork’ [ˈg̊ʁɛ:b̥ə] *grebe* ‘pitchforks’	*Duft* ‘(pleasant) smell’	[ˈd̥ɔfd̥n̩] *duften* ‘the smell’ [ˈd̥ɔfd̥ə] *dufte* ‘smells’
Hank ‘handle’	[ˈhɑŋˀg̊ŋ̍] *hanken* ‘the handle’ [ˈhɑŋg̊ə] *hanke* ‘handles’	*Dusk* ‘tuft’	[ˈd̥usg̊ŋ̍] *dusken* ‘the tuft’ [ˈd̥usg̊ə] *duske* ‘tufts’
Helt ‘hero’	[ˈhɛlˀd̥n̩] *helten* ‘the hero’ [ˈhɛld̥ə] *helte* ‘heroes’	*Flok* ‘flock’	[ˈflʌg̊ŋ̍] *flokken* ‘the flock’ [ˈflʌg̊ə] *flokke* ‘flocks’
Hingst ‘stallion’	[ˈheŋˀsd̥n̩] *hingsten* ‘the stallion’ [ˈheŋsd̥ə] *hingste* ‘stallions’	*Gift* ‘poison’	[ˈg̊ifd̥n̩] *giften* ‘the poison’ [ˈg̊ifd̥ə] *gifte* ‘poisons’
Hob ‘crowd’	[ˈhoˀb̥n̩] *hoben* ‘the crowd’ [ˈho:b̥ə] *hobe* ‘crowds’	*Hat* ‘hat’	[ˈhad̥n̩] *hatten* ‘the hat’ [ˈhad̥ə] *hatte* ‘hats’
Hvalp ‘puppy’	[ˈvalˀb̥n̩] *hvalpen* ‘the puppy’ [ˈvalb̥ə] *hvalpe* ’puppies’	*Hest* ‘horse’	[ˈhɛsd̥n̩] *hesten* ‘the horse’ [ˈhɛsd̥ə] *heste* ‘horses’
Kalk ‘chalice’	[ˈkʰalˀg̊ŋ̍] *kalken* ‘the chalice’ [ˈkʰalg̊ə] *kalke* ‘chalices’	*Hæk* ‘hedge’	[ˈhɛg̊ŋ̍] *hækken* ‘the hedge’ [ˈhɛg̊ə] *hække* ‘hedges’
Kamp ‘fight’	[ˈkʰɒmˀb̥ə] *kampen* ‘the fight’ [ˈkʰɒmb̥ə] *kampe* ‘fights’	*Kat* ‘cat’	[ˈkʰad̥n̩] *katten* ‘the cat’ [ˈkʰad̥ə] *katte* ‘cats’
Kilt ‘kilt’	[ˈkʰilˀd̥n̩] *kilten* ‘the kilt’ [ˈkʰild̥ə] *kilte* ’kilts’	*Kok* ‘chef’	[ˈkʰʌg̊ŋ̍] *kokken* ‘the chef’ [ˈkʰʌg̊ə] *kokke* ‘chefs’
Krank ‘crank’	[ˈkʰʁɑŋˀg̊ŋ̍] *kranken* ‘the crank’ [ˈkʰʁɑŋg̊ə] *kranke* ‘cranks’	*Kost* ‘broom’	[ˈkʰɔsd̥n̩] *kosten* ‘the broom’ [ˈkʰɔsd̥ə] *koste* ‘brooms’
Kælk ‘sledge’	[ˈkʰɛlˀg̊ŋ̍] *kælken* ‘the sledge’	*Krop* ‘body’	[ˈkʰʁʌb̥n̩] *krokken* ‘the body’

