Processing the L2 comprehension process:

Testing Processability Theory's predictions in an ERP study of adult learners of L2 Swedish

by Ariane Senécal

Under the supervision of Gisela Håkansson and Mikael Roll

Thesis submitted for approval for the

Master's Programme in General Linguistics

Centre for Languages and Literature

Lund University

24 May 2011

Abstract

This small-scale experiment was a first attempt at testing the theoretical framework of Processability Theory (PT), a well-established theory of Second Language Acquisition (SLA), against data from online brain imaging technology (Electroencephalography – EEG). Event Related Potentials (ERPs) were extracted for 15 advanced beginners/intermediate-level learners of Swedish who were presented with grammatical and ungrammatical sentences, each containing one of two grammatical forms evaluated in PT as belonging to two different stages of acquisition (attributive number agreement vs. predicative number agreement). In accordance with PT predictions and earlier studies of more advanced L2 learners, the hypothesis was that an increased LAN and/or P600 might be observed in the attributive agreement conditions, while more weakly increased or no increased components might show in the predicative agreement conditions. Although the expected effects were not observed, results do indicate significant effects of processability level (agreement type) on ERP distribution among electrodes, indicating that this factor does have an effect on the comprehension/perception processing of even less-advanced L2 learners. The observation of an increased frontal positivity beginning at approximately 500 ms for both attributive agreement conditions could be indicating that some integration of this type of agreement is being accomplished online; as predicative agreement sentences do not elicit this response, results appear to preliminarily be supporting PT predictions regarding an order of acquisition based on increasing complexity of structures, although further experimentation would be required to back this claim. Potential reasons for the fact the initial hypotheses were not confirmed as well as suggestions for further research, namely more ERP studies of SLA, are also discussed.

Keywords: Neurolinguistics, Event-Related Potentials, Second Language Acquisition, Processability Theory, Attributive Agreement, Predicative Agreement, Swedish as a Second Language

1. Introduction

Comprehension and perception remain lesser-explored areas in studies of Second Language Acquisition (SLA), although research interest has greatly increased in recent years (see thematic issue of *Lingua*, Hendriks and Koster 2010). This could be partially attributed to the relative difficulty of evaluating these less evident facets of language learning. The present project attempted a first foray into testing the theoretical framework of a well-established theory of SLA, Processability Theory (Pienemann 1998, 2003, 2005, 2007), with the imaging possibilities offered by Electroencephalography (EEG) and the study of Event Related Potentials (ERPs). This experiment therefore firstly aimed to evaluate to what extent some particular structures may be acquired and grammaticalized in the brain of adult L2 learners, and secondly, to verify whether the capability for recognizing the correctness of certain grammatical forms online could follow set and ordered discrete stages as predicted by PT research on production - therefore testing PT as a tool for analyzing online processing of a L2. By looking at ERP responses to two types of adjective agreement (attributive vs. predicative) evaluated in PT as belonging to two different stages of acquisition of L2 Swedish, the present experimental project proposed to explore whether there are differences and whether these correspond to the predicted PT stages of acquisition.

1.1 Research Question

The main question that guided this research can be phrased as follows: can online brain imaging technology reveal discrete grammatical stages in beginner and intermediate learners' processing of a L2? Furthermore, can PT be used as a relevant tool for studies of L2 perception and comprehension, more specifically for the analysis of ERP experimental results dealing with grammatical aspects of L2 acquisition?

1.2 Background and review of previous publications

1.2.1 Processability Theory

Processability theory (PT, Pienemann 1998) is a SLA theory that assumes a universal, necessarily hierarchical order of acquisition in interlanguage, with L2 learners assumed to inevitably build on previously acquired (processed) structures to learn and produce new ones, this according to a fixed developmental schedule. Language acquisition is seen as the acquisition of skills for language processing (*ibid.*). Information-matching procedures are used to gradually build phrases of increasing complexity. According to PT, though individual rate of acquisition and interlanguages may vary, the sequence of acquisition remains the same.[1]

The grammatical description used in PT has been modeled on Lexical Functional Grammar (see Bresnan 2001); the processability hierarchy is defined in terms of Levelt's (1989) Model of speech production. According to Levelt (1989), a native speaker's production of utterances consists of a set of implicatively related psycholinguistic procedures: (1) lemma access, (2) categorical procedure, (3) phrasal procedure, and (4) S-procedure. As the earlier procedures constitute a necessary input for later procedures, they are implicatively related in the following way: $4 \supset 3 \supset 2 \supset 1$. For instance, execution of a phrasal procedure (level 3) is possible only if a word has been retrieved from the mental lexicon (level 1), and its syntactic category has been accessed by the system (level 2).

While Levelt's Speech Production model (1989) deals with the oral production of a L1, Pienemann (1998, 2003, 2005) argues that the task of a L2 learner is to acquire the very same processing procedures. Pienemann includes both production and comprehension under the scope of PT: "[t]he logic underlying processability theory (PT) is the following: at any stage of development the learner can produce and comprehend only those L2 linguistic forms which the current state of the language processor can manage" (Pienemann 2003:686).

PT possesses great descriptive power and it has been tested for many different languages, proving itself as a reliable model of the gradual learning and processing of a L2 (Pienemann 2005). Though the process is assumed to be universal, language-specific grammatical features can belong to belong to different stages. Swedish, which was one of the first languages to be extensively examined under the PT framework, bears overt marks of morphological and syntactic procedures and is thus an interesting object of study to test the model's hypotheses. The processing prerequisite levels as applied to Swedish as a second/foreign language (SSL/SFL) were devised by Pienemann and Håkansson (1999) using a broad and in-depth survey of different studies of L2 Swedish. First exposing the core of the

[1] In the original formulation of PT, the order is as follows: 1) no procedure; 2) category procedure; 3) noun phrase procedure; 4) verb phrase procedure; 5) sentence procedure; and 6) subordinate clause procedure (Pienemann 1998: 7).

theory (including 4 basic premises[2]), the article reviews 14 studies of SFL/SSL to check findings against PT predictions. The researchers conclude that all the results possibly pertaining to the processability hierarchy confirm the model's sound bases. Consequently, they are able to apply PT's processing procedures to the morphology of Swedish. It is according to their table (1999: 398; see Table 1) that we can select the two main conditions we use for the present experiment: we can observe that NP marking (which includes attributive agreement) is seen to belong to the Phrasal procedure stage (level 3) while the adjective agreement in predicative constructions is seen as belonging to the S-procedure, which deals with interphrasal information (level 4).

Table 1. Processing procedures applied to Swedish morphology

Processing procedures	L2 structure	Swedish morphology
Clause boundary	Main and subordinate clause	-
S-procedure or word order rules	Interphrasal information	Adjective agreement in predicative constructions
Phrasal procedure	Phrasal information	Definiteness agreement, markings in NPs, compound tense markings in VPs
Category procedure	Lexical morphemes	Plural, definiteness on nouns, past or present tense on verbs
Word or lemma access	Words	Invariant forms

(From Pienemann and Håkansson 1999)

Though not originally designed for this use, being based partly on speech production models, for example (see discussion of Levelt 1989 above), PT should logically be applicable to the evaluation of perception and comprehension as well as production - that is, there is nothing in the theory that is an inherent obstacle to its further application. However, PT has not yet been purposefully used to examine this other aspect of language acquisition to any great extent.

One of the few such works known to date is that of Håkansson and Yager (under review). The study used picture identification and picture description to test participants' ability to comprehend and produce certain morphosyntactic structures, namely attributive vs. predicative adjective agreement. The researchers aimed to compare comprehension and production ability to shed some light on the relation between these two aspects of SLA as well as to verify the application of PT to the results of a comprehension task. To achieve this, Håkansson and Yager tested two groups of participants, one consisting of beginner learners of

[2] 1) Processing components e.g. phrase-building procedures are generally autonomous and automatic; 2) Processing is incremental; 3) Output is linear although the underlying mapping may not be; and 4) Processing of grammar requires access to a grammatical memory store (Pienemann and Håkansson 1999: 386-387).

Swedish and the other of advanced learners, in experimental conditions that sought to recreate a communication setting that was as natural as possible[3]. Using picture material based on Glahn *et al.* (2001), participants were tested first on production ability, then on comprehension. In the first case, they had to locate a picture among many and describe it in order to answer a question (e.g. to test attributive agreement, 'What stands in front of the little red house? - Two brown dogs'; to test predicative agreement, 'What color are the small cups? - They are brown'[4]), while in the second learners had to identify the correct picture only relying on cues from the adjective (e.g. to test attributive agreement, 'Point to "black sheep"'; to test predicative agreement, 'Point to "his sheep is/are black"'[5]). For both tasks, the researchers found that an implicational order could be drawn by which processing of attributive agreement precedes processing of predicative agreement (i.e. no learner was able to process predicative agreement without having processed attributive agreement), and likewise that comprehension precedes production (i.e. no participant was able to produce a structure without showing comprehension of that structure). Håkansson and Yager report that their experiment confirms the predictions of PT regarding order of acquisition and that the discrepancy in comprehension and production results points to different processes for these two aspects. They suggest PT as an explanatory model to other studies of SLA, including ERP studies, an idea we take up in the present project.

