Cognitive and neural mechanisms of inflectional morphology processing
Studies of native speakers and second language learners of Swedish
Schremm, Andrea

2018

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Cognitive and neural mechanisms of inflectional morphology processing
Studies of native speakers and second language learners of Swedish

ANDREA SCHREMM
CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY
Cognitive and neural mechanisms of inflectional morphology processing
Studies of native speakers and second language learners of Swedish

Andrea Schremm

LUND UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculties of Humanities and Theology, Lund University, Sweden.
To be defended at LUX Aula Övre, Friday September 14, 2018, 10:15-12:00.

Faculty opponent
Associate Professor Nuria Sagarra
Cognitive and neural mechanisms of inflectional morphology processing: Studies of native speakers and second language learners of Swedish

Abstract
The present dissertation investigates inflectional morphology processing in native speakers and second language (L2) learners of Swedish. Results of Study 1 suggest that two separate neural mechanisms might be available for native comprehension of inflected words, as reflected in event-related brain potentials obtained for visually presented verb forms. Overregularized verbs (e.g. *bär+de ‘bear + past tense’) yielded a left anterior negativity (LAN), indicating decompositional processing of the regular tense inflection versus whole-word retrieval of correct irregular verb forms (e.g. bär ‘bore’). Enhanced long-range neural oscillatory phase synchrony observed for familiar irregular words potentially reflected increased engagement of the ventral language processing stream during whole-word access. As Swedish is characterized by a predictive association between specific word stem tones and upcoming suffixes, facilitating speech processing, Study 2 examines the integration of tonal cues into the native morphological system. Correlational analysis was conducted between cortical thickness in selected brain regions and individual participants’ response time patterns for suffix recognition following the tonal cue in real words (e.g. hatt[^1]++[^1]en ‘hat+sg’) and pseudowords (e.g. kvut[^1]++[^1]en ‘kvut+sg’). Results suggest that the left planum temporale might play a role when tones are accessed as part of whole-word memory representations, whereas the pars opercularis of the left inferior frontal gyrus could potentially support rule-based decompositional analysis of cued suffixes when no stored full-form representations are present. Study 3 focuses on the L2 acquisition of the tonal aspects of Swedish inflectional morphology. Response time patterns to inflected verbs indicate facilitated processing of word endings validly cued by the preceding stem tone in proficient L2 learners of Swedish, who had not received any explicit information about the tested L2 regularity. As these results suggested gradual and slow implicit acquisition of tone-suffix associations through exposure to L2 input, Study 4 explores possibilities of training the L2 feature at earlier stages of learning. Performance data collected during a two-week-period of training with a game prototype show gradually faster and more accurate responses to suffixes cued by preceding tones, indicating that low proficient learners start to integrate Swedish word accents into their L2 morphological processing system.

Key words: morphology, inflection, linguistic tone, ERP, cortical thickness, oscillatory phase synchrony, left anterior negativity, second language acquisition, response times, implicit learning, computer assisted learning
Cognitive and neural mechanisms of inflectional morphology processing

Studies of native speakers and second language learners of Swedish

Andrea Schremm

LUND UNIVERSITY
## Contents

Acknowledgements .................................................................................................. 7  
Abstract .................................................................................................................... 9  
List of original papers ............................................................................................ 11  
1. Introduction ........................................................................................................ 13  
2. Background ........................................................................................................ 17  
   2.1 Inflectional morphology ......................................................................... 17  
   2.2 Dual route model of inflectional morphology processing .................. 18  
   2.3 Brain networks of language .................................................................... 20  
      2.3.1 Neural substrates of dual route processing................................. 21  
   2.4 Swedish morphology ............................................................................. 23  
   2.5 Swedish word accents ............................................................................ 25  
   2.6 The processing of word accent-suffix associations in native speakers .. 26  
   2.7 Investigating second language acquisition of Swedish tone-suffix associations .................................................. 28  
      2.7.1 Tone perception in L2 acquisition.............................................. 29  
      2.7.2 Implicit and explicit L2 learning................................................ 31  
3. Methods .............................................................................................................. 35  
   3.1. Response times...................................................................................... 35  
   3.2 EEG and ERPs ....................................................................................... 35  
      3.2.1 ERPs in language processing ...................................................... 38  
      3.2.2 Neural oscillations......................................................................... 40  
   3.3 MRI and cortical thickness measurements.......................................... 42  
      3.3.1 The MRI technique .................................................................... 43  
      3.3.2 Measuring cortical thickness......................................................... 44  
4. The investigations .............................................................................................. 47  
   4.1 Study 1 – Brain responses to morphologically complex verbs........... 47  
   4.2 Study 2 – Cortical thickness in native language tone processing ........ 49
4.3 Study 3 – Implicit acquisition of tone-suffix connections in L2 learners of Swedish

4.4 Study 4 – Training predictive L2 processing with a digital game

5. Conclusions

6. Outstanding issues and future directions

   6.1 Factors modulating inflectional morphology processing mechanisms in native speakers

   6.2 Issues related to L2 processing and acquisition of tone-suffix associations

   6.3 Neural oscillatory synchrony and the language processing streams

References
Acknowledgements

First and foremost, I wish to thank my supervisors, Mikael Roll and Merle Horne, for their continuous support and insightful guidance during the past four years. I am deeply grateful for the many thought-provoking discussions we have had and for everything they taught me about language and the brain. Without their encouragement and thorough knowledge, I seriously doubt that this thesis would ever have been completed.

I would also like to thank my fellow lab mates in the neurolinguistics group – Anna Hed, Frida Blomberg, Mikael Novén, Otto Ewald, Pelle Söderström and Sabine Gosselke Berthelsen – for their support, feedback and the fun moments we have shared. A special thanks goes to Anna, Mikael, Otto and Pelle for invaluable help with experimental materials and data collection. Thank you, Briana Van Epps, for being the best office mate and a true friend during these years.

Last, but certainly not least, I would like to thank my parents, who have always been there for me, and who made sure I got the most relaxing breaks from the work with this thesis.
Abstract

The present dissertation investigates inflectional morphology processing in native speakers and second language (L2) learners of Swedish. Results of Study 1 suggest that two separate neural mechanisms might be available for native comprehension of inflected words, as reflected in event-related brain potentials obtained for visually presented verb forms. Overregularized verbs (e.g. *bär+de ‘bear + past tense’) yielded a left anterior negativity (LAN), indicating decompositional processing of the regular tense inflection versus whole-word retrieval of correct irregular verb forms (e.g. bär ‘bore’). Enhanced long-range neural oscillatory phase synchrony observed for familiar irregular words potentially reflected increased engagement of the ventral language processing stream during whole-word access.

As Swedish is characterized by a predictive association between specific word stem tones and upcoming suffixes, facilitating speech processing, Study 2 examines the integration of tonal cues into the native morphological system. Correlational analysis was conducted between cortical thickness in selected brain regions and individual participants’ response time patterns for suffix recognition following the tonal cue in real words (e.g. hat\textsuperscript{Accent 1} +en ‘hat+sg’) and pseudowords (e.g. kvut\textsuperscript{Accent 1} +en ‘kvut+sg’). Results suggest that the left planum temporale might play a role when tones are accessed as part of whole-word memory representations, whereas the pars opercularis of the left inferior frontal gyrus could potentially support rule-based decompositional analysis of cued suffixes when no stored full-form representations are present.

Study 3 focuses on the L2 acquisition of the tonal aspects of Swedish inflectional morphology. Response time patterns to inflected verbs indicate facilitated processing of word endings validly cued by the preceding stem tone in proficient L2 learners of Swedish, who had not received any explicit information about the tested L2 regularity. As these results suggested gradual and slow implicit acquisition of tone-suffix associations through exposure to L2 input, Study 4 explores possibilities of training the L2 feature at earlier stages of learning. Performance data collected during a two-week-period of training with a game prototype show gradually faster and more accurate responses to suffixes cued by preceding tones, indicating that low proficient learners start to integrate Swedish word accents into their L2 morphological processing system.
List of original papers

Study 1

Study 2

Study 3

Study 4

¹ Reprinted with permission from John Benjamins Publishing Company
1. Introduction

The combinatorial nature of language underlies its unique expressive power. By arranging a limited set of building blocks into a practically unlimited number of combinations, language enables us to flexibly communicate meanings and even produce completely novel utterances (Chomsky, 1965; Jackendoff, 2003; von Humboldt, 1836). This combinatorial property is expressed across several levels of linguistic representation: phonemes are combined into morphemes; words are assembled into phrases, which are organized in turn to form sentences. In an attempt to understand the cognitive mechanisms and brain structures involved in such fundamental processes of language, research has often looked at the production and comprehension of complex word forms. The phenomenon the present thesis focuses on, inflectional morphology, has specifically been argued to capture important aspects of the general combinatorial property within the word level (Pinker, 1999). For example, regularly inflected words such as walked can be viewed as combinations of smaller building blocks, e.g. walk plus the past tense suffix -ed. The sequencing and choice of elements are clearly not arbitrary but appear to follow some underlying pattern, which might be expressed as rules. By contrast, many irregular words such as went seem to be simple word units rather than assemblies of morpheme parts. From this perspective, regularly inflected items could be produced and comprehended by combining the component parts “on the fly”, based on some specific regularities, whereas irregular words would have to be retrieved in their full form from memory. The question whether these descriptive differences between the two types of word forms actually translate into two separate processing routes in the brain has constituted one of the core issues of psycholinguistic literature on language processing (e.g. Pinker & Ullman, 2002; Regel, Opitz, Müller, & Friederici, 2015).

A unique property of Swedish adds an extra dimension to the investigation of inflectional morphology processing. Thus, in Swedish, each prosodic word carries a tonal pattern, a so-called word accent. Even though word accents are realized on word stems, they are associated with specific suffixes attached to the stem (Riad, 2014; Rischel, 1963), which enables native speakers to anticipate possible upcoming endings already as they hear the beginning of the word (Roll, 2015; Roll, Horne, & Lindgren, 2010; Roll, Söderström, & Horne, 2013; Roll et al., 2015; Söderström, Horne, & Roll, 2016b; Söderström, Roll, & Horne, 2012). Such endings include inflectional suffixes, a fact which makes word accents an
important part of Swedish inflectional morphology. Still, it is unclear how the interaction between tones signaling upcoming suffixes and the morphological processing system is implemented during on-line language comprehension in a way that enables predictive speech processing. Furthermore, the fact that the word accent-suffix association is largely unique to Scandinavian languages raises another important question: the acquisition of this feature by second language (L2) learners of Swedish. It is as yet not known whether adult L2 learners can integrate tonal information into their L2 morphological processing system and use word accent cues predictively. A related issue concerns the nature of the learning mechanisms that might facilitate L2 acquisition of word accent-suffix associations, which in turn carries important practical implications for effective L2 training strategies.

The present thesis investigated inflectional morphology processing in native speakers and L2 learners of Swedish. The first aim was to establish the nature of the processing routes available to native speakers for the comprehension of complex word forms and to identify brain structures and mechanisms that enable the integration of word accents as predictive cues into this system. Next, we addressed the question as to whether L2 learners can acquire this unique property of Swedish inflectional morphology, i.e. the tone-suffix associations and their predictive use, and how the learning processes might be supported with training. Four empirical studies were carried out, using a range of methodological approaches. Electrical activity of the brain, i.e. electroencephalography (EEG), was recorded from Swedish native speakers in order to gain insights into modulation of online neural activity during the processing of different types of inflected word forms. Thickness of the cerebral cortex was also measured in magnetic resonance (MR) images, to identify neuroanatomical correlates of processes involved in using stem tones as cues to morphological structure. With L2 learners of Swedish, the main approach was the measurement of response times to complex word forms. Patterns of response latency and accuracy to word endings cued by tones shed light on the way with which tone-suffix associations are processed after longer or shorter exposure to the L2 and were used to evaluate the efficiency of focused training in the L2 feature using a digital game.

Study 1 examined the general processing mechanisms involved in the comprehension of inflected words, by recording EEG as Swedish native speakers read regular versus irregular verb forms. Event-related potentials (ERP) and neural oscillatory patterns, extracted from the EEG, were analyzed to establish whether regular morphological patterns would trigger decompositional analysis and if familiar irregular items would be accessed as stored whole word forms in memory. Study 2 introduced the dimension of word accents, employing auditory stimuli, and looked at the way word accents might interact with the two proposed processing routes for inflected word forms. Therefore, the reliance on tonal cues to facilitate suffix processing was examined in two contexts: in real words, where
retrieval of stored memory representations might play a role, and in inflected pseudowords, where comprehension depends on the extraction of the grammatically meaningful element, the suffix, which could be expected to engage morphological parsing mechanisms. In order to identify specific brain regions that support morphological processing, individual differences in cortical thickness measurements were correlated with the degree to which participants used tonal cues predictively on the different word forms. A meaningful relation between variation of cortical structure and tone processing performance would point to the involvement of the specific brain region as a neuroanatomical substrate. The two brain areas examined were the pars opercularis of the inferior frontal gyrus and the planum temporale, which have previously been implicated in inflectional morphological analysis and tone processing on familiar lexical items, respectively.

Study 3 and 4 focused on the L2 acquisition of Swedish tone-suffix associations. Predictive features in language rely on co-variations between an anticipatory cue (e.g. the stem tone) and a target (e.g. a suffix attached to the stem). A powerful domain-general learning mechanism for picking up co-variations in the input is statistical learning, which is assumed to proceed like implicit learning, i.e. incidentally and without conscious awareness of the regularity to be learned (Perruchet & Pacton, 2006). Study 3 thus investigated the assumption that L2 learners at higher levels of proficiency in Swedish might have implicitly acquired the predictive use of word accents during a relatively long period of exposure to the L2. In order to assess the degree of learning that has taken place without any explicit instruction, we examined whether these advanced learners would show a processing advantage in their response times to verbal inflections that were cued by the preceding word accent, in the same way as it was observed in native speakers. Study 4 focused on L2 acquisition at earlier stages of learning. Since implicit learning might proceed relatively slowly, and low proficient L2 learners do not show signs of acquiring the predictive use of word accents (Gosselke Berthelsen, Horne, Brännström, Shtyrov, & Roll, 2018), we investigated possibilities of providing focused training in Swedish tone-suffix associations with the help of a digital game. L2 learners at a low proficiency level in Swedish played a prototype of the game, which combined features of explicit and implicit learning conditions. The task constituting the core of the game mechanics required participants to make a correct choice between two alternative inflectional suffixes, one of which was validly cued by a previously presented word accent. Game performance measures as well as accuracy improvements in tone production were analyzed in order to assess the effectiveness of the training method and to gain insights into important aspects of learning an L2 predictive strategy facilitating morphological processing.
2. Background

2.1 Inflectional morphology

The morphology of a language concerns the structure of words, including the way new words are formed (e.g. player from play) or how various forms of a given word are constructed (e.g. play, played). The focus of the present thesis is on the brain processes associated with the latter phenomenon, inflectional morphology. Morphemes constitute basic linguistic units in the internal analysis of words, often defined as the smallest linguistic elements with a lexical or grammatical meaning (Booij, 2007). Morphemes might be free or bound: free morphemes can appear on their own as separate words, whereas bound morphemes always need to attach to a base morpheme, a ‘stem’. In case a free (e.g. play) or bound morpheme (e.g. -ful in playful) has a meaning of its own, which is not tied to the grammar of the language, it is referred to as a lexical morpheme. Other types of morphemes have a grammatical function instead, and these may also be either bound or free, for example the past tense affix -ed (e.g. played) and the preposition of (e.g. apple of Eve), respectively. An affix added after the stem is called a ‘suffix’ (e.g. -ed in played), and an affix preceding the word stem is a ‘prefix’ (e.g. re- in replayed). The part of the stem that cannot be subdivided into further morpheme units is called the ‘root’ (e.g. play is the root and the stem at the same time, whereas replay, to which -ed is added, is a stem).