	[ˈkʰɛlg̊ə] *kælke* ‘sledges’		[ˈkʰʁʌb̥ə] *kroppe* ‘bodies’
Lort ‘crap’	[ˈloɐ̯ˀd̥n̩] *lorten* ‘the crap’ [ˈloɐ̯d̥ə] *lorte* ‘craps’	*Kusk* ‘driver (of horse-drawn carriage)’	[ˈkʰusg̊n̩] *kusken* ‘the driver’ [ˈkʰusg̊ə] *kuske* ‘drivers’
Læg ‘calf’	[ˈlɛˀg̊n̩] *læggen* ‘the calf’ [ˈlɛːg̊ə] *lægge* ‘calfs’	*Kvast* ‘taffel’	[ˈkʰvasd̥n̩] *kvasten* ‘the taffel’ [ˈkʰvasd̥n̩] *kvaste* ‘taffels’
Milt ‘spleen’	[ˈmilˀd̥n̩] *milten* ‘the spleen’ [ˈmild̥ə] *milte* ‘spleens’	*Kvist* ‘twig’	[ˈkʰvesd̥ə] *kvisten* ‘the twig’ [ˈkʰvesd̥n̩] *kviste* ‘twigs’
Pilk ‘jig’	[ˈpʰilˀg̊n̩] *pilken* ‘the jig’ [ˈpʰilg̊ə] *pilke* ‘jigs’	*Kæp* ‘stick’	[ˈkʰɛb̥n̩] *kæppen* ‘the stick’ [ˈkʰɛb̥ə] *kæppe* ‘sticks’
Pulk ‘pulk’	[ˈpʰulˀg̊n̩] *pulken* ‘the pulk’ [ˈpʰulg̊ə] *pulke* ‘pulks’	*Lugt* ‘smell	[ˈlɔg̊d̥n̩] *lugten* ‘the smell’ [ˈlɔg̊d̥ə] *lugte* ‘smells’
Salt ‘salt’	[ˈsalˀd̥n̩] *salten* ‘the salt’ [ˈsald̥ə] *salte* ‘salts’	*Ost* ‘cheese’ ** excluded*	[ˈɔsd̥n̩] *osten* ‘the cheese’ [ˈɔsd̥ə] *oste* ‘cheeses’
Skab ‘closet’	[ˈsg̊æˀb̥n̩] *skabet* ‘the closet’ [ˈsg̊æːb̥ə] *skabe* ‘closets’	*Pisk* ‘whip’	[ˈpʰisg̊n̩] *pisken* ‘the whip’ [ˈpʰisg̊ə] *piske* ‘whips’
Skalk ‘trickster’	[ˈsg̊alˀg̊n̩] *skalken* ‘trickster’ [ˈsg̊alg̊ə] *skalke* ‘tricksters’	*Rig* ‘rigging’	[ˈʁɛg̊n̩] *riggen* ‘the rigging’ [ˈʁɛg̊ə] *rigge* ‘rigs’
Skalp ‘scalp’	[ˈsg̊alˀb̥n̩] *skalpen* ‘the scalp’ [ˈsg̊alb̥ə] *skalpe* ‘scalps’	*Rist* ‘grating’	[ˈʁɛsd̥n̩] *risten* ‘the grating’ [ˈʁɛsd̥ə] *riste* ‘gratings’
Skank ‘shank’	[ˈsg̊ɑŋˀg̊n̩] *skanken* ‘the shank’ [ˈsg̊ɑŋg̊ə] *skanke* ‘shanks’	*Rok* ‘spinning wheel’	[ˈʁʌg̊n̩] *rokken* ‘the spinning wheel’ [ˈʁʌg̊ə] *rokke* ‘spinning wheels’
Skilt ‘sign’	[ˈsg̊eˀld̥ð] *skiltet* ‘the sign’ [ˈsg̊eld̥ə] *skilte* ‘signs’	*Skakt* ‘shaft’	[ˈsg̊ɑgd̥n̩] *skakten* ‘the shaft’ [ˈsg̊ɑgd̥ə] *skakte* ‘shafts’
Skænk ‘sideboard’	[ˈsg̊ɛŋˀg̊n̩] *skænken* ‘the sideboard’ [ˈsg̊ɛŋg̊ə] *skænke* ‘sideboards’	*Skat* ‘treasure’	[ˈsg̊ad̥n̩] *skatten* ‘the treasure’ [ˈsg̊ad̥ə] *skatte* ‘treasures’
Stab ‘staff’	[ˈsd̥æˀb̥n̩] *staben* ‘the staff’ [ˈsd̥æːb̥ə] *stabe* ‘staffs’	*Slot* ‘castle’	[ˈslʌd̥ð] *slottet* ‘the castle’ [ˈslʌd̥ə] *slotte* ‘castles’
Stank ‘stink’	[ˈsd̥ɑŋˀg̊n̩] *stanken* ‘the stink’ [ˈsd̥ɑŋg̊ə] *stanke* ‘stinks’	*Stak* ‘stack, pile’	[ˈsd̥ɑg̊n̩] *stakken* ‘the stack, the pile’ [ˈsd̥ɑg̊ə] *stakke* ‘stacks, piles’
Stilk ‘stalk’	[ˈsd̥elˀg̊n̩] *stilken* ‘the stalk’ [ˈsd̥elg̊ə] *stilke* ‘stalks’	*Stok* ‘walking stick, cane’	[ˈsd̥ʌg̊n̩] *stokken* ‘the walking stick, the cane’ [ˈsd̥ʌg̊ə] *stokke* ‘walking sticks, canes’
Sump ‘swamp’	[ˈsɔmˀb̥n̩] *sumpen* ‘the swamp’ [ˈsɔmb̥ə] *sumpe* ‘swamps’	*Stub* ‘stump’	[ˈsd̥ub̥n̩] *stubben* ‘the stump’ [ˈsd̥ub̥ə] *stubbe* ‘stumps’
Svamp ‘mushroom’	[ˈsvɑmˀb̥n̩] *svampen* ‘the mushroom’ [ˈsvɑmb̥ə] *svampe* ‘mushrooms’	*Sæk* ‘bag’	[ˈsɛg̊n̩] *sækken* ‘the bag’ [ˈsɛg̊ə] *sække* ‘bags’
Telt ‘tent’	[ˈtˢɛlˀd̥n̩] *teltet* ‘the tent’ [ˈtˢɛld̥ə] *telte* ‘tents’	*Top* ‘top’	[ˈtˢʌb̥n̩] *toppen* ‘the top’ [ˈtˢʌb̥ə] *toppe* ‘tops’
Tolk ‘interpretor’	[ˈtˢʌlˀg̊n̩] *tolken* ‘the interpretor’ [ˈtˢʌlg̊ə] *tolke* ‘interpretors’	*Tragt* ‘funnel’	[ˈtˢʁɑg̊d̥n̩] *tragten* ‘the funnel’ [ˈtˢʁɑg̊d̥ə] *tragte* ‘funnels’
Væg ‘wall’	[ˈvɛˀg̊n̩] *væggen* ‘the wall’ [ˈvɛːg̊ə] *vægge* ‘walls’	*Vægt* ‘scale’	[ˈvɛgd̥n̩] *vægten* ‘the scale’ [ˈvɛgd̥ə] *vægte* ‘scales’
Ulk ‘sculpin’	[ˈulˀg̊n̩] *ulken* ‘the sculpin’ [ˈulg̊ə] *ulke* ‘sculpins’	*Vask* ‘sink’	[ˈvasg̊n̩] *Vasken* ‘the sink’ [ˈvasg̊ə] *vaske* ‘sinks’