Another study that has looked at comprehension using the PT model is that of Keatinge and Keßler (2009). In their investigation of the acquisition of the passive voice in L2 English, the researchers tested learners of various language backgrounds in perception and production tasks. They hypothesized a universal gradual development from perception to production parallel to the hierarchy of stages set forth in PT. A picture-pointing task was used to evaluate comprehension, in which participants listened to sentences in both active and passive voice and should select the corresponding picture (e.g. 'The blue fish eats the green fish' and 'The green fish is eaten by the blue fish'). Four production task were used, including a film description task (the 'Fish film'), sentence completion tasks and a storytelling task. Keatinge and Keßler conclude that perception of passive forms does precede production of passive forms, and that there is gradual development of what they call 'Pseudo Passive' forms in the

---

[3] The authors indicate natural communicative situations offer optimal data for analysis under the PT framework

[4] In Swedish: '*Vad finns framför den lilla röda huset? - Två bruna hundar*'; '*Vilken färg är de små kopparna? - De är bruna*'.

[5] In Swedish, both the noun and verb remain the same in singular or plural, but the adjective takes an 'a' in the plural: '*Svart/a får*'; '*Hans får är svart/a*'.

interlanguage, which show early processing of the grammatical notion although they are not properly formed (e.g. *'The castle is visit the family' rather than 'The castle is visited by the family'). Again, the PT framework makes good use of implicational tables, which show the hierarchy of stages in the development and acquisition of a second language; in Keatinge and Keßler's as well as Håkansson and Yager's studies, such tables are useful in representing the set order of distinct grammatical stages.

The two known studies which deal with comprehension/perception under PT, described in the above paragraphs, use offline methods (picture-pointing) to evaluate participants' levels. As far as we know, no online studies had been conducted prior to this one: this was the first experiment combining the framework of PT with brain-imaging technology. Pienemann puts forth that a speaker "needs to develop procedures that can handle the job of storing and comparing grammatical information; this way, speakers can learn to decide which sentences are grammatically acceptable and which aren't" (Pienemann 2007: 16-17). He claims that a learner who has not processed the appropriate procedure will not detect a grammatical error related to that procedure. The present experiment sought to verify whether this could indeed be the case, and whether this process could be observed online using ERP responses.

1.2.2. ERP experiments and L2 research

The Event-Related Potentials technique offers very accurate temporal measurements of the brain's electrical activity, and thus allows to follow online activation patterns by time-locking encephalography readings to the onset of stimuli. It is well-suited to studies of comprehension, as it can show very small and rapid changes in scalp voltage measurements as they happen.

While more and more studies are looking at comprehension in L2 learners, including an increasing number of ERP experiments, many seem to focus mainly on questions of age of acquisition/critical learning period, comparison to native speakers, or cross-linguistic influence (Tokowicz and MacWhinney 2005, Sabourin and Stowe 2008, Gillon Dowens *et al.* 2010, Steinhauer *et al.* 2009, van Hell, J.G. and N. Tokowicz 2010 – see Kotz's 2009 review for more examples and Sabourin 2009 for a general introduction to the field). While interesting topics of investigation, these are not concerns under the PT model: focused on interlanguage, PT does not take native-like competence and performance as a benchmark, nor does it seek to explain issues of innateness or transfer. As Pienemann and Håkansson put it plainly, “it is the sole objective of processability theory to determine the sequence in which procedural skills develop in the learner” (1999: 386).

Still, many ERP experiments have contributed to advancing understanding of L2 acquisition, and many were relevant to the development of the present study. Gillon Dowens *et al.* 2010’s research investigated the online brain responses of native English speakers who are proficient late learners of L2 Spanish and compared their ERP results to those of a native Spanish control group. They tested violations to number agreement (a grammatical feature present in English) and gender agreement (a feature not present in English) in two sentence positions, at the beginning or at the middle of the sentence (corresponding to attributive vs. predicative agreement). For example, a sentence such as ‘*El suelo está plano y bien acabado*’ (The floor (m.sing.) is flat and well-finished) was modified to include attributive number agreement violations, ‘*Los suelo está plano y bien acabado*’ (‘The (m.plur.) floor’), attributive gender agreement violations, ‘*La suelo está plano y bien acabado*’ (‘The (f.sing.) floor’), predicative number agreement violations, ‘*El suelo está planos y bien acabado*’ (‘The floor is flat (m.plur.)’), and predicative gender agreement violations, ‘*El suelo está plana y bien acabado*’ (‘The floor is flat (f.sing.)’). Gillon Dowens *et al.*’s results show that for agreement violations, patterns of increased left-anterior negativity (LAN, a response to morphosyntactic mismatch) followed by a syntactic positive shift (P600, a response to syntactic violations) are observed in learners as well as native speakers, which leads the researchers to conclude that proficient L2 speakers can have ERP readings ‘qualitatively’ similar to those of native speakers even if the former have later ages of acquisition. However, results did vary according to conditions: only in the first sentence position (i.e. attributive agreement) did learners mark an increased LAN, and there were latency and amplitude differences in responses between gender and number agreement conditions. This study greatly influenced the design and hypotheses of the current experiment, although elements of the PT framework were included in the description of conditions.

One quite interesting recent article is that of McLaughlin *et al.* (2010), which reviewed studies of adult novice L2 learners to conclude that different, robust stages of acquisition can be observed in learners, which can even be used to identify sub-groups of learners. Direct comparisons to the present study are not possible, due to the different nature of the experiments and materials used. However it would appear that PT could offer some explanations as to why some learners appear to have a “discontinuous pattern over time”

(: 138), if the responses were to elements belonging to different acquisition stages, for example. This is a perfect example of the usefulness a well-founded theory of SLA could play in completing the analysis of L2 ERP studies, something we wished to consider in the current experimental design.

Another interesting paper is that of Steinhauer *et al.* (2009). The researchers set out to review ERP studies of SLA to look into possible age of acquisition-related effects. They conclude by arguing against a critical period for the acquisition of morphosyntactic structures: they suggest that proficiency is more of a determining factor than age of acquisition. Furthermore, they propose that "with increasing proficiency, L2 learners' brain activation profiles typically approximate that of native speakers **in a systematic way**[6]" (: 30). They add that although the timing of these 'transitions' varies between learner, "a prototypical second language learner will be more likely to pass through these putative stages than not" (*ibid*.). The authors suggest their own developmental schedule based on expected ERP responses (see Discussion below).

One research paradigm put forth in a paper by Osterhout *et al.* (2006) seeks to use EEG technique to identify the level of exposure required for adult novice learners to grammaticalize a L2 acquired in a classroom setting. With their suggested paradigm, the researchers aim to reduce variability as much as possible (a notably difficult issue in L2 studies, including the present one). They suggest studying learners longitudinally to explore processes underlying L2 comprehension. Although the current project cannot incorporate many of their suggestions, some elements are worth taking into account. For one, their working definition of 'learning' as "specifically the incorporation of L2 knowledge into the learner's online, real-time language processing system" (: 223) suits our needs. Another significant point relates to judgments of adults as having limited L2 acquisition potential: Osterhout *et al.* propose that this point of view is due to adults being assessed on their production, which often does contain grammatical errors. They note, however, that "one general rule about L1 learning is that a learner's ability to understand the language develops in advance of his or her ability to produce it. It seems likely that this maxim also applies to L2 learners" (: 224). If that is the case, it is certainly important to evaluate comprehension when evaluating SLA stages as suggested in PT, for example, which further confirms the relevance of the currently proposed experiment.

[6] Bold added for emphasis.

1.3 Hypothesis

It was hypothesized that there would be a difference between the waveforms following the onset of attributive adjective disagreement and those following the onset of a predicative adjective disagreement. More precisely, it was hypothesized that participants would show LAN and P600 (well-documented ERP-components elicited in response to syntactic violations, see Luck 2005) responses to attributive adjective agreement violations, but that the response would possibly be absent in predicative agreement violations, as learners may not have processed this latter form. It was also considered possible that more proficient participants would show a larger P600 response to the across-phrase disagreement due to increased processing demands (see Barber and Carreiras' 2005 study of Spanish native speakers). Considering the results exposed in McLaughlin *et al.* (2010), it was also considered possible that no LAN would be recorded in any of the contexts, as there is some debate as to whether these are always present, or that in less proficient learners a N400 might be elicited instead of the expected P600 due to difficulties in lexical retrieval.

2. Method

2.1 Participants

As PT does not consider L1 relevant to order of acquisition, participants were not recruited from any particular language group. 15 participants (9 female, 6 male) took part in this experiment. All were right-handed[7] learners of Swedish between the ages of 20 and 40 (mean age 27.2), with no history of severe brain damage[8], self-evaluated[9] as being advanced beginners or intermediate-level learners of Swedish. Native languages were distributed as follows: 4 German (3 from Germany and 1 from Austria), 3 Russian (of which 1 Russian and Ukrainian and 1 Russian and Belarusian), 2 French (1 from Canada and 1 from France), 1 Turkish, 1 Korean, 1 Nepali, 1 Italian, 1 Finnish, and 1 English (from the U.S.A.) All had learned other foreign languages before learning Swedish (all knew English, and all but one indicated they had some level of proficiency in at least 4 languages).