A further central concept in inflectional morphology is the ‘lexeme’, which denotes the word as an abstract entity, e.g. APPLE. The actual word realizing a lexeme is referred to as a ‘grammatical word’ (Booij, 2007; Stump, 2001), e.g. apple (the singular of APPLE) or apples (the plural of APPLE). The same lexeme might be therefore spelled out as different word forms, depending on the specific syntactic context in which the word appears, or the information that the speaker intends to express (e.g. one apple, three apples). Rules of inflection specify how such different forms of a lexeme might be created, a process which involves marking the word for specific morphosyntactic properties. Morphosyntactic properties encode relations among different constituents of the sentence via morphological marking. For instance, nouns in many languages can be inflected by the morphosyntactic category ‘number’. In English, ‘number’ has two values, singular and plural, where plural is regularly expressed with the inflectional suffix -s. If the noun is in a subject-predicate relation with another constituent of the
finite sentence, this relation is expressed through ‘agreement’ in English, via co-variation of the inflectional feature of the constituents involved: a finite verb with a singular subject thus gets an -s suffix (e.g. *One apple taste+s good*), whereas with plural subjects the verb carries no overt inflection (e.g. *Three apples taste good*). Another example of a morphosyntactic category is ‘tense’, which marks the temporal reference of a finite verb. Verbs and nouns in many languages can be grouped into different inflectional classes, called ‘declensions’ for nouns and ‘conjugations’ for verbs. Members of the same class share a specific pattern concerning the way the various forms of the word are created by inflection (Booij, 2007).

Another relevant notion for the present dissertation is that of the ‘lexicon’. The lexicon of a language is an abstract entity, listing idiosyncratic word-related information: existing words as well as other established expressions of the language, such as idioms or affixes. Information specified in the lexicon concerns arbitrary signs (Aronoff & Anshen, 2001), where the forms and/or the form-meaning pairings are unpredictable to some extent, hence the need to list these in a repository. The representation of (part of) this lexical repository in the brain of an individual speaker of the language is called the ‘mental lexicon’ (Booij, 2007).

### 2.2 Dual route model of inflectional morphology processing

A general distinction between regular and irregular inflectional classes is a well-known property of the morphological description of various languages, notably English, a language that has often dominated the focus of psycholinguistics research (Clahsen, 2016). It is no wonder therefore that the most influential psycholinguistic theories on inflectional morphology processing were formulated with specific focus on accounting for the production and comprehension of these two types of complex words. Proponents of the so-called dual-system model assume two qualitatively different morphological processing mechanisms (Clahsen, 1999; Pinker, 1999; Pinker & Ullman, 2002): regularly inflected words are suggested to be processed through productive rule-governed compositional operations whereas irregular items are assumed to rely on full-form representations in the mental lexicon. From this perspective, the production of an inflected word form involves access to the mental lexicon as well as engagement of grammatical operations. If the inflected form is stored in memory, together with the relevant morphosyntactic property as part of its lexical specification, the word is retrieved and further rule-based computations are blocked. In the absence of a stored representation a default morphological rule applies, combining the word stem with the regular inflectional suffix (Pinker & Ullman, 2002). Alternatively,
connectionist single system approaches reject a distinction along the regular/irregular dimension (e.g. Rumelhart & McClelland, 1985) and argue that the same information, phonological and semantic regularities, would underlie the production of all inflected word forms (e.g. Joanisse & Seidenberg, 1999). From this perspective, apparent processing distinctions between regular and irregular words would stem from a solely qualitative difference: phonology plays a relatively greater role in the processing of regulars, where the verb stem and the base of the inflected form tend to be identical and the form of the inflection is often phonologically conditioned, as compared to irregulars, which rely more on semantic relations among words (Joanisse & Seidenberg, 1999).

The dual-system approach originally received support in studies that uncovered differences in priming effects for regular versus irregular word forms (e.g. Münte, Say, Clahsen, Schiltz, & Kutas, 1999; Sonnenstuhl, Eisenbeiss, & Clahsen, 1999; Stanners, Neiser, Hernon, & Hall, 1979). The observation that regularly inflected items (e.g. walked) facilitated the subsequent recognition of the base form (e.g. the word stem walk), as efficiently as presentation of the base form itself, were interpreted in favor of the assumption that such complex word forms are parsed into their component morpheme parts (walk+ed) during language comprehension, and lexical access takes place via the stem. Accordingly, stems were found to dominate meaning analysis of inflected word forms, assumed to be processed through decomposition, during a semantic decision task (Laine, 1999). Findings for irregular items were much more inconsistent and often indicated limited facilitation for word stem recognition or complete absence of priming effects, which might suggest that these forms have separate representations in lexical memory (Sonnenstuhl et al., 1999). Furthermore, a series of electrophysiological studies reported different brain responses to regular versus irregular word forms, in line with the assumption that the underlying neural mechanisms might be different as well (Gross, Say, Kleingers, Clahsen, & Münte, 1998; Morris & Holcomb, 2005; Münte et al., 1999; Penke et al., 1997; Rodriguez-Fornells, Clahsen, Lleó, Zaake, & Münte, 2001; Weyerts, Penke, Dohrn, Clahsen, & Münte, 1997). Importantly, a left anterior negativity, a brain response pattern commonly associated with morphosyntactic processing, has been typically observed in connection with the presence of a regular inflection or stem formation pattern in the input (Gross et al. 1998; Morris & Holcomb, 2005; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Weyerts et al., 1997).

Nevertheless, the situation seems rather more complex than a simple regular versus irregular distinction when data from a wider range of languages are considered, and further factors such as frequency of word forms in the input are taken into account. To begin with, it is reasonable to assume that frequent regular forms might get encoded in memory in full form in order to support rapid access to regularly encountered items (Pinker & Ullman, 2002). In this case, it has been argued that access to the stored representation versus decomposition might vary
for the same word depending on, for instance, the nature of the task (Pinker & Ullman, 2002). Task requirements that draw attention to the formal features of the language input might then motivate decompositional analysis, whereas reading/listening for comprehension could be thought to favor whole-word access. At the same time, whole word storage of frequent regular forms might be less extensive in highly inflected languages due to long-term memory limitations related to the large number of inflected variants for the same stem (Gor & Cook, 2010). Also, in these languages even words belonging to non-default classes may show tendencies towards decomposition depending on the complexity and the degree of regularity of specific inflectional patterns. For instance, compositional processing of verbs might be preferred regardless of conjugational class if word forms can be easily analyzed into their stem and inflection parts, as it has been argued for Russian, a language with rich verbal morphology and several verb classes displaying different degrees of regularity in their inflectional paradigms (Gor & Jackson, 2013). In sum, it can be assumed that the modulation of the different processing routes might depend on a complex interplay with the morphological properties of the language as well as with various speaker- and context-specific factors. The core question, whether two different neural mechanisms are available during word form processing in native speakers, nevertheless still remains unanswered. This issue was investigated in the context of Swedish inflectional morphology processing in Study 1 and 2.

2.3 Brain networks of language

In this section, we will provide a brief overview of some of the core brain areas and major processing streams involved in language comprehension and production. Next, we will turn to some proposals concerning the neuroanatomical underpinnings of dual route mechanisms within this language network. According to current theoretical perspectives, cognitive functions, including language, are implemented in large-scale networks in the brain, encompassing several anatomically connected brain regions as specialized processing nodes (Saur et al., 2008). Within these systems, further functional specialization among anatomically distinct processing streams has been observed. For instance, the cortical organization of vision is characterized by two major streams of projections from primary sensory cortical areas: a ventral (“what”) stream towards temporal regions, associated with visual object recognition; and a dorsal (“how”) stream terminating in parietal areas, supporting visually guided actions involving such objects (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). Separate dorsal and ventral auditory processing streams have also been described in nonhuman primates, originating in non-primary auditory cortical regions (Rauschecker &
Tian, 2000; Romanski et al., 1999). On analogy with the visual and the auditory system, language processing has been proposed to involve two parallel, but potentially interacting, streams, which connect temporal and frontal language-relevant regions (e.g. Hickok & Poeppel, 2004, 2007; Saur et al., 2008). Functionally, the ventral stream has been related to aspects of semantic and conceptual processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Hagoort, 2013) or more generally to language comprehension (Hickok & Poeppel, 2007; Saur et al., 2008). Its anatomical substrate was reported to encompass regions of the temporal cortex and those of the ventrolateral prefrontal cortex (Brodmann areas (BA) 45/47), which are connected mainly via a ventral pathway running through the extreme capsule (Saur et al., 2008). As regards the dorsal stream, it has been attributed with various functions, including sensory-motor transformations underlying speech production (Hickok & Poeppel, 2004, 2007), hierarchical sentence structure processing (Friederici, 2009, 2012) and time-dependent aspects of combining and ordering linguistic elements in successively larger structures (Bornkessel-Schlesewsky & Schlesewsky, 2013). Brain regions activated for a prototypical task of dorsal language processing (pseudoword repetition) involved temporal lobe and premotor areas (BA 6/44) connected by a dorsal tract via the arcuate and the superior longitudinal fascicle (Saur et al., 2008). For higher level linguistic processing, a further functional subdivision of the dorsal pathway has been argued, where connections between the temporal cortex and the pars opercularis of the inferior frontal gyrus (BA 44) in Broca’s area would support hierarchical syntactic analysis via the arcuate fasciculus (Friederici, 2012).

2.3.1 Neural substrates of dual route processing

Proposals that maintain a distinction between decompositional and full-form processing routes have associated these functions with separate subsystems within the language processing network, relating them to different processing streams as well (Bozic, Fonteneau, Su, & Marslen-Wilson, 2015; Marslen-Wilson & Tyler, 2007; Ullman 2001a, 2004). A detailed neurocognitive model has been formulated by Ullman (2001a, 2004) in the context of two general memory systems that have been extensively studied for non-linguistic functions in humans and non-human animals. According to this proposal, the mental lexicon, containing stored memory representations of word forms, is associated with the declarative memory system, whereas the mental grammar, underlying the combinatory processing of representations, is implemented in procedural memory (Ullman, 2001a, 2004). Declarative memory is largely subserved by temporal areas of the brain, representing knowledge about facts and events, including arbitrary word-related information (Ullman, 2004). Whole-word access is therefore associated with this
system (Pinker & Ullman, 2002). Also, retrieval of stored auditory word form representations was related to the superior temporal cortex, in a meta-analysis of neuroimaging studies, which indicated that analysis of the incoming speech input in terms of increasingly more complex sound patterns appears to proceed along a ventral processing pathway in the temporal cortex (DeWitt & Rauschecker, 2012). Medial temporal lobe structures such as the hippocampus play an important role in the learning and consolidation of declarative knowledge. Long-term storage is mostly implemented in neocortical areas of the temporal lobe and the processing of this knowledge has been linked to the ventral stream (Ullman, 2004, 2016). Procedural memory is constituted by a highly interconnected brain network, encompassing the basal ganglia and frontal cortical regions, as well as potentially parts of parietal cortex, superior temporal cortex and cerebellum (Ullman, 2004).

Procedural memory underlies the acquisition and performance of sensory-motor and cognitive skills, including the processing of complex linguistic representations with sequential and hierarchical structures (Ullman, 2004). Decompositional analysis of inflected word forms would consequently rely on this system (Pinker & Ullman, 2002). Whereas the basal ganglia appear to play a central role in the acquisition of procedural skills (Ullman, 2006a), performance of consolidated procedures might be largely tied to neocortical structures and the dorsal processing stream (Ullman, 2004, 2016). Broca’s area (BA 44 and 45), and specifically the pars opercularis of the inferior frontal gyrus (BA 44), has been proposed to be important in the selection and maintenance of representations in working memory during the processing of sequential or hierarchical linguistic structures (Ullman, 2004).

Marslen-Wilson, Bozic and colleagues have also described different neural subsystems for whole-word access versus the processing of word forms with decomposable internal structure (Bozic et al., 2015; Bozic & Marslen-Wilson, 2010; Marslen-Wilson & Tyler, 2007). Their proposal differs from the dual route account by Clahsen (1999) and Pinker (1999) in several respects, importantly concerning assumptions about the input properties that trigger analysis along a decompositional route: regularly inflected word forms need to be decomposed in order to enable access to lexical representations, a process which takes place via the word stem and the affix, separately. All word forms that are potentially segmentable based on morphophonological cues are automatically parsed into component morphemes, and, therefore, this operation is not tied to the presence of a regular inflection and an associated morphological rule application (Marslen-Wilson & Tyler, 2007). Nevertheless, the distinction between two separate neurobiological subsystems is maintained. Empirical evidence suggests that structural decomposition of complex word forms crucially depends on the left inferior frontal cortex and on its connections with posterior temporal lobe areas, implicating a dorsal decompositional network (Bozic et al., 2015; Marslen-Wilson & Tyler, 2007). Among the frontal regions, Broca’s area, in particular, the pars
opercularis of the inferior frontal gyrus has been repeatedly observed to be involved in regularly inflected word form processing (Bozic et al., 2015; Fonteneau, Bozic, & Marslen-Wilson, 2015; Tyler, Stamatakis, Post, Randall, & Marslen-Wilson, 2005). Access to lexical-semantic content associated with the stem, as well as to stored representations related to the affix, have been reported to take place in regions of the superior and middle temporal cortex (Marslen-Wilson & Tyler, 2007). Whereas the proposed network for decompositional analysis is lateralized to the left hemisphere, whole-word access is assumed to rely on a bi-hemispheric subsystem (Bozic & Marslen-Wilson, 2010).

To sum up, regions of the temporal lobe can be assumed to play an important role for whole-word access during complex word form processing, which is also in line with views on the general neurological substrates of lexical, semantic and conceptual representations (e.g. Vigneau et al., 2006). The left inferior frontal gyrus, specifically pars opercularis, might be crucially involved in processes of decompositional analysis of inflected word forms. Study 2 examined this frontal versus temporal distinction in the context of Swedish tone processing in inflected nouns, scrutinizing specifically the role of cortical thickness in a frontal region, the pars opercularis of the inferior frontal gyrus, and in the planum temporale, which is a non-primary auditory region, constituting the core part of classical Wernicke’s area (Hickok & Saberi, 2012) in the temporal lobe. Importantly, encoding of intonational pitch contours, irrespective of speaker-specific pitch variations, has been localized to higher-order auditory areas including the planum temporale (Tang, Hamilton, & Chang, 2017).

2.4 Swedish morphology

In this section, we will provide a brief descriptive introduction to some of the characteristic features of Swedish morphology, relevant to the language materials used in the investigations. In terms of its general morphological properties, Swedish shares several characteristics with analytic languages of the world. Such languages typically use separate words or word order to express grammatical and semantic relations, which might otherwise be marked by affixation. Thus, in Swedish syntactic relations (e.g. subject, object) are indicated by word order, and function words, e.g. prepositions are used in several contexts to mark a number of grammatical and semantic functions, for instance in adverbials of time and space (e.g. till Lund ‘to Lund’, i affären ‘in the shop’) (Teleman, Hellberg, & Andersson, 1999). Nevertheless, even if Swedish does not have a rich inflectional system comparable to languages characterized by free word order, it does express several grammatical and semantic relations through inflection, creating morphologically complex words. For example, Swedish uses affixation for the morphosyntactic
categories of ‘number’ and ‘tense’. It even has some more unusual inflectional categories, which include marking definiteness on nouns as well as passive voice on verbs using suffixation. The inflectional system of Swedish is largely agglutinating: each suffix is associated with a specific meaning, and the boundaries between the morphemes as well as between the stem and the affixes tend to be clearly identifiable (e.g. pojk+ar+na+s: boy+PL+DEF+POSS, i.e. ‘the boys’) (Teleman et al., 1999).