Appendix B

Python: Extracting words from KorpusDK

```
# Requires all txt files from KorpusDK in a folder “KorpusDK” including
subfolders

import glob2

# create list of all txt files in the directory KorpusDK including subfolders
filenames = glob2.glob('KorpusDK/*/*.txt')

# opens all files, skip lines starting with “<” which mark next texts
with open('outfile_korpusdk900010.txt', 'w') as f:
    for file in filenames:
        with open(file) as infile:
            for line in infile:
                if line.startswith("<"):
                    continue
                if line.strip():
                    cols = line.split()

            # print cols[0] and cols[4], i.e. item and ePOS tag, to file
                    f.write(cols[0] + "\t" + cols[4] + "\n")
```

Unix: Calculating frequencies from KorpusDK

```
# requires outfile_korpusdk900010.txt
sort outfile_korpusdk900010.txt | uniq -c | sort -nr > korpusDK_pos_freq.txt
```

Python: Prepare NST lexicon in txt version for merge by replacing PoS-tags and lowercasing all orthographies

```
#Requires txt file with all words incl PoS tags from NST
import csv
import os

# define paths
word_pos_pron = os.path.join('Pronunciation',
'word_pos_pron_nopropernouns.csv')
word_pos_pron_replace_lower = os.path.join('Pronunciation',
'word_pos_pron_replace_lower_nopropernouns.csv')

# open files
with open(word_pos_pron, 'r') as file1, open(word_pos_pron_replace_lower,
'a',newline='') as file2:
   reader = csv.reader(file1, delimiter=',')
   writer = csv.writer(file2, delimiter=',')


# lowercase all in row[0] (orthography)
   for row in reader:
       lower = row[0].lower()
```

```
        row[0] = lower

# proper nouns have longer tags like “PM | Person”, this part removes the
part after |
        replace = row[1].split("|")
        if len(replace) > 1:
            row[1] = replace[0]

# this part replaces PoS tags with ePoS tags
        replaced = row[1].replace("NN", "N").replace("VB",
"V").replace("JJ","A").replace("PP", "T").replace("KN", "C").replace("SN",
"C").replace("PN", "P").replace("PS","P").replace("AB", "D").replace("RG",
"L").replace("RO", "L").replace("IN", "I").replace("IE", "U").replace("UO",
"X").replace("PM","N").replace("DT","P")

        row[1] = replaced

# write to file
        writer.writerow(row)          # write to writer file
```

Python: Merging pronunciation lexicon and frequency list

```
# requires: frequency list and pronunciation lexicon
import pandas as pd
import os

# define paths
KorpusDK_word_pos = os.path.join('Frequency', 'korpusDK_pos_freq.csv')
word_pos_pron = os.path.join('Pronunciation',
'word_pos_pron_replace_lower_nopropernouns.csv')

# import frequency list
frequency = pd.read_csv(KorpusDK_pos_freq)

#name columns
frequency.columns = ["freq", "word", "pos"]

# import pronunciation lexicon
pronunciation = pd.read_csv(word_pos_pron)

# name columns
pronunciation.columns = ["word", "pos", "pron1", "pron2", "pron3", "pron3",
"pron4", "pron5", "pron6", "pron7", "pron8", "pron9", "pron10", "pron11",
"pron12"]

# merge csv files based on word and word class (PoS-tag)
merged = frequency.merge(pronunciation, on=["word", "pos"])

# save as CSV
merged.to_csv("Pronunciation/freq_word_pos_pron_korpusDK_nopropernouns.csv",
index=False)
```

Python: Calculating continuations and entropy

Contact the author.

Appendix C

Experiment 1

Item	num	SAMPA	continuations	entropy
alken	sg	a-_l	246	5.11902177
alke	pl	a-_l	246	5.11902177
bolden	sg	b_6-	200	5.76309307
bolde	pl	b_6-	200	5.76309307
bænken	sg	b_E-	183	5.74632722
bænke	pl	b_E-	183	5.74632722
dampen	sg	d_A-	242	5.72859625
dampe	pl	d_A-	242	5.72859625
falken	sg	f_a-	451	5.50603778
falke	pl	f_a-	451	5.50603778
fjærten	sg	f_j_	149	5.0370972
fjærte	pl	f_j_	149	5.0370972
flaben	sg	f_l_	615	6.69482858
flabe	pl	f_l_	615	6.69482858
fonten	sg	f_6-	378	5.5091047
fonte	pl	f_6-	378	5.5091047
greben	sg	g_R_	891	6.54709065
grebe	pl	g_R_	891	6.54709065
hanken	sg	h_A-	256	6.23174115
hanke	pl	h_A-	256	6.23174115
helten	sg	h_E-	474	6.04733087
helte	pl	h_E-	474	6.04733087
hingsten	sg	h_e-	103	4.21430085
hingste	pl	h_e-	103	4.21430085
hoben	sg	h_o+	439	5.81957774
hobe	pl	h_o+	439	5.81957774
hvalpen	sg	v_a-	568	6.46133565
hvalpe	pl	v_a-	568	6.46133565
kalken	sg	k_a-	712	7.35681271
kalke	pl	k_a-	712	7.35681271
kampen	sg	k_A-	556	5.88431553
kampe	pl	k_A-	556	5.88431553