[7] Two participants actually completed the questionnaire using their left hand – when questioned about this, both claimed to be ambidextrous but favor writing using their left hand. After consideration, their data was included in the study.

[8] Two participants reported having suffered from mild concussions in their childhood, but both assured that no permanent damage had been observed. After consideration, their data was included in the study.

[9] During the experimental setup, participants completed a questionnaire to gather data on their linguistic background and language use, as well as to allow them to do a self-evaluation of their language skills.

2.2 Materials

The project consisted of an experiment evaluating online responses (using EEG recording and ERP responses) to different conditions corresponding to different stages in L2 grammar acquisition as defined in PT studies of Swedish as a L2. The method resembled that used by Barber and Carreiras (2005) and Gillon Dowens, Vergara *et al.* (2010), who gave an excellent description of their experimental methods.

Participants were presented with visual material consisting of 5-word Swedish sentences containing within (PT Level 3; Pienemann and Håkansson 1999) or across phrase (PT Level 4; *ibid*) number agreement violations (attributive agreement vs. predicative agreement). Target adjectives came on the first or second slide to avoid any sentence wrap-up effects. To limit the number of variables under evaluation, only common gender regular nouns and adjectives were used to compose the sentences. It has been observed that language learners are exposed to words differently than native speakers and that therefore special attention and multiple sources should be considered when choosing vocabulary for research or experiments (Davidson *et al*. 2008): sentences were composed using the beginner's book in a popular Swedish learning series, *På svenska!* (Göransson and Parada 1997), as principal source of inspiration, and target words were verified for frequency using the Stockholm-Umeå Corpus (Ejerhed *at al.* 1992). As a control condition, participants were presented with properly grammatical sentences following the same basic structure. Four conditions (including two control conditions) were therefore tested. Note that all conditions presented singular and plural sentences in equal distribution:

Condition 1 (c1): Within-phrase (attributive) number agreement
e.g. 'En snäll hund bor här' / 'Flera snälla hundar bor här'
Condition 2 (c2): Within-phrase (attributive) number agreement violation
*e.g. *'En snälla hundar bor här' / *'Flera snäll hund bor här'*
Condition 3 (c3): Across-phrase (predicative) number agreement
e.g. 'En hund är snäll och dyr' / 'Hundar är snälla och dyra'
Condition 4 (c4): Across-phrase (predicative) number agreement violation
*e.g. *'En hund är snälla och dyr' / *'Hundar är snäll och dyra'*

The full list of sentences used in the experiment may be found at Appendix A.

2.3 Procedure

Using the stimulus presentation software E-Prime 2.0 (Schneider, Eschman and Zuccolotto 2002), 40 sentences were presented for each condition (20 for each of the singular and plural subconditions), for a total of 160 sentences, divided into 4 experimental blocks using pseudo-randomization. Sentences were presented in random sequencing within each block, and blocks were also presented in random order from one participant to the next. The sentences were presented as visual material on a 85 Hz, 19-inch computer screen (white size 18 Courier New font on black background), word by word, at a slightly slower than average reading speed (600 ms/slide). Interstitial blank slides divided each word (300 ms) and preceded the response (800 ms). Each trial block began with a 5000 ms fixation cross. Prior to the experiment participants were instructed that they would have to read the words carefully, composing the full sentence in their mind and paying close attention to grammar[10], and then answer a yes/no grammatical judgment question. During the experiment, they were prompted to accomplish this task when the screen requesting them to do so appeared after the last word of the sentence. They then should respond by pressing numerical keys 1 for a properly grammatical sentence (using their right index finger) or 2 for an ungrammatical sentence (using their right middle finger). They were informed that their response time would not be taken into consideration, and that they should favor accuracy over speed without dwelling on answers for too long. The next sentence began once the response had been logged, following a fixation cross (800 ms). A practice block of 10 questions preceded the actual experiment, during which the researcher remained in the room to answer questions or make adjustments. Afterwards, participants were left to complete the experiment on their own, in a quiet room. Participants were able to take a break between blocks, controlling the time of onset of the following block.

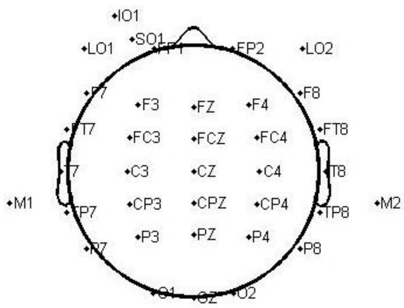
2.4 EEG Recordings

Participants were outfitted with an EASYCAP EEG Recording Cap, using a 36-electrode layout (30 scalp electrodes, namely O2, O1, OZ, PZ, P4, CP4, P8, C4, TP8, T8, P7, P3, CP3,

[10] In a pilot trial of the experiment, a participant had been instructed to try not to blink during the presentation of words, and to try to limit blinking to the periods in between sentences or during their response. However this participant expressed that it had been difficult to complete the task while remembering to blink as little as possible, i.e. that the effort required by remembering not to blink distracted from the experiment itself. Therefore, in further trials, participants were instead instructed to focus on reading each word carefully and concentrate on putting the sentence together in their mind. This strategy effectively aided in limiting the number of eye artifacts in the data while ensuring participants were attentive to the task.

CPZ, CZ, FC4, FT8, TP7, C3, FCZ, FZ, F4, F8, T7, FT7, FC3, F3, FP2, F7 and FP1, plus horizontal eye electrodes LO1-LO2 as well as vertical left-eye electrodes SO1-IO1 and mastoid electrodes M1-M2).

Figure 1. Electrode distribution



Impedance was maintained below 5 kΩ for all electrodes. The EEG was recorded continuously at a sampling rate of 500 Hz using the ACQUIRE application in Scan 4.5 (Compumedic Neuroscan 2009) with a SynAmps2 amplifier. An online band-pass filter of .05 and 70 Hz was applied. The data was referenced to a central, cap-mounted reference electrode during recording and an anterior ground electrode was used.

2.5 ERP Data Analysis

Using the EDIT application in Scan 4.5, participants' data was firstly down-sampled to a rate of 250 Hz, then re-referenced to both mastoids, and finally filtered using a low-pass band filter of 30 Hz. Further analysis was completed using the ERPLAB toolbox (Markley, Luck and Lopez-Calderon 2010) and EEGLAB (Delorme and Makeig 2004) in Matlab (The MathWorks 2009). ERPs were time-locked to the target words (i.e. the onset of the adjective that bore/should bear the mark of agreement) and epochs extended from -200 ms prior to target words to 800 ms after their onset. The period prior to the onset of the target word (-200 ms) was used as baseline. The EEGLAB Independent Component Analysis (ICA, Jung *et al.* 2000) was used to identify and remove eye artifacts. The ERPLAB artifact detection tool was further used to apply a +/- 100 µV simple voltage threshold. ERPs for each participant were averaged in ERPLAB, as was a grand average for all participants.

3. Results

3.1 Behavioral data

As participants were instructed to favor accuracy over speed, response times for not taken into consideration.

Although PT usually evaluates learners according to the emergence criterion (see Pienemann 2007 for the manner in which this criterion should be applied, and Håkansson and Yager for a discussion on adapting this criterion to comprehension analysis), this is obviously not feasible in the present experimental context. Therefore, a more traditional accuracy percentage summary is preferable for our current uses.

The accuracy for all participants was relatively high, at 83.2% overall (participant median for all conditions: 91.25%). Condition 1 had an average accuracy percentage of 80.5% (median: 95%); condition 2, 83.3% (median: 87.5%); condition 3, 90.7% (median: 90%); and condition 4, 78.3% (median: 95%).

Table 2. Accuracy percentages for each participant by condition

Participant #	Attributive agreement (c1)	Attributive agreement violation (c2)	Predicative agreement (c3)	Predicative agreement violation (c4)
1	0.85	0.65	0.825	0.875
2	0.95	1	0.975	1
3	0.325	0.775	0.825	0.25
4	1	1	1	0.975
5	0.475	0.675	0.65	0.6
6	0.875	0.825	0.9	1
7	1	0.95	1	1
8	1	0.975	0.975	0.975
9	0.425	0.575	0.9	0.15
10	0.975	0.875	0.9	0.9
11	0.975	0.825	0.875	0.925
12	1	1	0.975	0.95
13	0.875	0.925	0.95	0.975
14	0.35	0.525	0.85	0.2
15	1	0.925	1	0.975

We can observe that 7 participants had their best accuracy rate for condition 1; 5 did best for condition 3, while conditions 2 and 4 were the stronger accuracy for 4 participants each.

However, as the behavioral task demanded grammaticality judgments, the conditions can logically be combined according to agreement type: the average accuracy percentage is then of 81.9% for attributive agreement (median: 90%) and 84.5% for predicative agreement (median: 95%). A majority of participants (9) have better accuracy percentages for predicative conditions.

The standard practice in PT is to present data in implicational scales, which reflect the gradual acquisition of the target language. Table 3 presents the original accuracy percentages according to an implicational order matching PT predictions, i.e. attributive agreement acquisition should precede predicative agreement acquisition. The table does not represent a large scalar difference in accuracy rate for violation conditions, as these appear to yield quite similar results. In any case, a gap is identified as one participant receives a 100% accuracy rating on the predicative conditions without having 100% on the attributive conditions.