Each study in the present thesis focused on the processing of specific inflected word forms, which were either verbs or nouns, placed in sentence context. Relevant for the experimental manipulation used with nouns, Swedish morphology distinguishes between singular and plural forms, which largely corresponds to the semantic distinction ‘one versus many’ for count nouns (Teleman et al., 1999). As mentioned above, Swedish uses suffixation to mark definiteness on nouns, e.g. bil versus bil-en ‘car’ versus ‘the car’. When the noun is indefinite, only the plural form carries an overt suffix: e.g. bil versus bil-ar, ‘car’ versus ‘car-PL’. For definite forms, however, the suffix added varies with the number property, and even the singular noun gets an inflection: bil-en versus bil-ar-na, ‘bil-SG+DEF’ versus ‘bil-PL-DEF’. As a result, the morphological processing of the semantic distinction “one versus many” can be studied by contrasting comparable word forms, each made up of a word stem plus one inflectional suffix: bil-en versus bil-ar.

Finite Swedish verbs are inflected for tense, which is either a present or a preterite form. Verbs appearing as target words in the stimulus materials of the investigations constituted such morphologically complex forms. The present tense form is mainly used to indicate that an action takes place in the present or in the future, whereas the preterite denotes an action associated with a specific time point in the past. Further temporal aspects can be expressed with multi-word structures including an auxiliary verb, which will be not discussed here (Teleman et al., 1999).

Swedish verbs can be divided into several conjugational classes, largely based on the type of suffix they take in the preterite form. The main lexically specified distinction is between ‘weak’ and ‘strong’ verbs, where weak verbs are considered to be regular. The preterite of weak verbs is formed by attaching a -de/-te/-dde suffix to the stem. This group of words can be divided into three conjugational classes (1st, 2nd and 3rd conjugations) depending on the phonological form of the stem-final segment, which in turn specifies the suffix-variant to be used to build the present and the preterite form (Teleman et al., 1999). Thus, the subdivision within the weak verb class is based on a phonologically conditioned, and as such predictable, suffix-variation. Of the weak verb classes, 2nd conjugation verbs were used as target words in Study 1 and 3. These verb stems end in a consonant, and the suffix attached to the stem is -de/-te for the past tense (-de after voiced and -te after voiceless stem-final consonant), e.g. häll-de ‘pour-ed’, tänk-te ‘plann-ed’.
Strong verbs constitute the 4th conjugational class. Unlike in the case of weak verbs, the preterite of strong verbs is formed without attaching a tense suffix and it normally involves a vowel-change in the stem. These vowel alternations are typically not arbitrary but follow specific patterns, which can be often predicted from the vowel of the verb stem (the stem of the infinitive and present tense form). The vowel sequences associated with the most common inflectional subclasses are exemplified in Table 1, with present, preterite and supinum verb forms (supinum is used, for example, in constructions with auxiliary verbs) (Teleman et al., 1999). Therefore, members of the 4th conjugational class can be regarded as interesting intermediate instances between regularly and irregularly formed verbs. On the one hand, they are not completely regular to the same extent as the Swedish weak verbs or English regular verbs (e.g. talk-ed) are: the construction of the preterite cannot be described in one simple rule that automatically attaches a default past tense suffix to the stem, which is left largely unchanged by the operation. On the other hand, the 4th conjugational class preterite forms are not random either, and, even if the underlying regularities are somewhat more complex, specific predictable patterns clearly exist. The way the morphological processing system in Swedish native speakers handles these verb forms, relative to the 2nd conjugational verb class, is investigated in Study 1.

Table 1
Regular vowel sequences associated with some of the most common inflectional subclasses in the 4th conjugational class.

<table>
<thead>
<tr>
<th>Vowel Sequence</th>
<th>Present</th>
<th>Preterite</th>
<th>Supinum</th>
</tr>
</thead>
<tbody>
<tr>
<td>i [i] – a – u</td>
<td>binder</td>
<td>band</td>
<td>bundit</td>
</tr>
<tr>
<td>i [iː]– e – i</td>
<td>skriver</td>
<td>skrev</td>
<td>skrivit</td>
</tr>
<tr>
<td>u/y – ö – u</td>
<td>bjuder; flyger</td>
<td>bjöd; flög</td>
<td>bjudit; flugit</td>
</tr>
</tbody>
</table>

The present tense of both weak and strong verbs is formed according to the same regularities, determined by the phonological form of the verb stem in the majority of cases. For example, stems ending with a vowel receive the suffix -r (prata-r ‘work-PRES’), whereas stems ending in a consonant (with the exception of long vowel + ‘r’ or ‘l’) get the suffix -er (e.g. häll-er ‘pour-PRES’, skriv-er ‘write-PRES’) (Teleman et al., 1999).

2.5 Swedish word accents

A characteristic feature of spoken Swedish is the presence of a word accent on each prosodic word. The two word accents of the language are called ‘accent 1’ and ‘accent 2’ (Bruce, 1977). In Central Swedish, which is the dialect that was used in the investigations, accent 1 appears as a low tone on the stressed syllable...
whereas accent 2 is realized as a high tone. Importantly, the word accent associated with the same stem may vary depending on the suffix attached. For example, the word stem *bil ‘car’* receives a low tone (accent 1) when it is followed by the singular definitive article *-en* and a high tone (accent 2) if it ends in the plural suffix *-ar*. According to one line of analysis, this apparently suffix-induced variation is related to the fact that the lexically specified tone, accent 2, is associated with a set of derivational and inflectional suffixes, including the plural suffix *-ar* (Bruce, 1977; Riad, 2015; Rischel, 1963). When these suffixes combine with other morphemes, they assign their associated accent 2 to the stressed syllable, which is most often the root syllable. In case there is no lexical tone present, or its realization is inhibited by well-formedness constraints on the structure, accent 1 is assigned as the default (Riad, 2015). From this perspective, accent 1 is intonation, whereas accent 2 is stored in the mental lexicon, as part of the information associated with the representation of specific suffixes. Accent 2 might nevertheless also be assigned postlexically, mainly in compound words, where the appearance of accent 2 is motivated by the presence of secondary stress (Riad, 2015).

In the present dissertation, the association between word accents and suffixes is investigated from a speech processing perspective. During actual speech comprehension, the tone realized on the word stem precedes the suffix and, therefore, could be assumed to function as a predictive cue to an upcoming word completion regardless of the lexical or postlexical status of the tone (Roll et al., 2015). Hearing the word *bil* with accent 2, the listener might expect a continuation with the plural suffix *-ar*, *bil*<sub>accent2</sub>-ar, due to the association between the tone and the suffix. At the same time, pronouncing the same word stem with accent 1 makes a continuation with *-ar* highly unlikely, whereas the suffix *-en*, as in *bil*<sub>accent1</sub>-en, is much more expected. In other words, if accent 2 might be assumed to cue its associated suffixes, accent 1 could be thought to similarly cue another set of suffixes, those that are not associated with accent 2. We therefore did not make an explicit distinction between accent 1 as the default, and accent 2 as the suffixed-induced tone in the investigations.

### 2.6 The processing of word accent-suffix associations in native speakers

There is substantial empirical evidence indicating that native speakers of Swedish rely on word accents to anticipate upcoming word endings during online language comprehension. Relevant results come from a range of behavioral, electrophysiological and brain imaging studies (Gosselke Berthelsen et al., 2018; Roll, 2015; Roll et al., 2010, 2013, 2015; Söderström, Horne, Mannfolk, van
Westen, & Roll, 2017; Söderström et al., 2012, 2016b). When listeners were asked to judge suffix meaning in morphologically complex words in a response-time experiment, they were observed to process faster suffixes that were preceded by their associated word accent on the word stem, relative to suffixes that were presented after a different stem tone than the one they are related to (Söderström et al., 2012). Thus, participants decided more rapidly between the present tense versus past tense meaning of words such as lek\textsubscript{accent2}-te ‘played’, pronounced with the word accent associated with the past tense suffix -te, i.e. accent 2, as compared to verbs such as lek\textsubscript{accent1}-te, presented incorrectly with accent 1 on the stem. The same pattern was observed for the present tense suffix -er, associated with accent 1. Since all the experimental items were grammatically correct and semantically plausible complex words, it was arguably the word accent that generated an expectation in listeners for a likely upcoming continuation. When the actually experienced suffix in the input disconfirmed this expectation, processing time increased as reflected in longer response latencies.

Next, consistent with the assumption that word accents function as predictive cues to their associated suffixes, Swedish tones were reported to generate an increased negativity in the electrophysiological brain response, referred to as PrAN, around 136 ms after tone onset on the word stem (Roll et al., 2015; Söderström, Horne, Frid, & Roll, 2016a). This negativity was greater for accent 1 than for accent 2, which seems to stem from a difference between the two word accents concerning their predictive significance. As mentioned above, compound words receive accent 2 in Central Swedish. Therefore, hearing a word stem with accent 2 opens up possibilities for an almost unlimited number of continuations as compound words (e.g. bil-nyckel ‘car-key’, bil-dörr ‘car-door’, bil-bälte ‘car-belt’ etc.). However, word endings that might follow accent 1 constitute a much smaller and well-defined set of suffixes (e.g. bil-en). Accent 1 is therefore associated with more predictive certainty and can be assumed to pre-activate related suffixes to a greater extent, hence the enhanced negativity (Roll et al., 2015). Further studies showed that the negativity indeed increases as the number of possible word completions decreases and the more frequent words those completions constitute, in line with its interpretation as an index of predictive activation of memory traces, modulated by the certainty with which a specific continuation might occur (Roll, Söderström, Frid, Mannfolk, & Horne, 2017; Söderström et al., 2016a). Observing a PrAN for word accents in native speakers thus clearly supports the idea that Swedish stem tones generate expectations for upcoming suffixes.

The PrAN effect was obtained even for processing tones on inflected pseudoword stems, i.e. on items that were phonotactically legal word forms, but not actually existing words in the Swedish language (Söderström et al., 2016b, 2017). Participants were also able to guess the identity of suffixes in cases when the actual pronunciation of the word ending on the pseudoword stem was replaced with a cough, and the only cue to the ending present was the word accent. The fact
that native speakers seem to be able to make use of the predictive association between word accents and suffixes even in the absence of lexical content in the word stem suggests that a more abstract association exists between tones and suffixes, independently of stored lexical representations (Söderström et al., 2016b).

Neuroimaging studies have identified a left-lateralized brain network involved in the predictive use of word accents during speech comprehension (Roll et al., 2015; Söderström et al., 2017). Correlating functional brain activations with electrophysiological measures has provided insights into the time course of processing supported by different parts of this network (Roll et al., 2015). Based on these results, tones on inflected Swedish words are thought to be initially processed in temporal lobe areas, where discrimination of tone patterns in primary auditory cortex is followed by access to associated phonological representations of word accent categories in the superior temporal gyrus. Following word accent recognition, increased activation has been observed in frontal brain regions such as the inferior frontal gyrus (IFG), most likely related to some aspects of processing the suffix prediction generated by the tone (Roll et al., 2015). Specifically, left IFG activation associated with PrAN has been suggested to underlie lexical selection of likely word completions and inhibition of competing alternatives (Roll et al., 2017). Interestingly, when tones were realized on pseudowords, temporal areas have shown less prominent activations, and the pars opercularis in the left IFG has emerged as an important processing center instead (Söderström et al., 2017). This is different from the frontal activation pattern observed in Roll et al. (2015), where the area associated with the strongest activation in the IFG was BA 47, potentially indicating involvement of the ventral processing stream (e.g. Saur et al., 2008). It seems therefore that the involvement of frontal versus temporal nodes of this tone processing network might be modulated by the presence versus absence of lexical information in the word stem carrying the word accent. This hypothesis was further investigated in Study 2.

2.7 Investigating second language acquisition of Swedish tone-suffix associations

Swedish native speakers have been shown to rely on word accents to facilitate processing of upcoming word structure, but it is still unknown if adult second language (L2) learners of Swedish can acquire the morphological function of tones. This question was investigated in Study 3 and 4. Here we will discuss some of the basic assumptions and considerations related to L2 acquisition of tone-suffix associations that guided the focus of these investigations as well as some methodological choices in the studies.
2.7.1 Tone perception in L2 acquisition

It is reasonable to assume that learners’ ability to effectively exploit the predictive significance of Swedish word accents depends on the accurate perception of accent 1 versus accent 2 tone distinctions. Research on the second language acquisition of tonal aspects has largely focused on the learning of tone languages such as Mandarin Chinese, where, unlike in Swedish, each syllable is associated with a distinctive tone. Despite significant differences concerning the status of tones in these languages, results from lexical tone studies are indicative of certain factors that are expected to play a role in Swedish word accent acquisition as well, most importantly previous language experience. Studies on cross-linguistic tone processing have repeatedly found that native speakers of tone versus non-tone languages differ in their ability to discriminate tonal contrasts in an L2 (Francis, Ciocca, Ma, & Fenn, 2008; Lee, Vakoch, & Wurm, 1996; Wang, 2013; Wayland & Guion, 2004). In this context, one theoretical model that is frequently referred to is Flege’s (1995) speech learning model (SLM), applying its predictions originally formulated to involve segmental acquisition to the suprasegmental level. Importantly, the SLM argues against the existence of maturational constraints on L2 learners’ ability to establish new phonetic categories. However, as the native language (L1) and the L2 categories are assumed to occupy the same representational space, the developmental state of the different subsystems and the relative similarity of the already established sound representations influence the way new L2 sounds are integrated. Therefore, learners are likely to create a new category for an L2 sound that is perceived to be markedly different from any L1 representation, whereas an L2 sound that is perceptually similar to a fully-developed L1 category tends to be processed through the already established long-term memory representation of the L1 sound. Even in these cases, learners might be able to recognize subtle auditory differences between the two sounds, and the shared memory representation will be gradually modified to incorporate the L2 sound features.

Applying this perspective to the acquisition of suprasegmental features, the existence of native tonal categories can either constrain or facilitate the accurate representation of L2 tonal contours, depending on the degree of similarity between specific tonal categories in the L1 and the L2. Also, speakers of non-tone languages should presumably be able to form representations for L2 tonal patterns, even in the absence of comparable phonetic categories in the L1. Still, it needs to be considered that pitch contours (fundamental frequency ($f_0$) patterns) are generally used to convey a variety of meanings in non-tone languages as well, for instance in sentence level intonation. Therefore, lexical tone processing might be influenced by learners’ experience with native intonational categories (e.g. a rise in English yes-no questions), which, in contrast to lexical tones, lack associations with specific items at the syllable or morpheme level. Nevertheless, even if
learners might not be able to make the sharp distinctions that are necessary for the
native-like identification of tonal categories (Hallé, Chang, & Best, 2004) at the
initial stages of learning, several studies have reported significant improvement in
the perceptual identification of lexical tones as a result of focused training
involving native speakers of non-tone languages such as English (Francis et al.,
2008; Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno,
1999). Training has been argued to lead to the establishment of long-term memory
representations for tonal categories in these learners (Wang et al., 1999) and to
increased accuracy in production (Wang et al., 2003). Relevant findings on L2
perception of Swedish word accents indicated that, in the absence of training,
beginner learners with a non-tone L1 processed Swedish tones non-linguistically,
despite their presentation on words in sentence context (Gosselke Berthelsen et al.,
2018). These results might indicate that learners have dissociated word accents
from the L1 function of comparable tonal patterns (a pragmatic function in polite,
soothing requests in German), which might be considered as a prerequisite for
acquiring the morphological significance of Swedish tones. Also, a subset of
relatively more proficient learners showed signs of developing greater sensitivity
to pitch differences, which could be expected to support the creation of memory
representations for tonal patterns (Gosselke Berthelsen et al., 2018).