kilten	sg	k_i-	265	4.81774816
kilte	pl	k_i-	265	4.81774816
kranken	sg	k_R_	881	5.13292263
kranke	pl	k_R_	881	5.13292263
kælken	sg	k_E-	278	5.23334237
kælke	pl	k_E-	278	5.23334237
lorten	sg	l_o-	138	5.48264989
lorte	pl	l_o-	138	5.48264989
læggen	sg	l_E+	376	5.84505708
lægge	pl	l_E+	376	5.84505708
milten	sg	m_i-	848	6.18532529
milte	pl	m_i-	848	6.18532529
pilken	sg	p_i-	161	5.91905437
pilke	pl	p_i-	161	5.91905437
pulken	sg	p_u-	125	3.58770008
pulke	pl	p_u-	125	3.58770008
salten*	sg	s_a-	407	6.26401498
salte	pl	s_a-	407	6.26401498
skabet	sg	s_g_	2157	8.57759934
skabe	pl	s_g_	2157	8.57759934
skalken	sg	s_g_	2157	8.57759934
skalke	pl	s_g_	2157	8.57759934
skalpen	sg	s_g_	2157	8.57759934
skalpe	pl	s_g_	2157	8.57759934
skanken	sg	s_g_	2157	8.57759934
skanke	pl	s_g_	2157	8.57759934
skiltet	sg	s_g_	2157	8.57759934
skilte	pl	s_g_	2157	8.57759934
skænken	sg	s_g_	2157	8.57759934
skænke	pl	s_g_	2157	8.57759934
staben	sg	s_d_	2535	8.06505169
stabe	pl	s_d_	2535	8.06505169
stanken	sg	s_d_	2535	8.06505169
stanke	pl	s_d_	2535	8.06505169
stilken	sg	s_d_	2535	8.06505169
stilke	pl	s_d_	2535	8.06505169
svampen	sg	s_v_	393	6.71414808
svampe	pl	s_v_	393	6.71414808

sumpen	sg	s_O-	184	5.585752
sumpe	pl	s_O-	184	5.585752
tolken	sg	t_6-	337	6.26791597
tolke	pl	t_6-	337	6.26791597
teltet	sg	t_E-	468	6.54218013
telte	pl	t_E-	468	6.54218013
ulken	sg	u-_l	38	3.74657863
ulke	pl	u-_l	38	3.74657863
væggen	sg	v_E+	79	3.79837479
vægge	pl	v_E+	79	3.79837479

Experiment 2

Item	**num**	**SAMPA**	**continuations**	**entropy**
biksen	sg	b_e-	1232	7.6211704
bikse	pl	b_e-	1232	7.6211704
blokken	sg	b_l_	516	6.19002694
blokke	pl	b_l_	516	6.19002694
boksen	sg	b_6-	200	5.76309307
bokse	pl	b_6-	200	5.76309307
bukken	sg	b_O-	342	5.30303994
bukke	pl	b_O-	342	5.30303994
busken	sg	b_u-	237	5.48690087
buske	pl	b_u-	237	5.48690087
bækken	sg	b_E-	183	5.74632722
bække	pl	b_E-	183	5.74632722
disken	sg	d_e-	551	6.13485774
diske	pl	d_e-	551	6.13485774
drikken	sg	d_R_	410	6.08096401
drikke	pl	d_R_	410	6.08096401
duften	sg	d_O-	89	4.774846
dufte	pl	d_O-	89	4.774846
dusken	sg	d_u-	47	3.97000941
duske	pl	d_u-	47	3.97000941
flokken	sg	f_l_	615	6.69482858
flokke	pl	f_l_	615	6.69482858
giften	sg	g_i-	139	5.27348873
gifte	pl	g_i-	139	5.27348873
hatten	sg	h_a-	472	6.18080388

hatte	pl	h_a-	472	6.18080388
hesten	sg	h_E-	474	6.04733087
heste	pl	h_E-	474	6.04733087
hækken	sg	h_E-	474	6.04733087
hække	pl	h_E-	474	6.04733087
katten	sg	k_a-	712	7.35681271
katte	pl	k_a-	712	7.35681271
kokken	sg	k_6-	1387	7.90602386
kokke	pl	k_6-	1387	7.90602386
kosten	sg	k_O-	357	5.61489509
koste	pl	k_O-	357	5.61489509
kroppen	sg	k_R_	881	5.13292263
kroppe	pl	k_R_	881	5.13292263
kusken	sg	k_u-	503	6.52171368
kuske	pl	k_u-	503	6.52171368
kvasten	sg	k_v_	378	4.79776645
kvaste	pl	k_v_	378	4.79776645
kvisten	sg	k_v_	378	4.79776645
kviste	pl	k_v_	378	4.79776645
kæppen	sg	k_E-	278	5.23334237
kæppe	pl	k_E-	278	5.23334237
lugten	sg	l_O-	263	5.39761324
lugte	pl	l_O-	263	5.39761324
pisken	sg	p_i-	161	5.91905437
piske	pl	p_i-	161	5.91905437
riggen	sg	R_E-	401	4.90814283
rigge	pl	R_E-	401	4.90814283
risten	sg	R_e-	1298	7.12761993
riste	pl	R_e-	1298	7.12761993
rokken	sg	R_6-	346	5.20762111
rokke	pl	R_6-	346	5.20762111
skakten	sg	s_g_	2157	8.57759934
skakte	pl	s_g_	2157	8.57759934
skatten	sg	s_g_	2157	8.57759934
skatte	pl	s_g_	2157	8.57759934
slottet	sg	s_l_	479	6.54212266
slotte	pl	s_l_	479	6.54212266
stakken	sg	s_d_	2535	8.06505169

stakke	pl	s_d_	2535	8.06505169
stokken	sg	s_d_	2535	8.06505169
stokke	pl	s_d_	2535	8.06505169
stubben	sg	s_d_	2535	8.06505169
stubbe	pl	s_d_	2535	8.06505169
sækken	sg	s_E-	809	6.74086383
sække	pl	s_E-	809	6.74086383
toppen	sg	t_6-	337	6.26791597
toppe	pl	t_6-	337	6.26791597
tragten	sg	t_R_	1156	7.79801359
tragte	pl	t_R_	1156	7.79801359
vasken	sg	v_a-	568	6.46133565
vaske	pl	v_a-	568	6.46133565
vægten	sg	v_E-	916	6.35708224
vægte	pl	v_E-	916	6.35708224