Table 3. Implicational scale of accuracy results as should correspond to PT predictions

Participant #	Attributive agreement (c1 and c2)	Predicative agreement (c3 and c4)
4	1	0.9875
12	1	0.9625
8	0.9875	0.975
2	0.975	0.9875
7	0.975	1
15	0.9625	0.9875
10	0.925	0.9
13	0.9	0.9625
11	0.9	0.9
6	0.85	0.95
1	0.75	0.85
5	0.575	0.625
3	0.55	0.5375
9	0.5	0.525
14	0.4375	0.525

A 100% accuracy requirement being very high (for one, a few participants reported after the experiment they had pressed the wrong key by mistake on at least one occasion), it is worth considering setting a threshold at which the condition is considered to be answered correctly. If a threshold is applied through which all scores above 90% are calculated as having a 100% (Table 4), result still appear rather similar for both sets of conditions, but if anything, results appear to show that predicative agreement precedes attributive agreement, as no participant had a 100% accuracy rating on attributive agreement without having 100% accuracy for predicative as well, whereas the opposite is not true. This is contrary to PT predictions, according to which attributive agreement acquisition should precede predicative agreement acquisition.

Table 4. Implicational scale of accuracy results where 90+% = 100%

Participant #	Attributive agreement (c1 and c2)	Predicative agreement (c3 and c4)
2	1	1
4	1	1
7	1	1
8	1	1
10	1	1
11	1	1
12	1	1
13	1	1
15	1	1
6	0.85	1
1	0.75	0.85
5	0.575	0.625
3	0.55	0.5375
9	0.5	0.525
14	0.4375	0.525

The same happens if we take an approach setting a simple 50% threshold by which all results below 50% are considered null and all results above are considered 100%: a similarly balanced table is established, with predicative agreement appearing to precede attributive agreement rather than the reverse.

3.2 ERP data

For all subjects, less than 3% (63 out of 2400) of trials were rejected using the +⁄- 100 µV threshold applied after the ICA.

Upon close visual inspection of the Grand Average ERPs for all 30 scalp electrodes for all 4 experimental conditions (Figure 2), it would appear that a small but noticeable negative peak can be observed mainly in the electrodes along the midline. These peaks, occurring at approximately 580 ms in condition 2 (attributive agreement violation), appear to signal a potential increased P600 component (a closer view is given in Figure 3).

Furthermore, a divergence between the predicative conditions and attributive conditions is visible beginning at approximately 500 ms, extending until the end of the epoch, in the frontal and central regions (see Figure 4 below). No peak is observed (perhaps due to epoch length): it is rather a trend in which both predicative conditions maintain a quite stable relative negativity in relation to both attributive conditions, which gradually increase in positivity.

This divergence can also be seen in a 2D plotting of the ERP map (Figure 5): for example, if we plot the average difference between attributive agreement and predicative agreement conditions (condition 1 minus condition 3) from 400 ms to 800 ms, we can observe increased positivity in the frontal region for the attributive agreement condition.

Figure 2. Overview of the Grand Average ERPs for all 30 scalp electrodes for all 4 experimental conditions

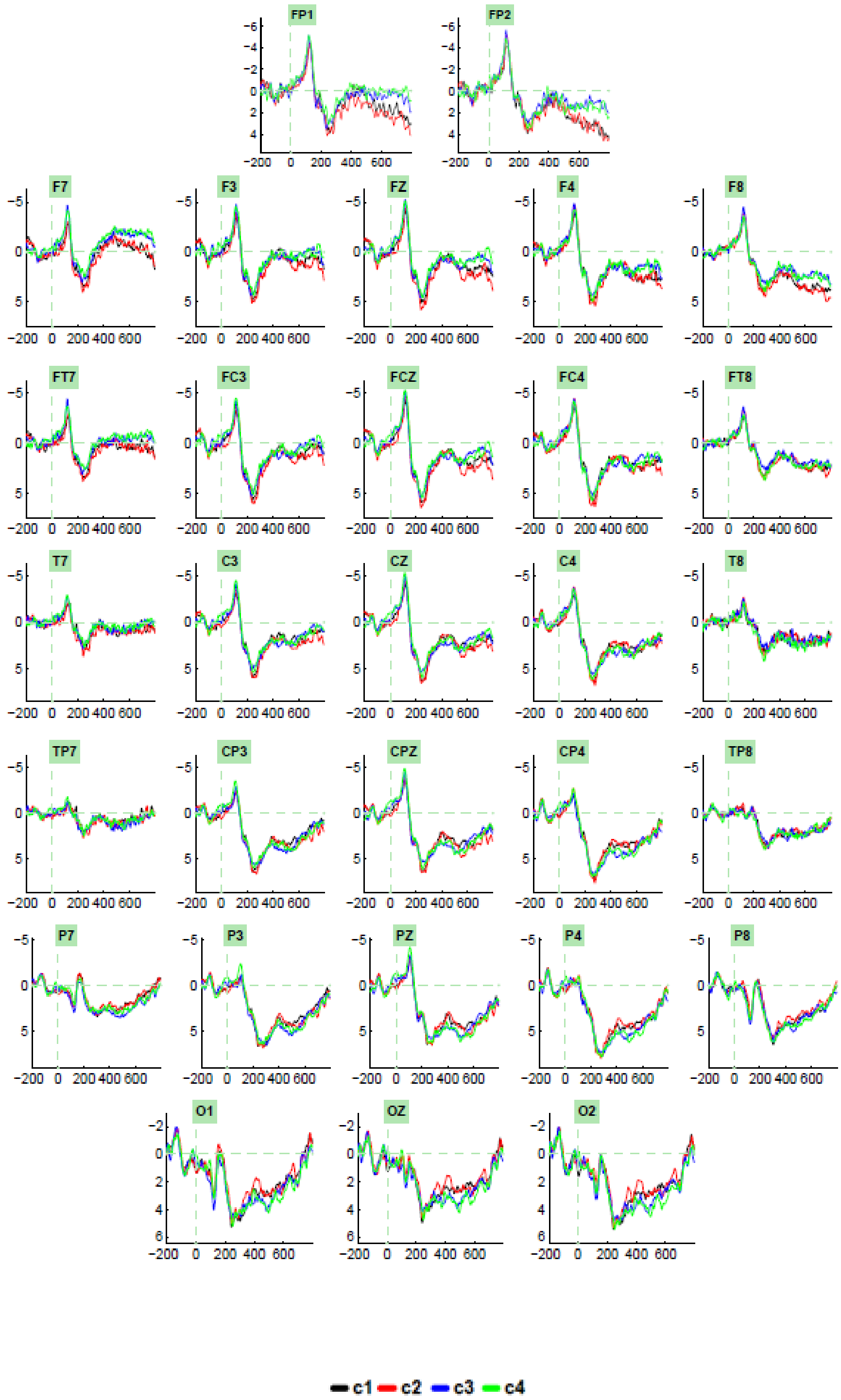


Figure 3. Grand Average ERPs of 3 midline electrodes for conditions 1 (attributive agreement) and 2 (attributive agreement violation) showing a small negative peak at approximately 580 ms

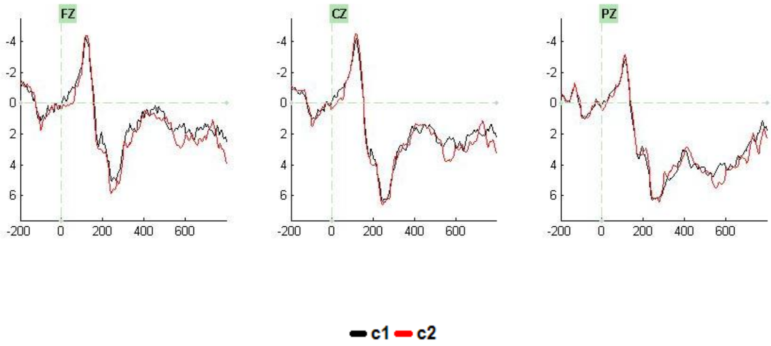


Figure 4. Grand Averaged ERPs of frontal and central electrodes showing a divergence between attributive (1 and 2) and predicative (3 and 4) conditions beginning at approximately 500 ms