The above considerations suggest that there do not seem to be absolute
maturational constraints on learning to discriminate tones in an L2, which can be
assumed to be a prerequisite to acquire the ability to rely on the predictive
significance of word accents. Nevertheless, providing perceptual training might be
beneficial for facilitating the acquisition process. Further, tone versus non-tone L1
background appears to be an important factor, and acquiring Swedish word
accents is expected to constitute different kinds of challenges for these different
learner groups. Native speakers of tone languages have extensive experience in the
accurate detection of $f_0$ variation patterns at the word level, which they might be
able to utilize for the discrimination of L2 tonal contrasts as well (Wayland &
Guion, 2004). Relative to non-tone L1 speakers, these learners might process word
accents as more salient and relevant linguistic features due to their predisposition
to direct attention to $f_0$ at the word level, which could possibly provide an
advantage in the acquisition of the tone-suffix associations. Even in these cases,
however, learners would presumably need to overcome the tendency to interpret
word accents in terms of native tone categories and learn to associate them with a
morphological function instead of a lexical one. As for learners with a non-tone L1
background, they would need to establish tonal categories that are well-defined
enough to underlie native-like discrimination of word accent patterns. Also, the
development of effective perceptual strategies is necessary to be able to track the
relevant features of the $f_0$ contour at the word level and to learn to form
associations between these pitch contours and word accent categories. Considering
these clear differences, studies in the present thesis focused exclusively on one of
these groups, L2 learners with a non-tone background, who might also be expected to find learning to discriminate word accents relatively more challenging.

2.7.2 Implicit and explicit L2 learning

Swedish tone-suffix associations are normally largely neglected in formal L2 instruction, as can be judged from the general absence of this L2 feature in Swedish language course curriculums and course books. It is also an aspect of the language that native speakers are generally unaware of. Consequently, it can be assumed that the average L2 learner of Swedish has not received any instruction or explicit information about tones as cues to word structure, and a relevant question is if it is possible for late L2 learners to acquire this feature of the language without such explicit training. Generally, learning is considered implicit when “we acquire information without intending to do so, and in such a way that the resulting knowledge is difficult to express” (Cleeremans, Destrebecqz, & Boyer, 1998, p. 406). Explicit learning is an intentional process, which usually involves conscious hypothesis testing (Cleeremans et al., 1998).

Certain aspects of L2s have been previously found to be learned rapidly solely through exposure, including word boundaries (Saffran, Newport, & Aslin, 1996), lexical information (Gullberg, Roberts, & Dimroth, 2012), grammatical word categories (Mintz, 2002), lexical subcategories such as gender classes (Sandoval, Patterson, Dai, Vance, & Plante, 2017) and morphological regularities (De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007). Also, there is some indication that implicit learning conditions that engage distributional learning mechanisms might specifically promote the development of strong enough associations between representations of language features, which would then underlie predictive processing. For instance, Grüter, Lew-Williams and Fernald (2012) suggested that for L1 learners of Spanish, the only way to discover the gender class of specific nouns from the input is by focusing on co-occurrence relations between nouns and gender-marked determiners. This distributional learning mechanism naturally results in strong associations between nouns and determiners in the mental lexicon, which, even after the development of a more abstract representation of gender information, enable native speakers to rely on gender cues in articles to anticipate upcoming nouns. Adult L2 learners, however, tend to have access to a much wider range of information concerning the gender class of different nouns (e.g. metalinguistic information), and are unlikely to rely on distributional learning to any comparable extent in typical learning situations. Indeed, highly proficient L2 learners were found to display native-like predictive processing only with novel words that were acquired under learning conditions that specifically promoted reliance on co-occurrence information between determiners and nouns (Grüter et al., 2012).
Nevertheless, as for the Swedish word accent-suffix associations, the available evidence suggests that mere exposure to the language might not be sufficient for acquisition at earlier stages of language learning in speakers of non-tone L1s: beginner to early intermediate German learners of Swedish (A1 to B1 level of the Common European Framework of Reference) showed no signs of implicit acquisition of the morphological function of word accents in their electrophysiological brain responses (Gosselke Berthelsen et al., 2018). Certainly, some degree of implicit acquisition of the L2 feature might eventually take place after a longer period of exposure when learners have reached higher levels of proficiency. Since no previous study has explored word accent-suffix association processing in more advanced learners, this question was investigated in Study 3. Importantly, the ability to use prosodic cues to anticipate word endings in online speech processing seems to develop with increasing L2 proficiency: advanced L2 learners of Spanish were found to rely on lexical stress to predict an upcoming present tense versus past tense suffix in a visual-world eye-tracking experiment (Sagarra & Casillas, 2018). These learners had presumably acquired the relevant prosody-suffix connections as well as the native-like use of lexical stress for suffix anticipation implicitly, despite the fact that the same prosodic feature had a weaker functional load in their native language, English. Beginner L2 learners, however, did not make use of stress cues for predictive processing of likely word endings (Sagarra & Casillas, 2018). Furthermore, it might be also assumed that advanced L2 learners are in principle able to gain native-like tone-suffix processing mechanisms, even if the grammar of their native language is not characterized by any similar tone-morphology associations. For instance, Sagarra and Herschensohn (2010) investigated L2 acquisition of gender-number agreement between nouns and adjectives in Spanish, in adult native English speakers who did not have comparable grammatical features in their native language. Gender marking has been observed to facilitate processing of upcoming items with congruent gender features in native speakers of Spanish and other gendered languages (e.g. Wicha, Moreno, & Kutas, 2004). As for L2 acquisition, intermediate learners, but not beginners, were sensitive to gender and number agreement violations during an online language-processing task, displaying increased response latencies to sentences with mismatching items, in a similar manner to native speakers. Apparently, intermediate learners had started to develop mental representations for grammatical features absent in their native language, and were able to rely on this knowledge for the online computation of adjective agreement during language comprehension (Sagarra & Herschensohn, 2010, 2012).

As mentioned in the previous section, learners with non-tonal L1 background are unlikely to be characterized by an initial predisposition to pay attention to the functional significance of word accents, especially considering the fact that stem tones are not indispensible for recovering meaning from suffixes. Sagarra and Ellis
(2013), for instance, argued that “learned attention”, i.e. the kind of cues that the L2 learner is accustomed to direct attention to as a result of experience with the L1 is an important factor determining the success with which a certain aspect of the L2 is acquired. Failure to attend to word accents might thus partly explain the apparent lack of automatic acquisition of word accent-suffix associations at lower L2 proficiency levels (Gosselke Berthelsen et al., 2018). Also, attention has been argued to play a significant role in general co-variation learning (Hoffmann & Sebald, 2005) and more specifically in L2 acquisition: according to the ‘Noticing Hypothesis’ (Schmidt, 2001), the registration of some stimulus in focal attention is the initial step in the acquisition process. Once some language feature has been encoded in memory, it may undergo further unconscious processing and non-conscious activations during the emergence of a specific language skill. Indeed, following implicit training in a miniature language, those learners who reported to pay conscious attention to grammar during training performed better at the grammaticality judgment task than those participants who only focused on lexical aspects of the input (Batterink & Neville, 2013).

A further relevant observation is that significant variation in learning outcome has been reported in studies examining artificial language and grammar learning under implicit conditions (e.g. Franco, Cleeremans, & Destrebecqz, 2011; Misyak & Christiansen, 2012; Morgan-Short et al., 2015), indicating that there might be individual differences as regards the ability of adults to pick up linguistic regularities through mere exposure to the input. While some L2 learners might learn successfully and rapidly under purely implicit conditions, others would possibly perform poorly without some explicit information to aid the acquisition process. Importantly, it has been argued that native-like or near-native procedural language processing mechanisms might emerge from the initial explicit learning of relevant rules and meanings, as a result of gradual automatization of declarative knowledge through extensive practice (DeKeyser & Criado-Sánchez, 2012).

Based on the above considerations, L2 learners of Swedish might benefit from some form of structured training, which combines features of implicit and explicit learning conditions, to help the acquisition of the morphological function of tones. First, extensive exposure to word accent-suffix co-variations, in a context that makes these dependencies relevant for the objectives of the learner, might promote the development of strong associations between suffixes and specific tones as predictive cues. Second, a more explicit element, such as a task to be performed, might be necessary to direct learners’ attention to word accent variations and their significance in cuing morphological structure. Descriptions of rules for the successful performance of the task might be made available for learners who require more explicit information. Such a learning tool is described in Study 4 in detail, and its efficiency in promoting the acquisition of word accent-suffix associations is tested with low proficient L2 learners.
3. Methods

3.1. Response times

The idea that response-time (RT) experiments can provide an insight into mental processes dates back to the investigations of F. C. Donders in the middle of the 19th century. Inspired by Helmholtz’s work on measuring the transmission speed of nerve impulses, Donders aimed at identifying the time required for the completion of different hypothesized processing stages of a mental task (Van Zandt & Townsend, 2012). With the advance of modern cognitive psychology and the development of increasingly precise time measurement techniques after the 1950s, RTs have become a crucial and extensively used dependent variable in psychology (Luce, 1991). In an RT experiment, the time it takes for participants to respond to a stimulus is measured, usually with an accuracy of a few milliseconds. RTs are generally assumed to reflect the duration of mental processes (Ratcliff, 2012), and in this sense their functional significance is quite clear: a specific increase in response latency for a stimulus presented in one experimental condition indicates how much longer the encoding, processing and acting on a stimulus takes in that condition relative to another (Luck, 2005). RTs are thus also often seen as a measure of processing ease or difficulty. Nevertheless, an important limitation of the method is that a single overt response might be an outcome of a multitude of underlying cognitive operations, and response latencies give little indication as to the nature of the specific mental processes involved (Luck, 2005). In order to gain insight into the “black box” of the mind, one might directly record the brain’s response to a stimulus or event, using electroencephalography (EEG) and the event-related potential (ERP) technique.

3.2 EEG and ERPs

EEG measures the electrical activity of the brain, picked up by electrodes placed on the scalp. From the on-going EEG, brain responses related to specific sensory or cognitive events can be extracted, by averaging over many trials of stimulus presentation associated with a given experimental manipulation (Luck, 2005). The resulting ERP waveform constitutes a series of positive- and negative-going
voltage deflections, as exemplified in Figure 1. Conventionally, negative voltages are plotted upwards. A voltage deflection associated with a given neural process occurring with a specific spatial distribution is referred to as an ERP component (Luck, 2012). Characteristic distribution of such ERP effects over the scalp are often illustrated with topographic maps, which show the voltage measured at specific electrode locations during the time window of the component on a color scale, interpolating values between the recording sites.

Figure 1
ERP waveforms (left) and topographic maps (right) from Study 1. Irregular incorrect verbs elicited a LAN effect 350-500 ms after verb presentation onset, shown at the left frontal electrode site F3 (top left). A P600 effect was also observed for irregular incorrect verbs at 600-1000 ms, exemplified at the posterior central site Pz (bottom left).

Whereas response times reflect the output of cognitive processes, ERPs can be used to track the modulation of neural activity as it happens, due to the excellent time resolution of the technique, and the observed ERP components can give information about the nature of cognitive processes underlying a specific behavioral response. Also, ERPs can be used to study neural processes that are not reflected in overt behavioral changes or responses (Luck, 2012). For instance, L2 learning effects might be detected in the electrical brain response even before any improvement in performing the L2 skill would occur and be observable in
behavioral measures (e.g. van Hell & Tokowicz, 2010). Furthermore, knowing the exact onset of mental operations associated with a stimulus, down to a few milliseconds accuracy, can help establish whether a specific experimental manipulation modulates early sensory activity or later higher-level processes (Luck, 2012). In the context of language processing, the timing of observed ERP effects could be indicative of the involvement of rapid automatic processes versus late and controlled language comprehension mechanisms. Nevertheless, one important limitation of the ERP technique is its poor spatial resolution: as electricity in the brain is conducted through the tissue between the generating neural population and the surface, it spreads out and gets further diverted by the high resistance of the skull. This property makes it difficult to draw conclusions concerning the brain regions where neural operations take place. Based on solely the observation of a specific topographic distribution, it is not possible to localize the ERP generators, since the so-called inverse problem has no perfect solution: any given voltage distribution on the scalp might in principle correspond to an infinite number of different brain sources (Luck, 2005). Furthermore, it is important to be aware of the fact that voltages recorded on the scalp reflect only part of the electrical activity that occurs in the brain. The current that eventually reaches the surface originates as postsynaptic potentials generated during neurotransmission, which must occur simultaneously in large populations of spatially aligned neurons in order to produce a measurable signal on the scalp. Based on these considerations, EEG is most likely the reflection of synchronized activity of pyramidal cells in the cortex (Luck, 2005, 2012).

Typical EEG recording and ERP analysis procedures are described below to provide a background to the methods used in the electrophysiological study on Swedish word form processing, Study 1. During an EEG experiment, voltages from a participant’s scalp are picked up by electrodes mounted on an elastic cap or net. These weak signals are then amplified and recorded in a digital format, as a series of voltage values corresponding to specific time points. The recorded EEG at this point is a mixture of brain activity as well as noise from various sources, including potentials generated by eye or muscle movements and external electrical appliances. Filtering during and after recording is often used to remove irrelevant parts of the signal that fall outside of the frequencies of interest associated with actual cognitive processes. Also, signal processing techniques have been developed to automatically remove common artefactual activities such as blinks. The averaging procedure typically used to extract ERP waveforms from the EEG relies on the following considerations: brain activity related to a stimulus is assumed to be more or less constant each time the given stimulus is presented, whereas unrelated noise is expected to vary from trial to trial in a random manner. Therefore, segments of EEG around the event of interest are extracted and aligned to the same time point relative to stimulus presentation. As many of these segments, also called epochs, are averaged together, randomly varying positive
3.2.1 ERPs in language processing

Three common language-related ERP effects are introduced in this section, which are also relevant for interpreting the results of the electrophysiological Study 1. These effects were originally observed and described as responses to various semantic or structural anomalies in language materials or to language input with increased complexity. By studying the brain’s response as language processing fails or becomes effortful due to specific manipulations, one can even gain insight into the nature of neural mechanisms during normal language comprehension, associated with the processing of features affected by the given experimental manipulation.

3.2.1.1 LAN

Different types of violations related to inflectional morphology have been observed to elicit a left anterior negativity (LAN) in the electrophysiological brain response, typically between 300-500 milliseconds (ms) following presentation of the incorrect word or morpheme (see Figure 1 (top)). The name of the ERP effect reflects its characteristic distribution on the scalp, with maxima over anterior (frontal) electrode sites on the left side of the brain. In addition, LAN is referred to as a ‘negativity’ since it constitutes a negative voltage deflection in the ERP waveform, relative to some control condition. For instance, LAN has been regularly observed for errors in number agreement between sentence constituents (Molinaro, Barber, & Carreiras, 2011), including incorrect inflectional marking of subject-verb agreement with the suffix -s in English, e.g. The elected officials *hopes to... (Osterhout & Mobley, 1995). Also, word forms to which productive inflectional suffixes were incorrectly applied have been reported to elicit LANs, such as in *bring+ed (Gross et al. 1998; Morris & Holcomb, 2005; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Weyerts et al., 1997). Generally, LAN has been interpreted as a signal of morphosyntactic anomaly detection (Friederici, 2002), or more specifically as a response to violations of regularities related to morphosyntactic structure building (Penke et al., 1997). It has been associated with sentence structure analysis processes operating on morphologically expressed cues, such as the presence of decomposable inflectional morphemes in constituents (Molinaro et al., 2011).
3.2.1.2 N400

Words that are structurally well-formed but semantically anomalous in the given context tend to elicit a different ERP effect, a so-called N400, which is a negative deflection peaking around 400 ms after stimulus presentation, typically displaying a centro-parietal scalp distribution. The N400 was first described as a response to sentence-final words with meanings that mismatched the sentence context such as *He spread the warm bread with socks* (Kutas & Hillyard, 1980). Later it became clear that the processing of each word in a sentence might be associated with an N400, but the magnitude of the component is affected by various manipulations. For instance, words with low frequency of occurrence in the language elicit larger N400 effects than more frequent items (Münte, Urbach, Düzel, & Kutas, 2000). Even pseudowords (pronounceable nonwords) yield N400s, which tend to be large for non-repeated presentation of pseudoword items in lists or word pairs (Kutas, Van Petten, & Klueener, 2006).