Experiment 1: Exploratory

Item	num	SAMPA	inhibition	Entropy decrease	SAMPA	inhibition	Entropy decrease
alken	sg	a-_l	1745	2.99257501	a-_l_?	196	2.40545391
alke	pl	a-_l	1745	2.99257501	a-_l	122	1.2745814
bolden	sg	b_6-	6796	4.1972285	b_6-_l_?	188	4.9835947
bolde	pl	b_6-	6796	4.1972285	b_6-_l	165	2.24240406
bænken	sg	b_E-	6813	4.21399436	b_E-_N_?	181	5.71990405
bænke	pl	b_E-	6813	4.21399436	b_E-_N	176	4.04940201
dampen	sg	d_A-	3101	2.82834443	d_A-_m_?	240	4.94924642
dampe	pl	d_A-	3101	2.82834443	d_A-_m	203	1.43074117
falken	sg	f_a-	7531	4.38459182	f_a-_l_?	445	4.58224084
falke	pl	f_a-	7531	4.38459182	f_a-_l	422	1.4681165
fjærten	sg	f_j_	7833	4.8535324	f_j_E-_6-_?	148	5.0370972
fjærte	pl	f_j_	7833	4.8535324	f_j_E-_6-	90	1.89219969
flaben	sg	f_l_	7367	3.19580102	f_l_a+_?	611	5.59703
flabe	pl	f_l_	7367	3.19580102	f_l_a+	606	4.53672604
fonten	sg	f_6-	7604	4.3815249	f_6-_n_?	350	2.42000815
fonte	pl	f_6-	7604	4.3815249	f_6-_n	371	4.25165172
greben	sg	g_R_	2287	1.90324015	g_R_e+_?	884	4.24531109
grebe	pl	g_R_	2287	1.90324015	g_R_e-	880	3.76373023
hanken	sg	h_A-	3590	2.81041573	h_A-_N_?	255	6.23174115
hanke	pl	h_A-	3590	2.81041573	h_A-_N	249	3.74441095

helten	sg	h_E-	3372	2.99482602	h_E-_l_?	471	4.88850378
helte	pl	h_E-	3372	2.99482602	h_E-_l	409	2.26779003
hingsten	sg	h_e-	3743	4.82785603	h_e-_N_?	102	4.21430085
hingste	pl	h_e-	3743	4.82785603	h_e-_N	97	2.48283064
hoben	sg	h_o+	3407	3.22257914	h_o+_?	438	5.81957774
hobe	pl	h_o+	3407	3.22257914	h_o+	2	0.01208194
hvalpen	sg	v_a-	3114	2.14699465	v_a-_l_?	563	5.87685245
hvalpe	pl	v_a-	3114	2.14699465	v_a-_l	395	0.67442319
kalken	sg	k_a-	7109	2.34147856	k_a-_l_?	705	5.85229756
kalke	pl	k_a-	7109	2.34147856	k_a-_l	674	3.43658219
kampen	sg	k_A-	7265	3.81397574	k_A-_m_?	516	4.54779776
kampe	pl	k_A-	7265	3.81397574	k_A-_m	465	3.00919805
kilten	sg	k_i-	7556	4.88054311	k_i-_l_?	264	4.81774816
kilte	pl	k_i-	7556	4.88054311	k_i-_l	263	3.82202071
kranken	sg	k_R_	6940	4.56536864	k_R_A-_N_?	880	5.13292263
kranke	pl	k_R_	6940	4.56536864	k_R_A-_N	880	5.13292263
kælken	sg	k_E-	7543	4.4649489	k_E-_l_?	272	3.01302331
kælke	pl	k_E-	7543	4.4649489	k_E-_l	267	3.09242465
lorten	sg	l_o-	4079	3.42878537	l_o-_6-_?	136	4.84393354
lorte	pl	l_o-	4079	3.42878537	l_o-_6-	136	4.74852184
læggen	sg	l_E+	3841	3.06637817	l_E+_?	347	1.90827751
lægge	pl	l_E+	3841	3.06637817	l_E+	29	0.06789882
milten	sg	m_i-	4738	2.75606699	m_i-_l_?	846	5.9139357
milte	pl	m_i-	4738	2.75606699	m_i-_l	843	4.96773391
pilken	sg	p_i-	4904	2.91844595	p_i-_l_?	159	5.06999994
pilke	pl	p_i-	4904	2.91844595	p_i-_l	152	3.3775275
pulken	sg	p_u-	4940	5.24980024	p_u-_l_?	114	0.48013712
pulke	pl	p_u-	4940	5.24980024	p_u-_l	107	0.86802512
salten*	sg	s_a-	12276	4.25735639	s_a-_l_?	309	3.05385186
salte	pl	s_a-	12276	4.25735639	s_a-_l	357	1.76461306
skabet	sg	s_g_	10526	1.94377202	s_g_a+_?	2147	7.37501654
skabe	pl	s_g_	10526	1.94377202	s_g_a+	2084	4.65078291
skalken	sg	s_g_	10526	1.94377202	s_g_a-_l_?	2156	8.57759934
skalke	pl	s_g_	10526	1.94377202	s_g_a-_l	2149	7.18236973
skalpen	sg	s_g_	10526	1.94377202	s_g_a-_l_?	2156	8.57759934
skalpe	pl	s_g_	10526	1.94377202	s_g_a-_l	2149	7.18236973
skanken	sg	s_g_	10526	1.94377202	s_g_A-_N_?	2156	8.57759934