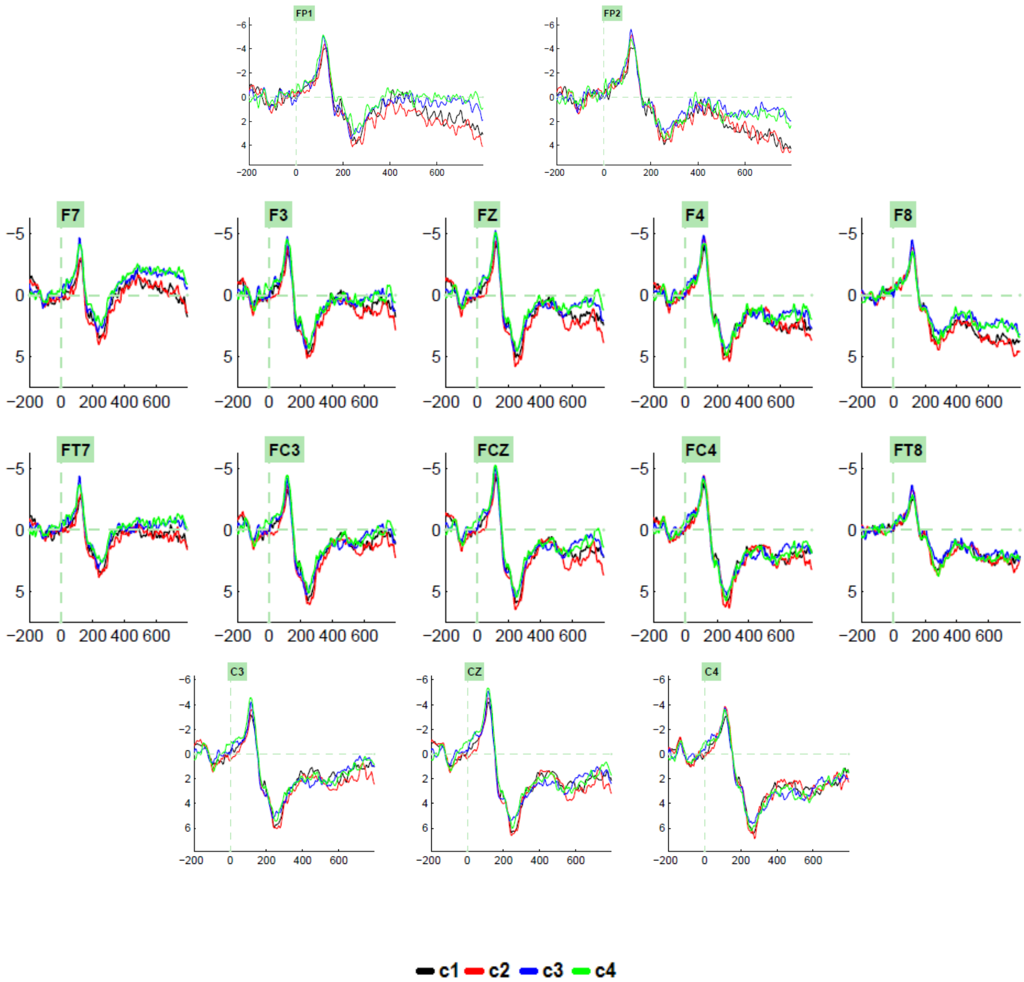
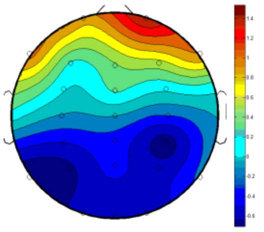


Figure 5. 2D ERP plot of mean amplitude at 400-800 ms for difference between attributive agreement (condition 1) and predicative agreement (condition 3)



2.6 Statistical analysis[11]

2.6.1 Behavioral data

Data was transferred to statistical software SPSS (SPSS 2010) for analysis. The original grammaticality judgment accuracy percentages described above (separate conditions) were submitted to a repeated measures ANOVA according to two factors: agreement type (attributive vs. predicative, for our present purposes also referred to as processability, level 3 or 4) and grammaticality (grammatical or ungrammatical). There was no significant interaction of processability and grammaticality, nor any significant effect of grammaticality, but the within-subject tests do reveal a significant effect of processability level ($F(1,14)=4.903$, $p=0.044$) on accuracy ratings.

2.6.2 ERP data

Data was transferred to statistical software SPSS (SPSS 2010) to verify the significance of results, namely the observation of a negative peak at approximately 580 ms for the attributive agreement violation condition (condition 2) in some electrodes' plotted ERPs (see Figure 3 above). A repeated measures ANOVA was therefore run using three factors: agreement type (processability level), grammaticality, and electrode ERP averages (for all 30 scalp electrodes). These tests did not reveal any significant effect or interaction (see Table 5).

[11] Where different sphericity corrections yielded different figures, the Greenhouse-Geisser correction was used.

Table 5. Tests of within-subject effects for timeframe 550-650 ms

Factor	*F*	*P*
Processability	.945	.347
Grammaticality	.421	.527
Electrodes	8.991	.000
processability * grammaticality	.004	.951
processability * electrodes	2.594	.075
grammaticality * electrodes	.743	.549
processability * grammaticality * electrodes	.463	.993

Significant effects of grammaticality for this same timeframe were not observed either when using separate repeated measures ANOVA for the attributive conditions (1- grammatical vs. 2- with violation) ($F(1,14)=.438$, $p=.519$) and the predicative conditions (3- grammatical versus 4- with violation) ($F(1,14)=.303$, $p.=591$) over all 30 electrodes. When the analysis was restricted to conditions 1 and 2 for the electrodes along the midline (FZ, FCZ, CZ, CPZ, PZ), where negative peaks for condition 2 seemed observable around 580 ms (see ERP results above), the effect of grammaticality still was not significant ($F(1,4)= .483$, $p=.525$).

An effect of agreement type/processability level on electrode voltages over a larger timeframe was however revealed to be significant (grammaticality remained a non-factor): between 500-800 ms, there is a distinct interaction ($F(29,406)=3.667$, $p=0.022$) between the processability and electrode factors. As this was observable mainly in the frontal and somewhat in the central electrodes (see ERP results above), a repeated measures ANOVA was run on just those electrodes but this still did not yield a significant effect from grammaticality.

Using data from different groups of learners, i.e. only those 90+% accuracy results in the grammatical judgment task or only those with results below 80% in more than two categories, for example, did not yield results with significant observable effects: no sub-group of learners can therefore be identified.

4. Discussion

4.1 Behavioral results

Following the statistical analysis, grammaticality was not found to have a significant effect on judgments (nor did an agreement type-grammaticality interaction). This is not surprising as it can be assumed that once a learner has processed a particular form, he or she knows when to recognize it or when it is missing or incorrectly manifested, as cited earlier in Pienemann (2007).

There are indications that agreement type, which within the current PT framework we can call processability level, does have a significant effect on participant's response accuracy percentages. This is interesting as it confirms PT's separation of the two forms, although we must consider this difference is not necessarily in conformance with PT criteria: it does not exclude previous suggestions such as those of Gillon Dowens *et al.* (2010), that the difference could relate to differences in determiner-adjective agreement vs. adjective-noun agreement (a difference not made in PT), or that other differences between the two types of agreement (sentence position and related memory loads, for example).

It is also not clear whether the statistical analysis supports the predictions made by Processability Theory regarding order of acquisition, as most participants appear to have performed slightly better with the predicative conditions. This was also the case in Gillon Dowens *et al*. (2010), but this fact was not discussed in any detail. As our present purpose was among other things to test PT as an analytical tool, we should address this discrepancy. Then again, PT does not concern itself with mastery of forms: as previously indicated, emergence is the usual criteria, and participants appeared to have acquired some competence with all conditions.

It has further been suggested that grammaticality judgments are simply incompatible with PT: this is the explanation put forth by Philipsson in his 2007 Ph.D. dissertation. When looking at different elicitation techniques, Philipsson found that grammaticality judgments stood out as the one method that did not support PT or the other theory he explored, the Markedness Differential Hypothesis: "an exclusive use of a grammaticality judgment task in this study would in fact most likely have resulted in a rejection of all the predictions of the two theories" (Philipsson 2007:185). He goes on to state that "[t]he anomalous character of the grammaticality judgment data in comparison with oral and written production gives rise to the question of what kind of cognitive operation grammaticality judgment actually is" (*Ibid.*). It is true that this formal exercise is not necessarily equivalent to any event in natural language.

It must be noted that Philippson contrasts grammaticality judgments, which evaluate grammatical perception (or intuition, he claims), and oral and written production. In this he returns to the prime role given to PT, that is, to serve as a description of L2 speech production stages. It is after all possible that PT does not correspond to acquisition of L2 comprehension procedures. Still, we must account for the present ERP results, which indicate an effect of processability level on voltage distribution among electrodes (see Discussion below), as well as the findings of Keatinge and Keßler (2009) and Håkansson and Yager (under review), where picture-pointing tasks did show learners to follow the same order of acquisition stages in comprehension and in production. Therefore it may be worthwhile to find other ways of testing comprehension[12] as behavioral tasks to accompany EEG recording when testing SLA theories. A picture identification task could possibly follow the written sentence, combining the present experimental design with that more similar to the experiment of Håkansson and Yager, for example.

There is of course something inherent to ERP experiments which cannot be helped, at least with current technologies and knowledge: the artificiality of the context, and the restrictions on the material and conditions. This artificiality could leave more room to strategies not as easily used in natural communication settings (e.g. the use of visual cues). The use of non-linguistic strategies could explain the discrepancy between expected ERP results and accuracy percentages (see section 4.2 below), though this is unlikely.

The experiment was designed mainly to control for ERP results, and this could also have been a factor affecting behavioral results: when formulating their offline judgments for sentences in the predicative conditions, participants had access to two coordinated adjectives (e.g. '*Hundar är* ***snälla*** *och* ***dyra***'/ *'*Hundar är* ***snäll*** *och* ***dyra***'), which could have been easier to take note of when formulating a judgment.