A number of different views exist concerning the processes that the N400 component reflects. For instance, the N400 has been associated with semantic unification, involving the integration of the meaning of a lexical item into a larger semantic representation constructed on the basis of the preceding context (Hagoort, Baggio, & Willems, 2009). Greater difficulties with this integration process would be then expected to yield larger N400 amplitudes. Alternative perspectives relate the N400 to accessing information in long-term semantic memory (Kutas & Federmeier, 2000, 2011). From this perspective, modulation of N400 might be related to the activation state of the semantic memory system, where larger changes generate greater N400 responses. However, if features corresponding to the meaning representation of an item are already more or less activated in memory, due to, for instance, processing a related sentence context, lexical access does not induce a significant increase in activation state, moderating the observed N400 amplitude (Kutas & Federmeier, 2011).

3.2.1.3 P600

A later, positive-going ERP effect with a largely posterior scalp distribution has been observed for syntactic anomalies and complex sentence structures (see Figure 1 (bottom)). The P600 has an onset around 600 ms, and it might appear after earlier language-related ERP effects in certain cases, constituting a biphasic pattern. For instance, a P600 has been reported for incorrect use of inflectional marking on sentence constituents (e.g. the past tense suffix *-ed* on an irregular verb *bringed*), following a LAN (Morris & Holcomb, 2005). Presenting an intransitive verb (e.g. *departed*) that does not take an object argument (*departed the banker*) together with an object noun in German has been shown to result in an enhanced N400 as well as a P600 effect (*Anna weiß, dass der Kommissar (NOM) den Banker (ACC) abreiste (V)*... 'Anna knows that the inspector (NOM) the
banker (ACC) departed (V).’) (Friederici & Frisch, 2000, p. 481). Completely well-formed sentences might produce a P600, if the reader/listener is forced to reanalyze an originally preferred structural interpretation at some point, such as at the auxiliary verb was in the garden path sentence The lawyer charged the defendant was lying (Osterhout, Holcomb, & Swinney, 1994). The P600 has thus been argued to reflect language comprehension processes associated with reanalysis and repair of syntactic structures (Friederici, 2002). Nevertheless, the P600 has even been reported for language materials where apparently no structural reanalysis is necessary: long distance dependencies, where related constituents – for instance the subject, object and the verb – are separated from each other by intervening words in the sentence, elicited P600 effects (Kaan, Harris, Gibson, & Holcomb, 2000). Based on these observations, the P600 has been proposed to reflect general syntactic integration difficulties associated with incorporating constituents into the emerging sentence structure (Kaan et al., 2000).

3.2.2 Neural oscillations

Already at the earliest observations of recordable electrical activity from the brain at the end of the 19th century, initially in animals, researchers noted the presence of rhythmic oscillations in the EEG (Bastiaansen, Mazaheri, & Jensen, 2012). For instance, high-amplitude 8-12 Hz oscillations, the alpha waves, are often easily detectable in the raw EEG, when subjects close their eyes (see Figure 2). Interest in neural oscillations, however, only started to grow in the 1980s and 1990s, when it became clear that ERPs constitute only one part of the event-related changes that take place in the EEG activity. Neural oscillations are ongoing phenomena, spontaneously present in the EEG even independently of a task; nevertheless, they might be modulated by experimental events. The related oscillatory changes are time-locked to the event of interest, but the phase of the oscillations (the position within the oscillatory cycle at a given moment) might vary from one presentation of the event to the next, due to the ongoing nature of rhythmic EEG activity. Consequently, non-phase-locked oscillatory event-related responses will be significantly reduced during the standard averaging procedure used to extract the ERP (Bastiaansen et al., 2012).
Figure 2
EEG represented in the time domain, between 629 and 634 seconds of a recording. Electrode names are shown along the vertical axis. Rhythmic alpha waves become especially prominent over parietal and occipital electrodes around 631 and 632 seconds.

The exact relationship between ERPs and neural oscillations is a matter of debate (e.g. Sauseng et al., 2007), and it is not clear whether these measures are the reflection of the same or independent phenomena. According to the additive model, ERPs are assumed to be produced by neural activity evoked by a stimulus that is additive to and independent of ongoing oscillations (e.g. Mazaheri & Jensen, 2006). The alternative phase-resetting view argues that ERPs are in fact the result of phase-related changes in ongoing oscillations: the onset of a stimulus in each trial causes a partial phase-resetting of the EEG, and the phase-locked rhythmic activity emerges as the ERP during averaging (e.g. Makeig et al., 2002). More recently it has also been suggested that both of these phenomena might contribute to ERP generation (Min et al., 2007).

Neural oscillations are often studied to gain insight into functional network formation in the brain (Bastiaansen et al., 2012). It has been argued that oscillatory synchronization among neuronal groups might constitute a mechanism for the dynamic coordination of distributed brain processes, underlying the emergence of unified concepts and experiences (e.g. Singer, 1999; Varela, Lachaux, Rodriguez, & Martinerie, 2001). For instance, visual objects are perceived as coherent entities, even though various attributes of an object are represented separately in different visual areas of the brain (Varela et al., 2001). Functional network formation can be assumed to rely on local neuronal synchrony within processing.
nodes of the network as well as on long-range synchronization between different nodes of the network (Bastiaansen et al., 2012). Elements constituting the same functional network will oscillate in synchrony at a specific frequency. Local synchrony will be reflected in power changes in a given frequency band: when a larger number of neurons within a population fire synchronously, the amplitude of field potentials recorded at a specific site increases (Bastiaansen et al., 2012). As for long-range synchronization, the phase relationship between oscillations picked up at different recording sites will be informative: with increased synchrony, the phase difference between the recorded rhythmic field potentials becomes more consistent (Bastiaansen et al., 2012). A number of methods exist for quantifying long-range neuronal synchrony (see e.g. Bastos & Schoffelen, 2016). In Study 1, Phase-Locking Value (PLV) analysis was conducted on the EEG data in order to gain insight into hypothesized modulation of the engagement of the language processing streams during the comprehension of different types of complex word forms. The PLV assesses the variance in the phase difference between two signals, e.g. oscillations recorded from two different electrodes, across trials of stimulus presentation. As input to the analysis, EEG data need to be represented in the time-frequency domain (Figure 2 shows EEG in the time domain), which involves estimating the magnitude and phase of oscillations at a given frequency for each time point in single-trial EEG epochs. Magnitude estimates are then unit-normalized (transformed to the same value) for the PLV analysis. Such magnitude normalized complex values are derived for signals obtained from both electrodes in question, and the two signals’ difference in phase is calculated for each trial. Quantifying the consistency of trial-to-trial phase differences will then yield the PLV index, ranging from 0 to 1, where 1 stands for completely stable phase differences (Roach & Mathalon, 2008). Software tools are available for semi-automatized processing of EEG data. For example, the open source Matlab toolbox, Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), contains predefined methods for calculating time-frequency representations and deriving PLV values.

3.3 MRI and cortical thickness measurements

As was mentioned in the previous section, EEG data provide limited information as to the brain regions that underlie language-related functions. Nevertheless, with the development of the magnetic resonance imaging (MRI) technique during the last few decades, it became a standard procedure to create images of different parts of the body in vivo, including the brain. Various methods have been introduced for mapping cognitive functions to neural substrates in humans in a non-invasive manner. For instance, functional MRI (fMRI) can be used to localize patterns of
brain activation, by measuring changes in blood oxygenation levels as blood flow to activated areas in the brain increases (Buxton, 2009). In addition, based on detailed anatomical images obtained with the MRI technique, one can examine the macrostructure of different brain areas and relate specific anatomical features to observed performance of cognitive functions. For instance, the thickness of the cerebral cortex has proved to be an especially informative measure in the study of various domains of human cognition. Pre-existing differences in the cortical thickness of higher-level association areas in the brain have been related to individual variation in the performance of associated cognitive abilities (Karama et al., 2009; Menary et al., 2013). Cortical thickness has been observed to change during an individual’s lifetime as a result of a number of factors, including experience – for example L2 acquisition (e.g. Mårtensson et al., 2012) – aging (e.g. Salat et al., 2004) and disease (e.g. Meyer et al., 2016). Study 2 relies on this measure of cortical structure to investigate the neural substrates of Swedish tone-suffix association processing. As a methodological background to this study, a brief description of the MRI technique is provided below, followed by an introduction to an automatized method for measuring cortical thickness based on MR images.

3.3.1 The MRI technique

MRI most commonly relies on the signal from the hydrogen proton, the nucleus of hydrogen-1 atoms, to create images. Such nuclei are abundant in the human body, which contains large amounts of water. Hydrogen protons spin on their axis, like the planet earth. Spin is an intrinsic property of the proton, which always has the same magnitude, and only the axis of spin can change. In addition, the proton has magnetic moment, being a rotating mass that has an electrical charge. The proton is therefore affected by external magnetic fields, similarly to a compass needle, and the motion of its magnetic axis generates a signal. The MRI scanner applies a strong magnetic field to the body placed into it, which causes the magnetic moments of the protons to line up with the direction of the field. The magnetic fields of the protons sum up to form a net magnetization. The magnetization of the protons precesses around the direction of the static magnetic field with a frequency that depends on the strength of the static magnetic field and the chemical environment of the nucleus. Precession refers to the wobbling motion that occurs when an external force acts on a spinning object, gradually changing the orientation of its rotational axis. Additional energy is then added to the magnetic field in the form of a radio frequency wave, referred to as an RF pulse, in resonance with the precession of the magnetization, which tips over the net magnetization, producing a measurable signal. To encode spatial information, two additional magnetic fields are applied in two directions that manipulate the
frequency of precession along one axis and the phase of the magnetizations along the other. The emitted signal is detected in the scanner using receiver coils, and its intensity is plotted on a grey scale to produce cross sectional images (Buxton, 2009; Weishaupt, Köchli, & Marineck, 2008). Importantly, protons in different tissues in the body relax back to their normal state at different rates, which creates the contrast in the resulting images. This relaxation time is measured in two ways: T1 refers to the recovery time of the net magnetization to reach equilibrium and T2 largely determines how much time it takes for the excited state to decay following excitation (Weishaupt et al., 2008). T1 and T2 times vary, for example, between white matter, grey matter and cerebrospinal fluid in the brain, and therefore such contrasts can be represented in the images (Buxton, 2009). For instance, as RF pulses are repeatedly applied to the imaged slice of the body with short intervals, the signal produced by a tissue with short T1 will be stronger, appearing bright in the images. Tissues with long T1 will produce relative weaker signals, appearing darker. The resulting digital MR image is a two-dimensional matrix of pixels, each standing for a value of signal intensity. Each pixel represents a corresponding three-dimensional tissue volume with a given slice thickness, referred to as a ‘voxel’ (Weishaupt et al., 2008).

The low energy of the radiation present in MRI experiments constitutes no biological hazards (Berger, 2002), but certain risks are involved largely related to the strong magnetic fields used, which can dislocate or heat up metal implants in the body. MR examinations are therefore preceded by rigorous screening and safety procedures. For clinical purposes, the magnetic field strength used is typically up to 3 Tesla, which is approximately 60 000 times the magnetic field of the Earth. For research applications, it is no longer uncommon to find 7 Tesla or even higher field systems (Duyn, 2012).

### 3.3.2 Measuring cortical thickness

The human cerebral cortex is a highly convoluted sheet of neurons, constituting two-thirds of the neuronal mass of the brain (Rakic, 1988). The thickness of the cortex varies between 1 and 4.5 mm, with an average of 2.5 mm over the whole brain (Fischl & Dale, 2000). Until the introduction of automatized methods around 2000, the measurement of cortical thickness from anatomical MR images was an extensively complex and labor-intensive task even for trained anatomists. One important difficulty is related to precisely identifying the three-dimensional folding of the cortex from a series of two-dimensional images. Therefore, studies on larger populations used to be rare, and cortical thickness was most commonly examined in post-mortem investigations (Fischl & Dale, 2000). Nowadays, however, open source software tools are freely available for the accurate and automated generation of cortical thickness measurements over the whole brain. In
Study 2, cortical thickness was measured using a surface-based automated method offered as part of the Freesurfer software package (Dale, Fischl, & Sereno, 1999). The input data constitutes MR images with sufficiently high spatial resolution and T1 contrast. The image processing method then involves the generation of accurate models of the grey and white matter surfaces as well as the pial surface, which is the boundary between the grey matter and the cerebrospinal fluid (see Figure 3). Cortical thickness is subsequently calculated as the shortest distance between the white matter and the pial surface (Fischl & Dale, 2000). To support identification of specific regions of interest in the brain, e.g. the pars opercularis of the inferior frontal gyrus, automated systems have been developed for labeling cortical (and subcortical) structures (e.g. Desikan et al., 2006). Mean cortical thickness values can be extracted for delimited brain regions and submitted to statistical analysis.

Figure 3
Illustration of the pial surface (red) and the white/grey matter boundary (blue). Cortical thickness can be measured as the distance between the white matter and the pial surface.
4. The investigations

4.1 Study 1 – Brain responses to morphologically complex verbs

Study 1 (Schremm, Novén, Horne, & Roll, submitted) focused on the question as to whether regularly inflected versus irregular verbs are processed by different neural mechanisms in Swedish during reading, as reflected in electrophysiological brain responses (ERPs). Previous studies have reported different ERP effects for regularly inflected words versus irregular forms in a range of other languages (Gross et al., 1998; Morris & Holcomb, 2005; Münte et al., 1999; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Weyerts et al., 1997), which have typically been interpreted to indicate the involvement of separate processing mechanisms, largely in line with dual route models (Clahsen, 1999; Pinker, 1999). Specifically, misapplication of the regular inflection to an irregular stem has been observed to elicit a left anterior negativity (LAN), which is commonly interpreted as an index of morphosyntactic anomaly detection. Therefore, it has been suggested that the regular inflection is apparently associated with a decompositional rule-based operation, underlying its segmentation from the stem, explaining the brain response associated with violations of morphosyntactic structure (e.g. Penke et al., 1997; Rodriguez-Fornells et al., 2001). Furthermore, the operation of this decompositional mechanism is apparently tied to the presence of the regular inflection, as irregularized regular words have yielded no or different effects relative to the correct variant, implicating full-form access for correct irregular words in the mental lexicon.

Due to the absence of any relevant electrophysiological results on regular/irregular verb processing in Swedish, it was unclear whether the incorrect application of the regular inflection, such as the past tense suffix -er/-te, would elicit a LAN in Swedish, similarly to findings in other languages. A further question was whether the observed response to the regular inflection would be any different from the effect for the misapplication of the irregular pattern. In the regular/irregular word processing research paradigm, the default, i.e. regular, word class is typically considered to be the one where a regular and productive inflection is applied, inducing no changes to the stem (e.g. Sonnenstuhl et al., 1999; Weyerts et al., 1997). From this perspective, the 4\textsuperscript{th} conjugation class of
strong verbs in Swedish can be considered as the non-default, irregular class. Nevertheless, these types of verbs are still characterized by largely predictable vowel changes in the verb stem, which raises the possibility that such word forms might in fact be processed based on morphological regularities, similar to regular verbs, instead of retrieving a stored whole-word representation from memory.

EEG was recorded from Swedish native speakers while they read sentences with overregularized irregular (e.g. *bär+de ‘bear + past tense’) and irregularized regular (e.g. *löft ‘lifted’) verbs, as well as corresponding correct irregular (e.g. bar ‘bore’) and regular (e.g. lyft+(t)e ‘lift+ed’) forms. Overregularized items were constructed by attaching the regular past tense (preterite) suffix to an irregular (4th conjugation class) verb, and irregularized words were formed by applying the predictable vowel change patterns of irregular past tense forms to regular (2nd conjugation class) verbs. The task was to judge the correctness of sentence form after each trial. In line with previous findings, overregularized verbs elicited a left-lateralized negativity between 350 and 500 ms following verb onset, interpreted as a LAN. Assuming that the source of this effect was the incorrect application of the default past tense rule to the stem, these results implicated rule-based decomposition for correct regular verbs in Swedish. Also, no difference was found between irregularized verbs and correct regular forms in the LAN time window, suggesting that the stem vowel alternations of the irregular class might not be perceived as productive morphological regularities with a comparable status to the regular past tense suffix, and correct irregular verbs might be retrieved in full form from the mental lexicon. Both incorrect conditions elicited a P600 as expected, potentially due to difficulties related to sentence structure processing with an incorrect verb form.