skanke	pl	s_g_	10526	1.94377202	s_g_A-_N	2155	7.63193904
skiltet	sg	s_g_	10526	1.94377202	s_g_e-_l_?	2138	6.09062937
skilte	pl	s_g_	10526	1.94377202	s_g_e-_l	2135	5.17858749
skænken	sg	s_g_	10526	1.94377202	s_g_E-_N_?	2156	8.57759934
skænke	pl	s_g_	10526	1.94377202	s_g_E-_N	2155	7.7409586
staben	sg	s_d_	10148	2.45631967	s_d_a+_?	2295	3.11260301
stabe	pl	s_d_	10148	2.45631967	s_d_a+	2497	5.55870583
stanken	sg	s_d_	10148	2.45631967	s_d_A-_N_?	2533	7.26133577
stanke	pl	s_d_	10148	2.45631967	s_d_A-_N	2530	6.51376092
stilken	sg	s_d_	10148	2.45631967	s_d_e-_l_?	2534	8.06505169
stilke	pl	s_d_	10148	2.45631967	s_d_e-_l	2525	5.85163352
svampen	sg	s_v_	12290	3.80722328	s_v_A-_m_?	392	6.71414808
svampe	pl	s_v_	12290	3.80722328	s_v_A-_m	378	5.08512915
sumpen	sg	s_O-	12499	4.93561936	s_O-_m_?	183	5.585752
sumpe	pl	s_O-	12499	4.93561936	s_O-_m	173	3.03994884
tolken	sg	t_6-	4782	2.88327767	t_6-_l_?	334	5.00116102
tolke	pl	t_6-	4782	2.88327767	t_6-_l	290	2.28192952
teltet	sg	t_E-	4651	2.6090135	t_E-_l_?	467	6.54218013
telte	pl	t_E-	4651	2.6090135	t_E-_l	450	3.8917916
ulken	sg	u-_l	1629	4.13978984	u-_l_?	33	2.55873371
ulke	pl	u-_l	1629	4.13978984	u-_l	8	0.37019085
væggen	sg	v_E+	3603	4.80995552	v_E+_?	63	1.69432058
vægge	pl	v_E+	3603	4.80995552	v_E+	18	0.21267722

Appendix D

Experiment 1: Continuations

```
            Df Sum Sq Mean Sq F value Pr(>F)
w_cont_hilo  1   7.43   7.430   5.745 0.0189 *
Residuals   78 100.87   1.293
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = w_C3_136_204 ~ w_cont_log, data = sb_regression_all)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6072 -0.7598 -0.1133  0.6686  3.0600

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.9436     0.7077  -6.985 8.41e-10 ***
w_cont_log    0.9523     0.2614   3.644 0.000482 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.089 on 78 degrees of freedom
Multiple R-squared:  0.1454,  Adjusted R-squared:  0.1345
F-statistic: 13.28 on 1 and 78 DF,  p-value: 0.0004824

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: w_C3_136_204 ~ 1 + cont_log + num + sentence + (1 + cont_log |
    Subject) + (1 | Item)
   Data: data_stodbasis_mag
Control: lmerControl(optCtr = list(maxfun = 1e+09), optimizer = "nloptwrap",
calc.derivs = FALSE)

REML criterion at convergence: 34286.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3882 -0.6101 -0.0161  0.6117  9.2120

Random effects:
 Groups   Name              Variance  Std.Dev. Corr
 Item     (Intercept)        0.405226 0.6366
```

```
 Subject  (Intercept)        1.536388 1.2395
          cont_log  0.001814 0.0426   -0.94
 Residual                   51.089047 7.1477
Number of obs: 5054, groups:  Item, 40; Subject, 16

Fixed effects:
                   Estimate Std. Error        df t value Pr(>|t|)
(Intercept)         -4.8346     0.8891   33.8120  -5.437 4.72e-06 ***
cont_log             0.9129     0.3080   36.9651   2.964  0.00529 **
num                 -0.2067     0.1005 4998.7902  -2.055  0.03989 *
sentence             0.1231     0.1434   37.1063   0.858  0.39637
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) cont    num
cont_log    -0.934
num          0.001 -0.001
sentence     0.118 -0.127 -0.001
```