[12] Philippson also looked at a "Receptive skills task" where SSL learners had to indicate whether a sentence was a question or not, a task requiring knowledge of Swedish word order. He found results from this experiment to support PT predictions. This shows that perhaps another question task could be found that would be more compatible with PT.

The high number of repetitions used in an ERP experiment could also magnify certain phenomena of interlanguage (see Selinker 1972). For example, if a learner has created a rule where *a*- endings signify a plural, the singular for cake (‘*kaka’*) could be misinterpreted as an agreement error. It will be a challenge to future researchers, especially those wishing to work with PT, to find tasks accessible to earlier L2 learners that have limited negative effects. Optimal data for PT is production in a natural setting (Håkansson and Yager, under review), so there is a definite adjustment when using ERP, which is a highly unusual context and format.

4.2 ERP results

Results do show significant effect of processability on scalp voltage distribution even though the hypothesized increased LAN or P600 could not be conclusively observed. Visual inspection shows an increasing frontal positivity for the attributive agreement conditions beginning at approximately 500 ms. This appears to correspond to a “slow, frontal, positive-going wave [that] appears to index successful integration” found in King and Kutas (1995), something the authors present as a little-discussed phenomenon exposed in other studies using syntactically simple sentences. It could correspond to the Positive Slow Wave (PSW), the presence of which has been linked to retrieval of information from working memory (García-Larrea and Cézanne-Bert 1998). The fact that this effect was observed for the attributive conditions and not the predicative conditions could indicate that, in conformance with PT predictions, participants are able to process attributive agreement online before they are able to process predicative agreement online. However, more experimentation would be required to sustain this claim.

Many studies show later than usual (native) ERP responses in less-advanced L2 learners, which could theoretically mean that the frontal positivity might in fact be leading to a delayed form of frontal P600, as found by Kaan and Swaab (2003): the authors suggest that such a P600 is not associated with reactions to ungrammaticality, which could explain why it is present in both the grammatical and ungrammatical attributive conditions. As for their conclusion that frontal P600 are elicited in structures of greater complexity requiring revision and repair, this would contradict the PT evaluation of attributive agreement as coming before predicative agreement, the frontal positivity being found only in the former. Then again, the fact that learners are not showing a response to an even more complex structure could confirm the theory that they have not yet reached that processing level. The present experiment does not allow us to speculate further, but it would be interesting to follow-up on these elements.

If we compare the actual data with expected results, the P600-like positivity peak observed in some electrodes at approximately 580 ms was not shown to be significant. If we are to trust the review by Steinhauer *et al*. (see Table 6 below), the lack of increased LAN/P600 could be due to participants being too early learners, i.e. novices.

Table 6. Steinhauer *et al*.'s summary of ERP effects for stages of L2 morpho-syntactic acquisition

Stage	Proficiency level	ERP pattern	Cognitive processes
1	Novice	No difference	Indifferent perception; performance at or near chance level
2	Very low proficiency	N400 or right-lateralized/posterior negativities	Difficulties during lexical access and integration; compensatory processing strategy, likely relyingon semantic plausibility and pragmatics
3	Low to intermediate	Small/delayed P600 (possibly preceded by N400s)	Beginning gram maticalization/ proceduralization
4	Intermediate	Larger/earlier P600	Late structural reanalysis/repair approaches native-like mecshnisms
5	Intermediate to high/near native-like	Bilateral AN + P600	Near-native processing; early automatic + late controlled processing
6	very high/ native-like	Lateralized LAN + P600	Native-like processing; early automatic + late controlled processing

(Steinhauer *et al*. 2009)

However, the high accuracy percentages of most participants for the behavioral task seems to indicate that though far from native-like, participant had a level of proficiency high enough to warrant, for example, a stronger N400 (as in McLaughlin *et al*. 2010) indicating efforts due to lexical access and integration. But what could be very weak N400 components are truly too weak to identify as such. There is also a possibility participants were using non-linguistics strategies (see section 4.1 above), but this is admittedly unlikely.

The weakness of the expected effects could simply be the result of too important of an inter-subject variability (see 4.3 below), as discussed in McLaughlin et al. (2010). The authors stress that such variability could reduce or obscure effects, which tend to be small in novice learners.

If variability is indeed the key to the fact that our hypotheses initially elaborated according to PT were not confirmed, this would mean that the significant interaction of processability/agreement type and electrode voltage distribution, as represented in Figures 4 and 5 above, is truly important. It would therefore be necessary to test this structure again with a more homogenous, perhaps slightly more proficient group of learners.

4.3 Participants

ERP experiments require a sample that is as homogenous as possible. Unfortunately this is a very difficult criterion to respect in studies of multilingualism. The term “Swedish as a second language” is a misnomer for all of the participants in the present experiment: all had learned at least one other foreign language prior to learning Swedish. In fact all but one participant indicated learning at least two foreign languages before having contact with SSL. Therefore there was quite a lot of variation in languages spoken by participants.

Another factor of variation among participants was competence level and experience of SSL. Participants were recruited through an announcement indicating that advanced beginners or intermediate-level learners of Swedish were required for an experiment. Participants therefore firstly self-evaluated by deciding to participate, and confirmed their self-evaluation when completing the Language History Questionnaire. Participants’ responses show a great variety of experiences with Swedish, including for duration of exposure to the language (months, years), methods used to acquire the language (courses in or outside of Sweden, daily contact, etc.) and daily contact and use of the language. As this was a small pilot-like experiment, and due to difficulties in recruiting participants, it was accepted that this was not an ideal way of forming a participant group. For more accurate ERP results, participants would preferably be evaluated in a more in-depth manner to ensure better homogeneity in competence levels, perhaps by expanding the study to include a picture-pointing experiment or simply more evaluation methods such as proficiency tests or interviews. However, we should be careful in using evaluation methods that evaluate production if we mean to test comprehension: the study by Håkansson and Yager (under review) shows that participants may be at different levels for comprehension vs. production.

A quick note should be made regarding those participants claiming to be ambidextrous: again to ensure a more homogenous sample of participants, it may be worthwhile to use a right-handedness assessment test as was used by Gillon Dowens *et al.* (2010).

There is a common belief that significant online ERP responses may be observable only in native or highly proficient L2 speakers (Gillon Dowens *et al.* 2010). However the review of studies by McLaughlin *et al*. (2010) shows that ERP responses in novices can be highly systematic, We believe the current experiment shows that significant effects of features belonging to different grammatical acquisition levels can be seen in less proficient speakers, and that it is worth pursuing further ERP studies on such speaker populations.

To limit the scope of the experiment, no control group of native speakers was used in this experiment. This would be relevant in future studies, not so much to set a target for learners' to reach, but rather to isolate effects which are particular to learners from those which are common to all language users.

5. Final comments and further research

A study by Tokowicz and MacWhinney (2005) shows that implicit (ERP) and explicit (behavioral task results) measures can differ. In our present case, it is not clear if behavioral results corroborate the findings regarding an apparent order of acquisition in conformance with PT predictions; however both the behavioral and ERP data mark an effect of agreement type even in these less advanced learners, showing this is a promising venue to pursue.

The initial idea to test and interpret ERP results using PT came from a belief that general SLA theory could hold a larger place in ERP studies of L2 learners, and that general theory could benefit from more practical and online testing. SLA theories could help explain what remains unclear in successful experiments like that of Gillon Dowens *et al*. (2010). It could lead to the creation of new experiments, whether these have the intended results or not.

In any case, shaping an experiment using a SLA theory is a good way to test said theory, to find its applications and limitations – as in the current study, where the hypotheses were not confirmed, or to analyze results when hypotheses are validated. For one, PT theory shows promise for ERP experiment contexts and should be tested again. Due to possibly slower responses in L2 learners, epochs could be extended for analysis purposes to ensure no components are missed, and slides could be presented at an even slower rate, for example.

The relation between SLA theory and experiments should work in the opposite direction as well: theories of SLA could perhaps take better note of experiments and conclusions such as those of McLaughlin *et al*. (2010) and Steinhauer *et al.* (2009).

To conclude by returning to our research questions, it appears that there is a difference in online response between the processing of attributive agreement and the processing of predicative agreement. Whether this is due to a hierarchical acquisition of forms, as put forth by PT, cannot be confirmed, but attributive agreement does appear to be processed online by learners in a manner not accessible for predicative agreement, However it would be worthwhile to test this further, for example by exploring the different suggestions and improvements presented in the above discussion and conclusion.

Acknowledgements

I would like to thank my supervisors, Gisela Håkansson and Mikael Roll, for their help, patience, and insight, and for their enthusiasm for the project – I hope my work honors the efforts and the faith they placed in me. I am also grateful to Merle Horne for opening the doors of this Master's program to me and for introducing the possibilities offered by neurolinguistics research. Thanks as well to Kristina Magnusson and Emelie Stiernströmer, as well as my colleague Hatice Zora, for their material and moral support in the laboratory. Bravo to Joanne Yager for her excellent Bachelor's thesis, which started it all. I must also express my boundless gratitude to my participants and friends: I thank them for their time, good humor and the interest they took in my experiment. *Merci* to my *sambo* Pac for initializing our return to Sweden, and for his understanding and encouragement throughout this intense process. *Merci aussi* to my parents and grand-parents for putting in time and mileage to come visit this spring; they brought the sun and the best reason to take breaks from the computer to enjoy it.