In addition, inspection of the ERP waveforms superimposed for all four conditions, instead of considering regular and irregular conditions separately, suggested the presence of a negativity in the LAN time window for not only overregularized verbs, but also for both regular conditions (regular correct and irregularized verbs) relative to correct irregular verbs. Indeed, increased morphological processing could have conceivably taken place in the conditions showing a negativity, which might be reflected in LANs, under a more general view of this ERP effect as an index of morphological analysis (Krott & Lebib, 2013). Thus, even irregularized verbs could have been processed based on morphological regularities of the vowel alternation pattern, since participants might have reanalyzed the stem during reading the sentences in order to recover the original correct regular form. This interpretation would implicate relatively more reliance on semantic information in processing the familiar irregular correct verbs, due to whole-word access, thus suggesting involvement of the semantic ventral stream (e.g. Hagoort, 2013). At the same time, relatively greater reliance on the dorsal stream might be assumed for the other conditions, where morphological processing for the verb is engaged as well (Bozic et al., 2015;
Marslen-Wilson & Tyler, 2007; Rolheiser, Stamatakis, & Tyler, 2011). In order to test these hypotheses, we conducted a follow-up analysis on the EEG data, comparing oscillatory phase synchrony between frontal and posterior electrodes across the experimental conditions. The rationale behind the analysis was that modulation of oscillatory phase synchrony between distant sites could indicate transient long-range functional network formation in the brain (von Stein & Sarnthein, 2000), and therefore might tap into patterns of engagement of the dorsal versus ventral processing streams for comprehension of the different verb forms. Results indicated increased synchrony for the irregular correct verb condition relative to overregularized verbs in the theta frequency range (4-7 Hz), which was confined to the left side of the brain during the time period of the LAN effect but later spread to both hemispheres in the P600 time window. This observation was tentatively interpreted to indicate a relative difference in the processing streams engaged, with greater reliance on the ventral, more semantic, route for irregular correct items. As expected, no difference was found between regular correct versus irregularized verb conditions, which might suggest similar involvement of the dorsal stream, assumed to be the decompositional route, for processing both of these word form types.

In sum, the proposed interpretations for both the ERP and the oscillatory phase synchrony results were consistent with the assumption that two separate processing routes might be engaged during the comprehension of visually presented complex word forms in Swedish. Processes associated with decompositional analysis versus whole-word access of inflected verb forms might potentially differ in their reliance on the ventral versus dorsal language processing streams. Furthermore, it seems that even regularities applying in the stem of the non-default class might be associated with morphological processing under specific conditions; nevertheless, the correct Swedish irregular verbs tested were apparently retrieved as whole words during reading.

4.2 Study 2 – Cortical thickness in native language tone processing

The results of Study 1 indicated two separate processing routes underlying the comprehension of different types of inflected word forms in Swedish. Nevertheless, as these findings were obtained with visually presented sentences, it was still unclear how Swedish native speakers process complex word forms when word accents on the stem cue upcoming inflections in speech stimuli. This question was investigated in Study 2 (Schremm et al., 2018), where the focus was on the way the cortical structure of brain regions implicated in Swedish tone
processing could support integration of the word accent cue into the morphological processing system.

Tones on word stems in Swedish elicit a pre-activation negativity (PrAN) in the brain response, suggesting anticipatory activation of memory representations of likely upcoming word endings, modulated by the certainty and frequency of a given continuation (Roll et al., 2017; Söderström et al., 2016a). Considering the gradual nature of this process, individual differences among Swedish native speakers might be observed as regards the degree to which they rely on word accents during suffix processing. Such differences could be reflected in response-time (RT) patterns to cued inflections: the more listeners rely on the word accent, the more their suffix processing is disrupted by word forms with invalid tone-suffix associations, resulting in longer decision latency as regards the meaning of the inflection, relative to the case when a valid tonal cue facilitates processing of upcoming word structure. Furthermore, reliance on word accents might be associated with macrostructural variation in core brain areas implicated in Swedish tone processing. Specifically, cortical thickness has been associated with individual differences in cognitive abilities, implemented in those brain regions that displayed variation in the measure (Karama et al., 2009).

Based on previous neuroimaging studies on functional brain activation during speech comprehension, the planum temporale (PT) and the pars opercularis of the inferior frontal gyrus (IFGpo) in the left hemisphere were identified as crucial areas for word accent processing in inflected Swedish words (Roll et al., 2015) and in pseudowords carrying regular Swedish suffixes (Söderström et al., 2017), respectively. Pseudowords clearly do not have any memory representations in the mental lexicon that would enable processing through stored full forms, and comprehension would therefore presumably involve segmenting the only familiar and grammatically meaningful element, the suffix, from the stem. Previous studies indeed reported left IFGpo involvement in decompositional processing of regularly inflected words (Bozic et al., 2015; Tyler et al., 2005). At the same time, speakers might potentially store more frequent (real) Swedish words in lexical memory (Lehtonen, Niska, Wande, Niemi, & Laine, 2006), in which case the stem tone could be thought to be incorporated into the stored whole-word representation. Study 2 investigated whether cortical thickness could support these two hypothesized tone-suffix processing mechanisms, by examining the relationship between listeners’ reliance on word accent cues, quantified by their RTs, and the cortical thickness of the PT as well as the IFGpo. Magnetic resonance imaging and RT data were analyzed from two previous experiments, testing real words (Roll et al., 2015) and pseudowords (Söderström et al., 2017). Participants listened to sentences with words (real or pseudoword stems) inflected with the singular or plural suffix, and made a decision to the question as to whether the target words referred to “one” or “many” things. RTs were measured from suffix onset. In the Valid condition, the tone on the word stem (accent 1 or
accent 2) was followed by its associated singular or plural suffix (e.g. $hattr_{\text{accent1}}+en$ ‘hat$\text{sg}’$, $kvut_{\text{accent1}}+en$ ‘kvut$\text{sg}’$), and in the Invalid condition an incorrect stem tone-suffix sequence was presented (e.g. *$hattr_{\text{accent1}}+ar$, *$kvut_{\text{accent1}}+ar$). The analysis in Study 2 involved extracting mean cortical thickness of bilateral PT and IFGpo using Freesurfer. Subsequently, Invalid minus Valid RT differences, quantifying the relative processing advantage for correctly cued suffixes, were correlated with the obtained cortical thickness measurements for each participant.

For real word processing, results showed that reliance on the tonal cue correlated positively with cortical thickness in left PT, but not in right PT or bilateral IFGpo. For pseudowords, RT advantage for validly cued suffixes correlated positively with cortical thickness in left IFGpo instead. These results suggest that the way cortical thickness of left PT supports tone processing seems to be specific to real words. Importantly, the tested pseudowords consisted of Swedish phonemes, but larger chunks such as syllables of the pseudowords did not occur in actual words. Based on these considerations, the following possible mechanism was proposed underlying tone pre-activation, when inflected words are accessed in full form in the mental lexicon: native speakers could be assumed to process the speech input in terms of stored memory representations for frequent speech sound patterns, e.g. syllables, which could even incorporate tone patterns in Swedish. Given the previously observed role of left PT in linguistic tone processing (Xu et al., 2006), greater cortical thickness of PT might enable more efficient analysis of tone information and in turn rapid activation of the stored sound patterns incorporating tone, which would facilitate pre-activation of the whole word form with the tone-associated ending. For pseudowords, however, the pre-activation process cannot rely on stored syllable chunks incorporating tones and might take place via a morphological rule instead, as indicated by the correlation found with left IFGpo. Thus, thicker cortex in left IFGpo might enable more efficient use of an abstracted association between the tone itself and the inflectional suffix, and could potentially support the segmentation process as well, separating the meaningful suffix from the stem.

In general, it seems that two mechanisms are available for processing inflected Swedish word forms in speech, and cortical structure of specific areas in the left hemisphere might modulate the efficiency of these processes. Cortical thickness in the left PT might play an important role in the processing of tonal cues when stored whole-word representations with the associated suffix are retrieved, and cortical thickness in left IFGpo could facilitate rule application specifying the relevant tone-suffix association during decompositional analysis.
4.3 Study 3 – Implicit acquisition of tone-suffix connections in L2 learners of Swedish

Study 3 (Schremm, Söderström, Horne, & Roll, 2016) examined L2 acquisition of the morphological function of word accents by adult learners of Swedish. The main question was whether learners with a non-tone native language background would implicitly acquire the predictive use of tone-suffix associations during L2 speech processing after relatively long exposure to Swedish and when they have reached higher levels of proficiency. It was assumed that any learning effects found would be the outcome of implicit acquisition, since Swedish tone-suffix associations are not normally taught at language courses and they are unlikely to be pointed out by native speakers who are typically unaware of the regularity.

In order to test whether relatively advanced L2 learners use word accents predictively, we analyzed response-time patterns to regular verbal inflections, validly or invalidly cued by the preceding tone on the word stem. The experimental conditions were similar to the ones used in Study 2, but the target words presented in sentences were inflected verbs instead of nouns, and exclusively real Swedish words. In the Valid condition, accent 1 or accent 2 was followed by its related present tense -er or past tense (preterite) -de/-te suffix in regular 2nd conjugation class verbs, e.g. lek_{accent1}+er ‘play+s’, lek_{accent2}+te ‘play+ed’. In the invalid condition, the suffix was preceded by an incorrect stem tone, e.g. *lek_{accent1}+te, *lek_{accent2}+er. The task was to decide on the present tense versus past tense reference of the sentence after each verb presentation, as quickly as possible. The participants were L2 learners of Swedish with a non-tone, non-Scandinavian native language background, who were all at an intermediate to upper intermediate proficiency level in the L2 and had spent on average 1.9 years in Sweden. Their response-time patterns were compared to data collected with native speakers in Swedish in a previous study (Söderström et al., 2012).

Results showed that, similarly to native speakers, L2 learners responded significantly faster to suffixes validly cued by the stem tone, relative to invalidly cued suffixes. This finding suggests that the learners had acquired the tested tone-suffix associations and they relied on tones to anticipate likely upcoming inflections. Encountering a suffix that mismatched the tone-based expectation in the Invalid condition generated greater processing difficulty, reflected in increased response times. It seems therefore that it is possible to acquire the predictive use of tones as cues to morphological structure without instruction. Presumably, L2 learners acquired this feature via implicit learning mechanisms that operate on statistical regularities, such as the transitional probability of the suffix to follow a specific word accent on the stem. Such regularities are abundant in the input given the pervasiveness of Swedish word accents, and the co-varying features (tone-suffix) are often realized locally, e.g. on adjacent syllables in disyllabic words.
Nevertheless, L2 learners displayed a smaller processing advantage for validly cued suffixes than the control group, suggesting that they still did not rely on the predictive significance of word accents to the same extent as native speakers do, possibly due to weaker tone-suffix associations in their mental representations. However, a marginal positive correlation obtained between time spent in Sweden and the response-time advantage for validly cued suffixes suggested that processing might become more native-like with increased exposure to the L2.

Interestingly, L2 learners responded faster and more accurately to the suffixes than native speakers did, which was tentatively attributed to the fact the learners were on average younger, which might have constituted a slight general cognitive processing advantage. Furthermore, suffix-processing patterns were also modulated by group: native speakers responded faster to present tense than to past tense suffixes, while no such difference became significant for L2 learners. As a possible explanation, it was suggested that learners’ L2 tense system might be still somewhat less complex than that of native speakers, perceiving both the tested present tense and preterite suffixes as default ways of expressing the given tense reference, whereas native speakers might have been more influenced by the fact that a larger number of complex tense constructions can be used to grammatically refer to the past than to the present in Swedish.

4.4 Study 4 – Training predictive L2 processing with a digital game

The results of Study 3 indicated that it is possible for adult L2 learners to implicitly acquire the Swedish tone-suffix associations, through exposure to the language. Nevertheless, even relatively advanced learners displayed more limited predictive use of word accents as compared to native speakers, suggesting that implicit acquisition of this L2 feature is a gradual and slow process. Therefore, in Study 4 (Schremm, Hed, Horne, & Roll, 2017) we explored the possibility of training low proficient L2 learners of Swedish in the anticipatory use of tones as cues to upcoming suffixes. Based on previous findings of L2 acquisition studies, we formulated a number of hypotheses as regards the learning conditions that could facilitate acquisition of the targeted L2 skill. For instance, extensive exposure to the tone-suffix associations in naturalistic native speech input was assumed to be beneficial for promoting the development of predictive associations between tones and suffixes in the learners’ L2 system. Also, introduction of a more explicit task requiring the actual performance of a tone-based prediction was considered necessary, in order to direct attention to the stem tone as an anticipatory cue. Finally, a task encouraging fast performance of the tone-based suffix prediction was assumed to facilitate the emergence of an L2 predictive
processing skill that can underlie the rapid generation of suffix predictions under the time-constraints of online speech processing.

Motivated by the above considerations, a prototype of a digital L2 learning game was developed and tested. Game play consisted of presentations of sentence contexts (both auditory and visual), up until a predictive tonal cue on the target word stem is heard. For each sentence item, the player was required to decide how the target word stem would continue by selecting one of two alternative inflectional suffixes, as quickly as possible. Both suffixes constituted grammatically correct continuations, but only one of them was cued by the preceding word accent. The player then received immediate feedback on the accuracy and speed of the choice, and the correct continuation was always pronounced and displayed, in order to foster the development of correct associations between word accents and suffixes. Target words included nouns with singular versus plural inflections (e.g. *bil-en* ‘car-the’ versus *bil-ar* ‘car-s’) as well as regular verbs with present versus past tense (preterite) inflections (e.g. *lek-er* ‘play-s’ versus *lek-te* ‘play-ed’), similar to the word forms tested in Study 2 and 3. Only valid tone suffix associations were presented, in the context of real and relatively frequent Swedish words. The language material was organized into levels of increasing difficulty, and an accuracy of 80% for a whole round (18-20 sentence items) was required for progression to the next level. In order to test the proposed L2 training method, L2 learners of Swedish with non-tone native language background played the prototype of the game for 10 days, 15 to 60 minutes each day, while response-time and accuracy data were continuously logged. Learners were at low to intermediate proficiency levels and had lived in Sweden for an average of 5.8 months. A sentence production test was conducted before and after training, to see if there would be any improvements in learners’ pronunciation of tone patterns, correctly preceding the trained noun and verb inflections.

Analysis of game performance data indicated clear learning effects by the end of the training period: L2 learners’ accuracy in selecting the correct suffix continuation cued by the tone showed a general increase throughout the levels of the game, at the same time as they became gradually faster at making the correct choice. Thus, the game seems to have promoted the acquisition of the tone-suffix associations as well as the development of the anticipatory use of tones. More time spent on the final level of the game, practicing the complete language material, was associated with greater accuracy gains, further indicating that performance improvement was actually due to training. Interestingly, response-time reduction was not straightforwardly related to practice time, as suggested by an absence of correlation between the measures. This observation might be indicative of the role of qualitative changes, such as automatization, in significantly speeding up the performance of the predictive L2 skill, possibly explaining the lack of a simple gradual learning curve. In addition, a transient increase in response times was
associated with the introduction of mixed target word types in the game, i.e. when sentences with either inflected noun or verb targets were first presented within the same sequence. This apparent increase in processing difficulty was possibly related to the fact that acquisition of the predictive use of word accents is not limited to the development of suffix pre-activations, but one should also learn to generate predictions that are contextually relevant. For instance, a word accent on a verb stem preferentially pre-activates only a subset of its associated suffixes. These would include word endings that constitute grammatical continuations, excluding therefore noun inflections. Also, pre-activation of associated suffixes might be modulated by further contextual (such as semantic or pragmatic) constraints. The L2 acquisition of this mechanism, i.e. inhibition of possible continuations according to dynamically changing contextual constraints, might temporarily increase processing difficulty. Finally, accuracy in pronouncing tone patterns improved significantly from pre- to post-training, indicating that L2 perceptual training focusing on tonal cues might beneficially affect production. Lack of any significant correlations with game performance gains pointed to a complex relationship between perception and production improvements.