Experiment 1: Entropy

```
            Df Sum Sq Mean Sq F value  Pr(>F)
w_ent_hilo   1  13.93   13.93   11.52 0.00109 **
Residuals   78  94.36    1.21
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = w_C3_136_204 ~ w_ent, data = sb_regression_all)

Residuals:
     Min       1Q   Median       3Q      Max
-2.41587 -0.81475 -0.03421  0.65677  3.15205

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.35081    0.55999  -7.769 2.63e-11 ***
w_ent        0.31103    0.08728   3.563 0.000628 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.093 on 78 degrees of freedom
Multiple R-squared:   0.14,   Adjusted R-squared:  0.129
F-statistic:  12.7 on 1 and 78 DF,  p-value: 0.0006282

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
```

```
Formula: w_C3_136_204 ~ 1 + entropinum_wif2 + num + sentence + (1 +
entropinum_wif2 ||      Subject) + (1 | Item)
   Data: data_stodbasis_mag
Control: lmerControl(optCtr = list(maxfun = 1e+09), optimizer = "nloptwrap",
calc.derivs = FALSE)

REML criterion at convergence: 34288.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3863 -0.6082 -0.0188  0.6132  9.2187

Random effects:
 Groups    Name            Variance  Std.Dev.
 Item      (Intercept)      0.399303 0.63190
 Subject   entropinum_wif2  0.002964 0.05444
 Subject.1 (Intercept)      1.160498 1.07726
 Residual                  51.084544 7.14735
Number of obs: 5054, groups:  Item, 40; Subject, 16

Fixed effects:
                 Estimate Std. Error        df t value Pr(>|t|)
(Intercept)       -4.3180     0.7034   41.9322  -6.139 2.52e-07 ***
entropinum_wif2    0.3064     0.1022   34.0762   2.999  0.00504 **
num               -0.2067     0.1005 4996.2441  -2.056  0.03986 *
sentence           0.1635     0.1418   37.1334   1.153  0.25618
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) entr_2 num
entrpnm_wf2 -0.894
num          0.001 -0.001
sentence     0.029 -0.032 -0.001
```

Experiment 2: Continuations

C3

```
             Df Sum Sq Mean Sq F value Pr(>F)
w_cont_hilo  1  10.31  10.311   5.819 0.0183 *
Residuals   76 134.66   1.772
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cz

```
             Df Sum Sq Mean Sq F value Pr(>F)
w_cont_hilo  1  16.26  16.260   5.895 0.0175 *
Residuals   76 209.63   2.758
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

  Call:
lm(formula = w_C3_136_204 ~ w_cont_log, data = nsb_regression_all)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8416 -0.9248  0.0438  1.0810  2.8556

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.4376     1.0665  -4.161 8.27e-05 ***
w_cont_log    0.6339     0.3896   1.627    0.108
Signif. Codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.358 on 76 degrees of freedom
Multiple R-squared:  0.03367, Adjusted R-squared:  0.02096
F-statistic: 2.648 on 1 and 76 DF,  p-value: 0.1078

   Call:
lm(formula = w_Cz_136_204 ~ w_cont_log, data = nsb_regression_all)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2913 -0.9098  0.0969  0.9562  3.7854

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.7652     1.3418  -3.551 0.000662 ***
w_cont_log    0.5855     0.4901   1.195 0.235962
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.708 on 76 degrees of freedom
Multiple R-squared:  0.01843, Adjusted R-squared:  0.005515
F-statistic: 1.427 on 1 and 76 DF,  p-value: 0.236

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: C3_136_204 ~ 1 + w_cont_log + sentence + num + (1 + w_cont_log |
Subject) + (1 | Item)
   Data: data_nonstodbasis_mag
Control: lmerControl(optCtr = list(maxfun = 1e+09), optimizer = "nloptwrap",
calc.derivs = FALSE)

REML criterion at convergence: 16727

Scaled residuals:
```

```
    Min      1Q  Median      3Q     Max 
-4.7411 -0.6042 -0.0138  0.6202  5.5875 

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Item     (Intercept)  0.12985 0.3603       
 Subject  (Intercept)  0.48064 0.6933       
          w_cont_log   0.05239 0.2289   1.00
 Residual             47.96261 6.9255       
Number of obs: 2488, groups:  Item, 40; Subject, 16

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)    
(Intercept)   -3.9423     1.0159   37.7776  -3.881 0.000404 ***
w_cont_log     0.4500     0.3705   39.4076   1.215 0.231677    
sentence       0.4559     0.1506   35.9308   3.028 0.004538 ** 
num           -0.4142     0.1388 2432.8813  -2.983 0.002882 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
           (Intr) w_cnt_ sentence
w_cont_log -0.936                
sentence    0.048 -0.052         
num         0.000  0.000  0.001  
```

Experiment 2: Entropy

C3

```
             Df Sum Sq Mean Sq F value Pr(>F)  
w_ent_hilo   1   9.09   9.093   5.086  0.027 *
Residuals   76 135.88   1.788                 
```

Cz

```
Signif. Codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
             Df Sum Sq Mean Sq F value Pr(>F)  
w_ent_hilo   1   14.3  14.296   5.135 0.0263 *
Residuals   76  211.6   2.784                 
Signif. Codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

   Call:
lm(formula = w_C3_136_204 ~ w_ent, data = nsb_regression_all)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8637 -0.8784 -0.0177  0.8643  2.8222 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
```

```
(Intercept)  -3.9765     0.8565  -4.643 1.41e-05 ***
w_ent         0.2002     0.1342   1.491     0.14
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.361 on 76 degrees of freedom
Multiple R-squared:  0.02843, Adjusted R-squared:  0.01564
F-statistic: 2.224 on 1 and 76 DF,  p-value: 0.14

  Call:
lm(formula = w_Cz_136_204 ~ w_ent, data = nsb_regression_all)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0289 -0.9010  0.1048  0.9327  3.7245

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.6204     1.0715  -4.312  4.8e-05 ***
w_ent         0.2297     0.1680   1.367    0.176
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.703 on 76 degrees of freedom
Multiple R-squared:  0.02401, Adjusted R-squared:  0.01117
F-statistic:  1.87 on 1 and 76 DF,  p-value: 0.1755

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: C3_136_204 ~ 1 + w_ent + sentence + num + (1 + w_ent | Subject) +
(1 | Item)
   Data: data_nonstodbasis_mag
Control: lmerControl(optCtr = list(maxfun = 1e+09), optimizer = "nloptwrap",
calc.derivs = FALSE)