References

Barber, H. and M. Carreiras. 2005. "Grammatical gender and number agreement in Spanish: an ERP comparison". *Journal of Cognitive Neuroscience* 17 (1): 137-153.

Bresnan, J. 2001. *Lexical Functional Syntax*. Malden, MA: Blackwell.

Compumedic Neuroscan. Scan Version 4.5. 2009. Compumedic Neuroscan.

Davidson, D.J., P. Indefrey and M. Gullberg. 2008. "Words that second language learners are likely to hear, read, and use". *Bilingualism: Language and Cognition* 11(1): 133-146.

Delorme, A. and S. Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134: 9-21.

Ejerhed, E., G. Källgren, O. Wennstedt and M. Aström. 1992. *The linguistic annotation system of the Stockholm-Umeå Corpus project*. Technical Report 33, Department of General Linguistics, Umeå University. Umeå, Sweden.

García-Larrea, L. and G. Cézanne-Bert. 1998. "P3, Positive slow wave and working memory load: a study on the functional correlates of slow wave activity". *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 108(3): 260-273.

Gillon Dowens, M., M. Vergara, H. A. Barber, and M. Carreiras. 2010. "Morpho-syntactic Processing in Late Second-Language Learners". *Journal of Cognitive Neuroscience* 22 (8): 1870-1887.

Glahn, E., G. Håkansson, B. Hammarberg, A. Holmen, A. Hvenekilde and K. Lund, 2001. "Processability in Scandinavian Second Language Acquisition". *Studies in Second Language Acquisition* 23: 389-416.

Göransson, U. and M. Parada. 1997. *På svenska!*. Lund: Folkuniversitetets förlag.

Håkansson, G., and Hansson, K. 2000. "Comprehension and production of relative clauses: A comparison between Swedish impaired and unimpaired children". *Journal of Child Language* 27: 313–333.

Håkansson, G. and C. Norrby. 2006. "Processability Theory applied to written and spoken L2 Swedish". In Mansouri, F. (Ed.) *Bilingualism and theory-driven second language acquisition research*. Newcastle upon Tyne: Cambridge Scholars Publishing. 81-117.

Håkansson, G. and J. Yager. Under review. "Processability = produçability? Processability Theory and comprehension vs. production of Swedish agreement morphology". Under review by *Second Language Research.*

Hendriks, P. and C. Koster (eds.). 2010. *Special issue on asymmetries in language acquisition*. *Lingua* 120(8).

Ingram, J. 2007. *Neurolinguistics: An introduction to spoken language processing and its disorders*. Cambridge: Cambridge University Press.

Jordan, G. 2004. *Theory Construction in Second Language Learning*. Amsterdam: John Benjamins Publishing Company.

Kaan, E. and T.Y. Swaab. 2003. “Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation”. *Journal of Cognitive Neuroscience* 15(1): 98-110.

Keatinge, D. and J.-U. Keßler. 2009. “The Acquisition of the Passive Voice in L2-English: Perception and Production.” In J.-U. Keßler and D. Keatinge (eds.), *Research in Second Language Acquisition: Empirical Evidence Across Languages*. Newcastle upon Tyne: Cambridge Scholars Publishing. 67-92.

King, J.W. and M. Kutas. 1995. “Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading”. *Journal of Cognitive Neuroscience* 7(3): 376-396.

Kotz, S.A. 2009. “A critical review of ERP and fMRI evidence on L2 syntactic processing”. *Brain and Language* 109: 68-74.

Levelt, W.J.M. 1989. *Speaking: From intention to articulation.* Cambridge. Massachusetts: MIT Press.

Luck, S. J. 2005. *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: The MIT Press.

Markley, C. S. Luck and J. Lopez-Calderon. 2010. “Introduction to ERPLAB Toolbox”. *ERPLAB Toolbox User's Manual*. http://erpinfo.org/erplab/erplab-documentation/erplab-manual/Intro_to_ERPLAB.htm. Retrieved 22 April 2011.

McLaughlin, J., D. Tanner, I. Pitkanen, C. Frenck-Mestre, K. Inoue, G. Valentine and L. Osterhout. 2010. “Brain potentials reveal discrete stages of L2 grammatical learning”. *Language Learning* 60(s2): 123-150.

Osterhout, L., J. McLaughlin, I. Pitkänen, C. Frenck-Mestre, and N. Molinaro. 2006. “Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing”. *Language Learning* 56(s1): 199-230.

Osterhout, L. and L.A. Mobley. 1995. “Event-related brain potentials elicited by failure to agree”. *Journal of Memory and Language* 34: 739-773.

Philipsson, A. 2007. *Interrogative Clauses and Verb Morphology in L2 Swedish: Theoretical Interpretations of Grammatical Development and Effects of Different Elicitation Techniques*. Ph.D. Dissertation. Stockholm University, Centre for Bilingualism.

Pienemann, M. 2007. “An introduction to Processability Theory”. In F. Mansouri (ed.), *Second Language Acquisition Research: Theory-Construction and Testing*. Newcastle upon Tyne: Cambridge Scholars Publishing. 13-38.

Pienemann, M. 2005. “An introduction to Processability Theory”. In M. Pienemann (ed.), *Cross-Linguistic Aspects of Processability Theory*. Amsterdam: John Benjamins. 1-62.

Pienemann, M. 2003. “Language Processing Capacity”. In C. Doughty M. Long (eds.), *The Handbook of second language acquisition*. Oxford: Blackwell. 679 – 714.

Pienemann, M. 1998. *Language Processing and Second Language Development: Processability theory*. Amsterdam: John Benjamins.

Pienemann, M. and G. Håkansson. 1999. “A unified approach towards a theory of the development of Swedish as L2. A processability account.” *Studies in Second Language Acquisition* 21: 383-420.

Roll, M., M. Horne and M. Lindgren. 2007. “Object shift and event-related brain potentials”. *Journal of Neurolinguistics,* 20: 462-481.

Sabourin, L. 2009. “Neuroimaging and research into second language acquisition”. *Second Language Research* 25(1): 5-11.

Sabourin, L and L.A. Stowe. 2008. “Second language processing: when are first and second languages processed similarly?”. *Second Language Research* 24(3): 397-430.

Schneider, W., A. Eschman and A. Zuccolotto. 2002. *E-Prime user’s guide*. Pittsburgh: Psychology Software Tools Inc.

Selinker, L. 1972. “Interlanguage”. *International Review of Applied Linguistics in Language Teaching* 10(3): 209.

SPSS for Windows, Rel. 19.0.0. 2010. Chicago: SPSS Inc.

Steinhauer, K. E.J. White and J.E. Drury. 2009. “Temporal dynamics of late second language acquisition: evidence from event-related brain potentials”. *Second Language Research* 25(1): 13-41.

The MathWorks, MATLAB Version 7.9.0 (R2009b). 2009. The MathWorks Inc.

Tokowicz, N. and B. MacWhinney. 2005. “Implicit and Explicit Measures of Sensitivity to Violations in Second Language Grammar”. *Studies in Second Language Acquisition*, 27: 173-204.

van Hell, J.G. and N. Tokowicz. 2010. “Event-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 proficiency levels”. *Second Language Research* 26(1): 43-74.

Appendix A – Sentence material

Condition 1: Attributive agreement
Condition 2: Attributive agreement violation

Condition 3: Predicative agreement
Condition 4: Predicative agreement violation

Slide 1	Slide 2	Slide 3	Slide 4	Slide 5	Grammaticality	Condition	Trial Block
En	snäll	hund	bor	här	Y	1	1
En	duktig	student	läser	här	Y	1	1
En	varm	dryck	hjälper	alltid	Y	1	1
En	färsk	apelsin	ligger	här	Y	1	1
En	stor	björn	sover	där	Y	1	1
Flera	goda	vänner	kommer	hit	Y	1	1
Flera	fina	cyklar	kostar	mycket	Y	1	1
Flera	starka	kvinnor	campar	där	Y	1	1
Flera	svåra	kurser	finns	det	Y	1	1
Flera	centrala	restauranger	serverar	lunch	Y	1	1
En	röda	blommor	står	där	N	2	1
En	vanliga	datorer	fungerar	bra	N	2	1
En	glada	pojkar	leker	där	N	2	1
En	dyra	bilar	kör	fort	N	2	1
En	billiga	bussar	kör	direkt	N	2	1
Flera	ny	bild	hänger	där	N	2	1
Flera	trevlig	kompis	arbetar	här	N	2	1
Flera	kall	vinter	kommer	ofta	N	2	1
Flera	rolig	fest	passar	bra	N	2	1
Flera	lång	gata	delar	staden	N	2	1
En bild	är	ny	och	vacker	Y	3	1
En kompis	är	trevlig	och	fantastisk	Y	3	1
En vinter	är	kall	och	vit	Y	3	1
En fest	är	rolig	och	trevlig	Y	3	1
En gata	är	lång	och	bred	Y	3	1
Blommor	är	röda	och	sköna	Y	3	1
Datorer	är	vanliga	och	populära	Y	3	1
Pojkar	är	glada	och	pigga	Y	3	1
Bilar	är	dyra	och	dåliga	Y	3	1
Bussar	är	billiga	och	snabba	Y	3	1
En vän	är	goda	och	viktig	N	4	1
En cykel	är	fina	och	lätt	N	4	1
En kvinna	är	starka	och	snygg	N	4	1
En kurs	är	svåra	och	jobbig	N	4	1
En restaurang	är	centrala	och	populära	N	4	1
Hundar	är	snäll	och	dyra	N	4	1
Studenter	är	duktig	och	kloka	N	4	1
Drycker	är	varm	och	goda	N	4	1
Apelsiner	är	färsk	och	saftiga	N	4	1
Björnar	är	stor	och	hungriga	N	4	1
En	god	vän	kommer	hit	Y	1	2
En	fin	cykel	kostar	mycket	Y	1	2
En	stark	kvinna	campar	där	Y	1	2
En	svår	kurs	finns	det	Y	1	2
En	central	restaurang	serverar	lunch	Y	1	2
Flera	nya	bilder	hänger	där	Y	1	2