In conclusion, only after ten days of playing the game prototype, L2 learners’ performance indicated significantly improved anticipatory suffix processing based on word accent cues. These results indicate that even low proficient L2 learners might begin to integrate Swedish tones as predictive cues into their L2 inflectional morphology processing system, if they receive focused training.
5. Conclusions

Inflectional morphology processing in Swedish was investigated in four studies. Studies 1 and 2 focused on native speakers and examined core processing mechanisms involved in the comprehension of complex word forms, in reading as well as in speech, when tones on word stems function as predictive cues to upcoming suffixes. Studies 3 and 4 addressed issues related to the L2 acquisition of a unique aspect of Swedish inflectional morphology processing: the word accent-suffix associations and reliance on this feature to anticipate word structure in online speech comprehension, at higher as well as lower L2 proficiency levels.

The results of Study 1 suggest that two separate processing routes are available for the comprehension of complex Swedish verb forms: rule-based decompositional analysis and whole-word access in lexical memory, in line with dual route models (Clahsen, 1999; Pinker, 1999). Generally, decomposition was associated with the presence of the regular past tense inflection, indicating that verbs such as *lekte ‘played’ might be processed by analyzing the verb form into stem and suffix parts (i.e. lek+te). Thus, similar to results in a number of previously investigated languages, incorrect application of the regular past tense inflection in Swedish elicited a LAN in the electrophysiological brain response, commonly interpreted as an index of morphological anomaly detection. Interestingly, LAN-like negativities were apparently also present for the incorrect use of the irregular verb formation pattern and for correct regularly inflected verbs in Study 1. This observation would be consistent with a more general interpretation of the LAN effect, signaling morphological processing as such (Krott & Lebib, 2013). Therefore, when no corresponding stored whole-word representations exist, it seems that decompositional mechanisms might be engaged even for processing regular patterns associated with non-default word classes, i.e. the predictable vowel alternations in Swedish irregular verb stems. Participants reading irregularized verbs such as *löft might have made an attempt to recover the correct word form, potentially based on changing the stem vowel (*löft > lyft- > lyfte ‘lifted’). The only experimental condition that showed no signs of morphological processing of target words in Study 1 constituted of correct irregular verbs, suggesting that native speakers reading an irregular verb such as *bar ‘bore’ might directly access a corresponding stored inflected word form in lexical memory. Facilitated semantic processing associated with whole-word access to irregular word forms was proposed to be reflected in the increased
posterior-frontal oscillatory synchrony obtained for that experimental condition. Increase in synchronization between these distant regions of the brain might indicate engagement of a functional network along the ventral language processing stream, potentially in relation to direct lexical access to complex word forms.

Based on the findings of Study 2, Swedish stem tones can be assumed to be used predictively to anticipate upcoming inflectional suffixes during both whole-word access and decomposition of complex word forms. Predictive reliance on tones might be implemented via incorporation of the tone pattern into stored representations of whole inflected word forms. When such stored items are unavailable, e.g. for pseudowords, the application of a regularity specifying the relevant abstracted tone-suffix association could facilitate decompositional processing of the cued word ending. These proposals were formulated considering the functional significance of the investigated brain regions where cortical thickness showed an association with anticipatory tone processing performance in Study 2. Thus, left planum temporale (PT) seems to play an important role when stored full form representations are accessed, as reflected in a positive correlation between response-time advantage for cued suffixes in real words and cortical thickness in this region. Thicker left PT cortex might in general enable more efficient left-hemispheric tone processing due to greater prevalence of cortical organization tuned for the accurate representation of slower changing acoustic cues (e.g. Harasty, Seldon, Chan, Halliday, & Harding, 2003; Poeppel, 2003).

Assuming that Swedish native speakers have stored phonological representations for certain inflected word forms specifying stem tones as well, processing a stem tone would automatically generate greater pre-activation of those complex word forms that are inflected with the suffix the specific tone pattern is associated with. A significant positive correlation obtained between cortical thickness and response times to inflected pseudowords suggests that the pars opercularis of the left inferior frontal gyrus (IFGpo) might be involved when predictive suffix processing relies on abstract stem tone-suffix associations. Thicker cortex in this area could conceivably support effective morphological rule application, given that left IFGpo was previously seen to be implicated in decompositional processing of inflected word forms (Bozic et al., 2015; Tyler et al., 2005). It seems therefore that the abstracted association between Swedish tones and suffixes might be captured in terms of a morphological regularity, operating independently of specific lexical items, as has also been proposed by Söderström et al. (2016b). Even though these results were obtained for pseudowords, similar tone-suffix processing mechanisms and neural substrates might be assumed for the comprehension of less frequent regularly inflected real words that are not stored in lexical memory in their various inflected forms.

The findings of Study 2 that the presence or absence of stored whole-word representations might modulate the role of specific brain regions in inflectional morphology processing are in line with the apparent differential engagement of
language processing streams observed in Study 1, for full-form access versus morphological analysis of inflected verbs. Moreover, the involvement of frontal (IFGpo) versus temporal (PT) areas is consistent with proposed dual-route brain systems, associating decomposition with a fronto-striatal network and full-form access largely with temporal regions (Pinker & Ullman, 2002; Ullman, 2004).

Turning to Swedish inflectional morphology processing in adult L2 acquisition, Study 3 showed that tone-suffix associations are acquired implicitly by more proficient learners as a result of extensive exposure to the L2. In a manner similar to that of native speakers, the learners tested in this study responded relatively faster to those suffixes that were cued by the preceding word accent. At the same time, incorrect tone-suffix associations increased processing time, potentially due to the need to re-evaluate a disconfirmed suffix prediction. Nevertheless, the predictive use of tones appeared to be less extensive in learners as compared to native speakers, even at relatively advanced L2 proficiency level: it can be assumed that the emergence of strong tone-suffix associations in mental representations and/or the acquisition of relevant predictive processing skills in adult L2 learners take a long time to develop under implicit learning conditions.

However, L2 training combining features of implicit and explicit learning conditions can effectively promote the acquisition of the anticipatory use of tone-suffix associations even at lower proficiency levels in Swedish, as shown by Study 4. During a two-week-period of training using a game prototype, learners displayed a gradual reduction in their response times to suffixes cued by preceding tones. Together with the observed increase in accuracy in selecting the cued endings, these findings suggest that learners acquired the relevant tone-suffix associations and started to use these to predict upcoming suffixes. Moreover, speakers of non-tone native languages could apparently learn to discriminate Swedish tone patterns well enough to begin to anticipate suffixes accurately, and the provided perceptual training even led to improved production of word accents.

A number of features of the tested game mechanics could be assumed to have contributed to the observed learning effects in Study 4. First, ample exposure to the tone-suffix associations in native speech input was assumed to engage learning mechanisms operating on the transitional probabilities between the predictive tonal cue and the target, the suffix. Such learning mechanisms could be thought to have promoted the development of strong mental associations between tones and suffixes in learners, supporting predictive processing. Second, it was considered important to direct attention to the stem tone as an anticipatory cue, in the form of a task that requires the performance of an actual prediction as regards the upcoming inflection. Finally, repeated and extensive performance of the suffix-prediction, under circumstances that motivated rapid responses, could have contributed to speeding up, and potentially automatizing, learners’ emerging predictive L2 processing skill.
To sum up, Swedish native speakers seem to process inflected word forms either by morphological processes of decomposition or by direct access to stored full form representations in the mental lexicon. Swedish stem tones appear to be integrated into this system so that they can effectively cue upcoming suffixes via both processing routes. Cortical thickness of specific brain regions might facilitate this process, in the left PT during whole-word access and in the left IFGpo during decompositional analysis, respectively. Furthermore, results of the L2 studies generally indicate that adult learners of Swedish can learn to rely on tonal cues to upcoming suffixes during L2 morphological processing, even in the absence of comparable tonal features in their native language. The use of tones to facilitate suffix processing seems to emerge implicitly after a longer exposure to the L2 and when higher levels of proficiency have been reached, whereas learners at lower proficiency levels might need focused training in the predictive L2 skill.
6. Outstanding issues and future directions

6.1 Factors modulating inflectional morphology processing mechanisms in native speakers

The results of Studies 1 and 2 raise questions as to the factors that might modulate the morphological processing route in native speakers for regularly inflected items in different contexts. In Study 1, correct regular verb forms were apparently processed through decomposition (yielding a LAN-like negativity in the brain response). In Study 2, inflected real nouns (items in the real word experiment) were associated with full-form access. As Pinker and Ullman (2002) argued, even certain regular word forms might be stored in declarative memory, and the processing route chosen in a given context might then depend on item-, task- or speaker-specific factors. In the present case, the most relevant differences between the real-word stimuli of Study 1 and Study 2 seem to concern the task (grammaticality judgement versus identifying singular/plural meaning), the presentation modality (visual versus auditory) and the word class of the target items (verbs versus nouns). First, in Study 1, the task (“Correct or Incorrect”), as well as the presentation of incorrectly inflected items could have motivated a general focus on word form and in turn decompositional processing. In addition, word-by-word visual presentation, which makes it possible to start processing the word stem and the inflection almost simultaneously, could have also favoured decompositional analysis. In Study 2, the task was more semantic in nature (“One or Many”) and the target words were presented as spoken stimuli, which further raises the question whether the presence of tonal cues to suffixes might have specifically promoted whole-word access, if such tonal patterns could be assumed to be incorporated into stored full-form representations. Finally, it is possible that the nominal suffixes involved in Study 2 did not have the same status in participants’ grammatical system as the regular past tense inflection tested in Study 1. In fact, Roll et al. (2010) noted that Swedish has no default plural formation rule, comparable to the -s plural suffix in English or German, and nouns can be classified into several declension classes. Nevertheless, a specific plural suffix is associated with each of the seven declension classes (Teleman et al.,
(1999), and it could be assumed that the inflected form might be processed decompositionally following identification of the declensional class of a given noun stem. Still, Roll et al. (2010) observed an increased N400, as well as a P600, for plural suffixes that were incorrectly applied to noun stems of a different declension class (e.g. *mink+or instead of correct mink+ar ‘mink+s’). The obtained N400 effect indicates that the incorrectly inflected nouns were treated as unfamiliar lexical items, suggesting the absence of morphological decomposition related to the plural suffix. This is in contrast to the LAN effect obtained for the incorrect application of the regular past tense inflection in Study 1, indicating morphological analysis. Roll et al. (2010) employed auditory presentation, just like Study 2, whereas the task involved relatively greater focus on form (acceptability judgements: “OK or Wrong”), similarly to Study 1. In general, it seems that all the discussed factors could have potentially modulated the processing route for the inflected word forms in Studies 1 and 2, possibly in a complex interaction with each other. An EEG experiment, similar to that in Study 1, systematically manipulating factors such as presentation modality, task requirement, target word class and the presence versus absence of tones as task-relevant cues to suffixes could further clarify the nature of neural mechanisms involved in native inflectional morphology processing.

6.2 Issues related to L2 processing and acquisition of tone-suffix associations

More research is necessary to identify the neural mechanisms and brain structures that subserve L2 processing of complex Swedish word forms and the predictive use of word accents. Advanced L2 learners display native-like tonal cue processing tendencies in their behavioural responses as shown in Study 3, but it is not clear whether the underlying representations and mechanisms are native-like as well and how these might develop with increasing L2 proficiency. First of all, it is still unknown whether two separate processing routes are available to adult L2 learners for the comprehension of complex word forms in Swedish and whether they function like they do in native speaker processing. In fact, late L2 learners have been argued to tend to rely on lexically stored whole-word representations, even for regularly inflected words, until high levels of proficiency have been reached (Clahsen, Felser, Neubauer, Sato, & Silva, 2010; Ullman, 2001b). From this perspective, the procedural memory system is characterized by reduced availability during adult L2 acquisition due to, for instance, hormonal changes in adolescence and the development of the declarative system (Ullman, 2006b). One might speculate whether the fact that different response-time patterns were observed for present tense versus past tense suffixes in native speakers as
compared to L2 learners in Study 3 could possibly reflect an associated difference in the morphological processing route between these participant groups. Native speakers processed present tense suffixes faster than past tense ones, while there was no significant difference in L2 learners. One highly tentative explanation could be that learners used whole-word access for the comprehension of the inflected verbs, whereas native speakers relied on decompositional processing of the regular present and past tense suffixes, in a manner similar to the processing strategies observed in Study 1, potentially motivated by the task requirements (decision on grammatical tense). Direct retrieval of stored full forms could partially explain the generally faster response times and the limited influence of suffix type in L2 learners. Certainly, native speakers could have showed a relative processing advantage for the present tense inflection solely due to its association with accent 1, which is the more predictive word accent (Roll et al., 2015). In that case, however, it is not clear why the learners tested in Study 3, who implicitly acquired tone-suffix associations through exposure to the L2, did not show a similarly facilitated processing of the accent 1-associated suffix. The available data do not enable us to draw any conclusions as to the questions whether results obtained in Study 3 could actually reflect whole-word access in L2 learners and whether this strategy would be due to the limited availability of the procedural memory system for L2 processing.

Nevertheless, previous neuroimaging studies have reported activations in brain areas related to native grammatical computations in adult L2 learners, even at low proficiency levels, a finding which challenges the assumption that the procedural system might generally be unavailable at earlier stages of L2 acquisition (Abutalebi, 2008). Also, ERP responses have been obtained in low proficient L2 learners of Swedish that were indicative of decompositional analysis of inflected word forms: following training with the game prototype presented in Study 4, L2 learners showed a left anterior negativity (LAN) for singular/plural suffixes invalidly cued by the preceding word accent (e.g. *hatt\textsubscript{accent2}+en, ‘hat+sg’) (Hed, Schremm, Horne, & Roll, submitted). A similar ERP effect has been observed in Swedish native speakers for invalidly cued suffixes in pseudowords (Söderström et al., 2016b). It seems that L2 learners might not have been familiar with the tested lexical items, and therefore processed these in a manner similar to inflected pseudowords, but they detected the incorrect tone-suffix associations as they decomposed the word forms (Hed et al., submitted). The earliness of the LAN effect, 225-300 ms following suffix onset, might indicate the involvement of automatic processes, which would be consistent with the assumption that decomposition observable after L2 training could actually be implemented in the procedural brain system, instead of reflecting conscious rule-application. Nevertheless, further neuroimaging studies are necessary to identify the exact brain regions underlying L2 processing of complex Swedish word forms.
A left-lateralized brain network has been identified in Swedish native speakers for the predictive use of word accents (Roll et al., 2015) and the brain regions where cortical structure was associated with tonal cue processing were also localized to the left hemisphere in Study 2. However, non-lexical tonal information in general is preferentially processed in the right hemisphere (Zatorre & Gandour, 2008), and native speakers of English with no experience in tonal languages have been found to differ from Chinese listeners in that they do not show left dominant processing of Chinese lexical tones (Wang, Jongman, & Sereno, 2001). It could therefore be assumed that at the earliest stages of L2 acquisition of Swedish, word accents are processed largely in the right hemisphere in learners with non-tone native language background. Indeed, low proficient L2 learners of Swedish displayed a centrally distributed late negativity in their brain responses to accent 1, which became more prominent at right-hemispheric electrode sites with increasing L2 proficiency, as learners presumably started to develop greater sensitivity to pitch differences (Gosselke Berthelsen et al., 2018). The development pattern of functional brain activation as a result of L2 training in Swedish tone-suffix associations could be tracked using fMRI in order to establish if and when a shift to left-hemispheric tone processing might take place in learners. Examining the timeline of this change relative to the emergence of signs of predictive use of word accents in response-time patterns as well as the ERP (PrAN) could also help answer the question whether left-lateralization of tone processing is a prerequisite or a consequence of integrating word accent cues into the L2 morphological system. In addition, it could be examined whether L2 training of the tone-suffix associations would lead to increased cortical thickness in brain regions such as the left PT and IFGpo or, alternatively, if pre-existing differences in the macrostructure of these areas would predict learning success. The results would contribute to our understanding of the plasticity of the adult brain and potentially shed light on the question whether experience with a tonal language could result in an increase in the cortical thickness of specific brain areas. Also, observing training-related changes in the left IFGpo would be suggestive of the involvement of the procedural memory system in adult L2 acquisition.