REML criterion at convergence: 17125.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7231 -0.6057 -0.0216  0.6198  5.6498

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Item     (Intercept)  0.12576 0.3546
 Subject  (Intercept)  0.08063 0.2840
          w_ent        0.02718 0.1649   1.00
 Residual             47.83759 6.9165
Number of obs: 2548, groups:  Item, 40; Subject, 16
```

```
Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)    
(Intercept)   -3.4902     0.6288   35.6987  -5.551 2.84e-06 ***
w_ent          0.1236     0.1074   27.0933   1.151  0.25985    
sentence       0.4557     0.1488   37.0145   3.063  0.00407 ** 
num           -0.4175     0.1370 2492.3381  -3.047  0.00234 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
       (Intr) w_ent  sentence
w_ent  -0.848                
sebtence0.096 -0.092         
num    -0.001  0.001  0.001  
```

Experiment 1, exploratory: Entropy change, word onset

```
               Df Sum Sq Mean Sq F value  Pr(>F)   
w_entfall_hilo  1  13.31  13.308   10.93 0.00143 **
Residuals      78  94.99   1.218                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:
lm(formula = w_C3_136_204 ~ w_entfall_p_w, data = sb_regression_all)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.80620 -0.71624 -0.08638  0.67969  2.65559 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -0.9827     0.4025  -2.441 0.016895 *  
w_entfall_p_w  -0.4275     0.1155  -3.702 0.000397 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.087 on 78 degrees of freedom
Multiple R-squared:  0.1495,  Adjusted R-squared:  0.1385 
F-statistic: 13.71 on 1 and 78 DF,  p-value: 0.0003969

Linear mixed model fit by REML. t-tests use Satterthwaite's method
[lmerModLmerTest
]
Formula:
```

```
w_C3_136_204 ~ 1 + w_ent_change_p_w + num + sentence + (1 + w_ent_change_p_w
|      Subject) + (1 | Item)
   Data: data_stodbasis_mag
Control: lmerControl(optCtr = list(maxfun = 1e+09), optimizer = "nloptwrap",
    calc.derivs = FALSE)

REML criterion at convergence: 34287.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3947 -0.6085 -0.0193  0.6110  9.2339

Random effects:
 Groups   Name             Variance  Std.Dev. Corr
 Item     (Intercept)      3.875e-01 0.62247
 Subject  (Intercept)      1.083e+00 1.04047
          w_ent_change_p_w 7.376e-04 0.02716  1.00
 Residual                  5.109e+01 7.14767
Number of obs: 5054, groups:  Item, 40; Subject, 16

Fixed effects:
                  Estimate Std. Error        df t value Pr(>|t|)
(Intercept)        -1.0041     0.5344   30.2893  -1.879  0.06990 .
w_ent_change_p_w   -0.4200     0.1341   37.0659  -3.131  0.00339 **
num                -0.2068     0.1005 4998.8494  -2.056  0.03980 *
sentence            0.1523     0.1409   37.1235   1.080  0.28693
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr)  ent___ num
w_nt_chng__ -0.807
num         -0.001  0.001
sentence    -0.047  0.056 -0.001
```

Experiment 1, exploratory: inhibition, stød/non-stød onset

```
                Df Sum Sq Mean Sq F value  Pr(>F)
s_inhibited_hilo 1  15.79  15.794   9.732 0.00254 **
Residuals       78 126.58   1.623
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = s_FC2_136_204 ~ s_inhibited_log, data = sb_regression_all)

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-3.10625 -0.94285  0.08318  1.09198  2.19462

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.5071     0.6532  -2.307   0.0237 *
s_inhibited_log -0.5211     0.2466  -2.113   0.0378 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.314 on 78 degrees of freedom
Multiple R-squared:  0.05414, Adjusted R-squared:  0.04201
F-statistic: 4.465 on 1 and 78 DF,  p-value: 0.0378

Linear mixed model fit by REML. t-tests use Satterthwaite's method
['lmerModLmerTest']
Formula: s_FC2_136_204 ~ 1 + s_inhibited_log + sentence + num + stod + (1 +
    s_ruledout_log | Subject) + (1 + s_inhibited_log | Item)
   Data: data_stodbasis_mag
Control: lmerControl(optCtr = list(maxfun = 1e+09), optimizer = "nloptwrap",
calc.derivs = FALSE)

REML criterion at convergence: 35754

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9728 -0.6364 -0.0018  0.6384  5.3404

Random effects:
 Groups   Name            Variance Std.Dev. Corr
 Item     (Intercept)      4.7400  2.1771
          s_inhibited_log 0.2991  0.5469   -1.00
 Subject  (Intercept)      5.1946  2.2792
          s_inhibited_log 0.1312  0.3622   -1.00
 Residual                 67.2935  8.2033
Number of obs: 5064, groups:  Item, 40; Subject, 16

Fixed effects:
                Estimate Std. Error        df t value Pr(>|t|)
(Intercept)      -1.2617     1.0204   28.2681  -1.236  0.22646
s_inhibited_log  -0.6054     0.3104   40.7905  -1.950  0.05806 .
sentence         -0.5333     0.1609   34.4318  -3.314  0.00217 **
num              -0.1694     0.1153 5009.2056  -1.470  0.14175
stod             -0.1328     0.1164 4946.6677  -1.141  0.25374
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Correlation of Fixed Effects:
            (Intr) s_inh_ sentence num
s_inhibi_lg -0.942
```

```
sentence       0.040 -0.056
num            0.000  0.000 -0.001
stod          -0.084  0.104 -0.024  0.000
```