Flera	trevliga	kompisar	arbetar	här	Y	1	2
Flera	kalla	vintrar	kommer	ofta	Y	1	2
Flera	roliga	fester	passar	bra	Y	1	2
Flera	långa	gator	delar	staden	Y	1	2
En	snälla	hundar	bor	här	N	2	2
En	duktiga	studenter	läser	här	N	2	2
En	varma	drycker	hjälper	alltid	N	2	2
En	färska	apelsiner	ligger	här	N	2	2
En	stora	björnar	sover	där	N	2	2
Flera	röd	blomma	står	där	N	2	2
Flera	vanlig	dator	fungerar	bra	N	2	2
Flera	glad	pojke	leker	där	N	2	2
Flera	dyr	bil	kör	fort	N	2	2
Flera	billig	buss	kör	direkt	N	2	2
En blomma	är	röd	och	skön	Y	3	2
En dator	är	vanlig	och	populär	Y	3	2
En pojke	är	glad	och	pigg	Y	3	2
En bil	är	dyr	och	dålig	Y	3	2
En buss	är	billig	och	snabb	Y	3	2
Hundar	är	snälla	och	dyra	Y	3	2
Studenter	är	duktiga	och	kloka	Y	3	2
Drycker	är	varma	och	goda	Y	3	2
Apelsiner	är	färska	och	saftiga	Y	3	2
Björnar	är	stora	och	hungriga	Y	3	2
En bild	är	nya	och	vacker	N	4	2
En kompis	är	trevliga	och	fantastisk	N	4	2
En vinter	är	kalla	och	vit	N	4	2
En fest	är	roliga	och	trevlig	N	4	2
En gata	är	långa	och	bred	N	4	2
Vänner	är	god	och	viktiga	N	4	2
Cyklar	är	fin	och	lätta	N	4	2
Kvinnor	är	stark	och	snygga	N	4	2
Kurser	är	svår	och	jobbiga	N	4	2
Restauranger	är	central	och	populär	N	4	2
En	ny	bild	hänger	där	Y	1	3
En	trevlig	kompis	arbetar	här	Y	1	3
En	kall	vinter	kommer	ofta	Y	1	3
En	rolig	fest	passar	bra	Y	1	3
En	lång	gata	delar	staden	Y	1	3
Flera	röda	blommor	står	där	Y	1	3
Flera	vanliga	datorer	fungerar	bra	Y	1	3
Flera	glada	pojkar	leker	där	Y	1	3
Flera	dyra	bilar	kör	fort	Y	1	3
Flera	billiga	bussar	kör	direkt	Y	1	3
En	goda	vänner	kommer	hit	N	2	3
En	fina	cyklar	kostar	mycket	N	2	3
En	starka	kvinnor	campar	där	N	2	3
En	svåra	kurser	finns	det	N	2	3
En	centrala	restauranger	serverar	lunch	N	2	3
Flera	snäll	hund	bor	här	N	2	3
Flera	duktig	student	läser	här	N	2	3
Flera	varm	dryck	hjälper	alltid	N	2	3

Flera	färsk	apelsin	ligger	här	N	2	3
Flera	stor	björn	sover	där	N	2	3
En hund	är	snäll	och	dyr	Y	3	3
En student	är	duktig	och	klok	Y	3	3
En dryck	är	varm	och	god	Y	3	3
En apelsin	är	färsk	och	saftig	Y	3	3
En björn	är	stor	och	hungrig	Y	3	3
Vänner	är	goda	och	viktiga	Y	3	3
Cyklar	är	fina	och	lätta	Y	3	3
Kvinnor	är	starka	och	snygga	Y	3	3
Kurser	är	svåra	och	jobbiga	Y	3	3
Restauranger	är	centrala	och	populära	Y	3	3
En blomma	är	röda	och	skön	N	4	3
En dator	är	vanliga	och	populär	N	4	3
En pojke	är	glada	och	pigg	N	4	3
En bil	är	dyra	och	dålig	N	4	3
En buss	är	billiga	och	snabb	N	4	3
Bilder	är	ny	och	vackra	N	4	3
Kompisar	är	trevlig	och	fantastiska	N	4	3
Vintrar	är	kall	och	vita	N	4	3
Fester	är	rolig	och	trevliga	N	4	3
Gator	är	lång	och	breda	N	4	3
En	röd	blomma	står	där	Y	1	4
En	vanlig	dator	fungerar	bra	Y	1	4
En	glad	pojke	leker	där	Y	1	4
En	dyr	bil	kör	fort	Y	1	4
En	billig	buss	kör	direkt	Y	1	4
Flera	snälla	hundar	bor	här	Y	1	4
Flera	duktiga	studenter	läser	här	Y	1	4
Flera	varma	drycker	hjälper	alltid	Y	1	4
Flera	färska	apelsiner	ligger	här	Y	1	4
Flera	stora	björnar	sover	där	Y	1	4
En	nya	bilder	hänger	där	N	2	4
En	trevliga	kompisar	arbetar	här	N	2	4
En	kalla	vintrar	kommer	ofta	N	2	4
En	roliga	fester	passar	bra	N	2	4
En	långa	gator	delar	staden	N	2	4
Flera	god	vän	kommer	hit	N	2	4
Flera	fin	cykel	kostar	mycket	N	2	4
Flera	stark	kvinna	campar	där	N	2	4
Flera	svår	kurs	finns	det	N	2	4
Flera	central	restaurang	serverar	lunch	N	2	4
En vän	är	god	och	viktig	Y	3	4
En cykel	är	fin	och	lätt	Y	3	4
En kvinna	är	stark	och	snygg	Y	3	4
En kurs	är	svår	och	jobbig	Y	3	4
En restaurang	är	central	och	populär	Y	3	4
Bilder	är	nya	och	vackra	Y	3	4
Kompisar	är	trevliga	och	fantastiska	Y	3	4
Vintrar	är	kalla	och	vita	Y	3	4
Fester	är	roliga	och	trevliga	Y	3	4
Gator	är	långa	och	breda	Y	3	4

En hund	är	snälla	och	dyr	N	4	4
En student	är	duktiga	och	klok	N	4	4
En dryck	är	varma	och	god	N	4	4
En apelsin	är	färska	och	saftig	N	4	4
En björn	är	stora	och	hungrig	N	4	4
Blommor	är	röd	och	sköna	N	4	4
Datorer	är	vanlig	och	populära	N	4	4
Pojkar	är	glad	och	pigga	N	4	4
Bilar	är	dyr	och	dåliga	N	4	4
Bussar	är	billig	och	snabba	N	4	4

Appendix B – Language History Questionnaire

Date: _______________________________ *Participant (to be completed by researcher):* _________

Below are questions about your background and your language use. Please answer these questions as completely as possible.

Background:

Age: _______________ Gender: M F Dominant hand: Left Right

Highest completed level of education: ____

Occupation/Field of work or study: ____

Where did you grow up (city/region, country, ages 0-5)? ____

Do you still reside in this location ? Y N If no, where is your permanent residence now?

(City/region, country) ____

Have you ever suffered any brain injuries or undergone brain surgery? Y N

If so, please explain: **____**

Language History*:**
What is/are your native language/s? **____**

Please list all languages you know below (including native language/s). For each, rate how well you can use the language (Not Good 1 2 3 4 5 6 Very Good):

Language	Speaking	Listening	Writing	Reading	Grammar	Pronun-ciation
1						
2						
3						
4						
5						

For the languages you listed, how many hours a week do you spend doing the following activities?

Language	Reading, email, internet	Watching TV/ videos	Listening radio/ music	Talking with others	Studying (in class or self-study)
1					
2					
3					
4					
5					

Please explain briefly your experience of the Swedish language: when you started learning it/for how long, in what context, what you did/do to learn it (classes, self-study, etc.):

____

******* If you have any other remarks about your language history that you think may be important for understanding your learning or use of these languages, please feel free to write them here:

____

Questionnaire developed using the following resources:

Gullberg, M. and P. Indefrey. 2003. “Language Background Questionnaire”. *The Dynamics of Multilingual Processing*. Nijmegen: Max Planck Institute for Psycholinguistics. Web. 10 February 2010.

Li, P., S. Sepanski, and X. Zhao. 2006. “Language history questionnaire: A Web-based interface for bilingual research”. *Behavior Research Methods* 38(2): 202-210.