6.3 Neural oscillatory synchrony and the language processing streams

Investigating neural oscillatory synchrony in order to gain insight into dynamic functional language network formation constitutes a promising research direction. In Study 1, increased theta (4-7 Hz) synchrony between posterior and frontal sites was tentatively suggested to reflect engagement of the ventral language processing
stream for whole-word access. Recording magnetoencephalographic (MEG) activity during an experimental task similar to that in Study 1 would enable localization of the brain areas that participate in the oscillatory synchrony in order to ascertain increased engagement of the ventral stream during the processing of familiar irregular verbs. Furthermore, it would be possible to examine whether the comprehension of regularly inflected items is associated with increased synchrony among nodes of the dorsal route. Interestingly, long-range synchrony in the gamma (and higher beta) frequency band (20-60 Hz) along the dorsal stream – between left posterior superior temporal gyrus and left IFGpo (BA 44) – has been reported to specifically underlie the processing of regularly inflected English verbs in spoken stimuli (Fonteneau et al., 2015). The observation of both synchronized slower theta as well as faster gamma oscillations in relation to complex word form processing raises intriguing possibilities as regards the mechanisms by which distant nodes of the language network might dynamically integrate their activity. In general, the phase of theta oscillations has been observed to modulate power in the gamma band in the human neocortex, and interactions between theta and gamma frequency oscillations have been proposed to facilitate communication within distributed brain systems during cognitive tasks (Canolty et al., 2006). One hypothesis, specifically formulated in the context of memory processes, is that nesting several gamma cycles within slower theta activity could provide a neural code for transmitting multi-item messages, where the theta phase encodes the onset of the message and the gamma band signifies separate items in an ordered sequence (Lisman & Jensen, 2013). Modulation of gamma band synchronization by theta phase has been reported in relation to expressive language processing (covert verb generation) (Doesburg, Vinette, Cheung, & Pang, 2012), and it would be interesting to investigate the potential role of theta-gamma interaction in dynamically coordinating engagement of the dorsal and ventral processing streams during the comprehension of complex word forms.
References


Luce, R. D. (1991). Response times: Their role in inferring elementary mental organization. New York: Oxford University Press.


Study 1
Brain responses to morphologically complex verbs: an electrophysiological study of Swedish regular and irregular past tense forms

Andrea Schremm, Mikael Novén, Merle Horne, & Mikael Roll

Abstract: The present electrophysiological study investigated irregular versus regular verb form processing in Swedish. In line with previous results, overregularized verbs, i.e. incorrect irregular stem + regular past tense suffix combinations (e.g. *stjäl+de ‘steal + past tense’), elicited a left-lateralized negativity (LAN) relative to correct irregulars (stal ‘stole’), suggesting rule-based decomposition of regularly inflected words. Lack of a similar effect for misapplication of the irregular stem formation pattern on regular verbs (e.g. *löft ‘lifted’ instead of lyfte) indicates the involvement of different processing mechanisms, presumably reflecting whole word access for irregular items. Visual inspection of ERP waveforms raised the possibility that all conditions except for well-formed irregular verbs elicited LANs, which might index morphological analysis of correct and incorrect items, including verbs irregularized through predictable stem vowel alternations. Follow-up analysis indeed detected increased long-range oscillatory synchrony for correct irregular verbs, tentatively interpreted to indicate facilitated processing via the ventral semantic stream, whereas, in the other conditions, morphological analysis might have also placed demands on the parallel dorsal stream.

Keywords: morphology; inflection; event-related potentials; oscillatory phase synchrony; left anterior negativity; P600
1. Introduction

Cognitive mechanisms and brain networks involved in processing regular and irregular word classes during language comprehension have been the subject of intense study and controversy. Investigating the issue in different languages with varying morphological properties could provide valuable insights into the way the brain handles idiosyncratic word forms as well as apparently rule-governed sequences. In the present study, we looked at the as yet unexplored question as to whether regular versus irregular verb forms in Swedish elicit distinctly different electrophysiological brain signals, which would suggest different processing mechanisms. Specifically, it is unknown whether misapplication of the Swedish regular verbal inflection would yield a left anterior negativity (LAN), an effect previously argued to be consistent with rule-based decompositional analysis of regular complex word forms in a number of languages (e.g. Morris & Holcomb, 2005; Penke et al., 1997; Rodriguez-Fornells, Clahsen, Lleó, Zaake, & Münte, 2001). Furthermore, due to the presence of predictable stem vowel patterns in the past tense formation of irregular verbs, Swedish provides an interesting test case of the question as to whether regularities characteristic of a non-default word class would engage processing mechanisms that are similar to or different from those associated with the default inflection.

In English, the past tense of the majority of verbs is predictively formed by attaching an -ed suffix to the stem (e.g. walk - walked), whereas a smaller set of words displays more idiosyncratic morphological variation (e.g. go - went). Much of the controversy in the psycholinguistic literature has focused on the question as to whether the neural processing of such morphologically complex forms actually involves two distinct mechanisms, as proposed by dual route models (Clahsen, 1999; Pinker, 1999), or alternatively, whether the production and comprehension of all inflected words are processed by a single cognitive system (e.g. Joanisse & Seidenberg, 1999). Findings of electrophysiological studies reporting different ERP patterns for regular versus irregular word form processing (Gross, Say, Kleingers, Clahsen, & Münte, 1998; Morris & Holcomb, 2005; Münte, Say, Clahsen, Schiltz, & Kutas, 1999; Newman, Ullman, Pancheva, Waligura, & Neville, 2007; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Weyerts, Penke, Dohrn, Clahsen, & Münte, 1997) have been typically interpreted to be consistent with the distinction assumed by dual route models: irregular forms are proposed to be directly accessed as whole word representations in the mental lexicon, whereas regular items might undergo rule-governed decomposition (Clahsen, 1999; Pinker, 1999). Specifically, the observation of a LAN in the electrophysiological brain signal for overregularization of verbs or nouns (go > *goed), an effect commonly seen as an index of morphosyntactic anomaly detection (Friederici, 2002), has
been argued to be suggestive of the productive operation of a combinatorial rule. Such a mechanism appears to be associated with the regular inflection, as irregularization of regular items (*walk > *welk) tends to elicit different, or no effects.

Nevertheless, the cross-linguistic generalizability of the proposed distinction between morphological processing routes is still unclear as electrophysiological data are available only for a limited set of languages (mainly English, German, Italian and Catalan), and there has been some degree of variation across studies and languages concerning, for instance, the distribution of the observed LAN effect (Krott, Baayen, & Hagoort, 2006) and the exact conditions that elicit anterior negativities. For example, in Catalan where verb stems are composed of a root as well as a thematic vowel, rule violations associated with incorrectly using the default thematic vowel within a stem (i.e. the incorrect application of the 1st conjugation theme vowel -a- to a 2nd or 3rd conjugation verb: e.g. *dorm-a-t ‘slept’ instead of correct dorm-i-t), but not the misapplication of the regular inflection (i.e. using the regular participle inflection -t in a verb with irregular participle form, e.g. *admetat ‘admitted’ instead of correct admès), resulted in a LAN (Rodriguez-Fornells et al., 2001). The specific morphological properties of the investigated language might indeed modulate the processing route engaged in a given context. Importantly, in Swedish, even members of the non-default verb class exhibit largely regular patterns in stem vowel alternations that distinguish present and past tense forms, a fact which raises the question as to whether regular versus irregular classes would still display distinctly different brain responses as seen in previous studies. Findings so far are inconclusive as regards this question, pointing towards the role of language- and morphology-specific factors: whereas in Catalan, misapplication of the non-default stem formation did not produce a LAN (Rodriguez-Fornells et al., 2001), violation of subregularities associated with irregular verb stem formation in German was reported to yield a widespread negativity preceding the syntax-related P600 effect (Regel, Opitz, Müller, & Friederici, 2015).

Furthermore, dual system views have been originally formulated with reference to English (Pinker, 1999), a language that is characterized by an especially clear-cut distinction between regular and irregular forms. Languages with relatively richer morphology, however, tend to have several noun or verb classes displaying varying degrees of regularity. In such cases, it has been argued that only one class can be considered as the default, which is in turn processed through decomposition as opposed to direct lexical access (Rodriguez-Fornells et al., 2001; Sonnenstuhl, Eisenbeiss, & Clahsen, 1999). An important characteristic of the default class has been suggested to be the lack of changes in the word stem form, and in turn a straightforward applicability of a productive suffix concatenation rule (e.g. Weyerts et al., 1997). Non-default classes might display changes in the word stem such as vowel alternations (e.g. *sit > *sat). Nevertheless, considering that
misapplication of regularities pertaining to the morphological form of stems of the default verb class yielded a left-lateralized negativity in Catalan (Rodriguez-Fornells et al., 2001), there is some indication that LANs might not only be elicited by violations of rules associated with attaching the regular suffix to an invariable stem. More recent observation of LAN-like responses to well-formed regular participles has challenged the traditional view of LAN as a signal of rule violation, raising the possibility that such negativities might in fact generally index morphological rule application (Krott & Lebib, 2013). From this perspective, LAN effects that were observed for regularized irregular words could be interpreted to reflect morphological processing as such, presumably triggered by a regular inflection.

In the present study, we investigated Swedish verb form processing by recording ERP responses to overregularized irregular versus irregularized regular verbs, visually presented in sentences (see Table 1 for the experimental conditions and examples). In Swedish, a distinction can be made between default (i.e. regular) versus non-default (i.e. irregular) verb classes based on the above described considerations. Regular verbs are inflected in the past tense with a productive suffix (-te/-de), which induces no changes in the verb stem (e.g. hjälp+te ‘help+ed’, base form: hjälp-a ‘to help’). In line with previous results, overapplication of the regular past tense suffix to an irregular stem (irregular incorrect (overregularized): *bär+de ‘bear + past tense’) is expected to elicit a LAN relative to the correct form (irregular correct: bar ‘bore’). Verbs in the irregular 4th conjugational class in Swedish display stem vowel alternations, a past tense formation pattern typically not generalized to novel words. Nevertheless, vowel changes in the stem tend to follow predictable patterns in the majority of cases (e.g. y > ò: flyger ‘fly’ > flög ‘flew’; i > a: binder ‘bind’ > band ‘bound’) (Teleman, Hellberg, & Andersson, 1999). Regular verbs were thus irregularized in the present study by changing the stem vowel on analogy to the vowel alternations in the non-default class (regular incorrect: lyfter ‘lift’ > *löft). If the resulting regular incorrect (irregularized) forms constitute unfamiliar words for Swedish speakers, with an internal structure opaque to rule-based decomposition, no LAN would be expected in this condition relative to the correct variant (regular correct: lyfte ‘lifted’) as previously seen in several studies (Gross et al. 1998; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Weyerts et al., 1997). Alternatively, the presence of a potentially recognizable regularity within the stem of regular incorrect (irregularized) words, i.e. the stem vowel alternation patterns of the irregular class, might promote some degree of morphological processing, which, considering the proposal of LAN as a general index of morphological rule application (Krott & Lebib, 2013), could conceivably be reflected in an increased negativity. In addition, since verbs were presented in sentence context, a P600 effect was predicted for both incorrect conditions, reflecting reanalysis processes
or increased difficulties integrating an incorrect verb into the sentence structure (Friederici, 2002; Kaan, Harris, Gibson, & Holcomb, 2000).

Table 1
The four experimental conditions with examples and predictions concerning the expected LAN effect.

<table>
<thead>
<tr>
<th>Correct</th>
<th>Incorrect (Irregularized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>Incorrect (Overregularized)</td>
</tr>
<tr>
<td><em>lyft+le</em> 'lift+ed'</td>
<td><em>löft</em> 'lifted' (no LAN)</td>
</tr>
<tr>
<td><em>bär+de</em> 'bear + past tense' (LAN)</td>
<td></td>
</tr>
</tbody>
</table>

In addition to analysing the ERP, a follow-up analysis was conducted in order to further elucidate the degree to which different conditions involved morphological analysis versus semantic processing based on direct lexical access. This analysis compared the experimental conditions as regards oscillatory phase synchrony between distant frontal and posterior sites in the theta (4-7 Hz) and alpha (8-12 Hz) frequency bands. Communication between two distant neural populations is assumed to be facilitated if they produce oscillatory electrophysiological activity in relative synchrony (Fries, 2005), i.e. if they oscillate with more or less stable phase differences, which could in turn reflect the formation of functional networks encompassing the anatomically connected neural groups in question (von Stein & Sarnthein, 2000). Language-relevant networks have been reported to underlie information flow between posterior and frontal sites via two parallel streams (Hickok & Poeppel, 2007; Marslen-Wilson & Tyler, 2007; Saur et al., 2008), where direct lexical access might preferentially engage the more semantic ventral stream (Hickok & Poeppel, 2007; Saur et al., 2008) and regular inflectional processing might place greater demands on the dorsal route (Marslen-Wilson & Tyler, 2007). Therefore, variation between the experimental conditions considering the observed posterior-frontal oscillatory phase synchrony could potentially reflect such relative differences in the dynamically established language processing networks involved.

2. Methods

2.1 Participants
Twenty-three right-handed Swedish native speakers (12 females) participated in the experiment. Mean age was 24.3 years, SD = 3.0. All participants gave
informed consent in accordance with the Declaration of Helsinki. The study procedures were approved by the Lund Regional Ethical Review Board (approval number 2016/871).

2.2 Stimuli

Each of the four test conditions involved 30 unique verbs with a monosyllabic stem, embedded in sentences. In the Regular correct condition, a verb from the regular (2nd) conjugational class was correctly inflected with the past tense suffix -te/-de (e.g. skrämd+de ‘frightened’). Verbs in the Regular incorrect (irregularized) condition appeared in an incorrect form (*skram+de), created by changing the vowel of the infinitive stem (‘ä’ in skrämd), on analogy to the vowel alternation (e.g. ä > a) characterizing irregular past tense verbs. The Irregular correct condition involved verbs in the irregular (4th) conjugational class in correct past tense forms (e.g. stäl ‘stole’). Verbs appearing in the Irregular incorrect (overregularized) condition were created by inflecting the infinitive stem of the irregular verb (e.g. stjäl+a ‘to steal’) with the regular past tense suffix -te/-de (stjäl+de). There was no difference between verbs in the Regular/Irregular conditions considering mean lemma frequency (t(58) = .276, p = .784) based on the Stockholm-Umeå corpus (Stockholm-Umeå Corpus SUC 1.0, 1996). The word form frequencies of correct past tense verbs in the Regular (M = 25.60 per million words, SD = 40.63) and Irregular conditions (M = 44.37 per million words, SD = 60.56) showed no significant difference (t(58) = -1.409, p = .165) in the PAROLE corpus (https://spraakbanken.gu.se/eng/resources/corpus).

Even though stem vowel changes in 4th conjugational class verbs tend to follow specific patterns, the regular-irregular distinction is made here between 2nd and 4th conjugational verbs based on the following considerations. First, belonging to the 4th conjugational class is an idiosyncratic property of the verb, which cannot be determined based on the infinitive form of the word. Regular verbs, however, always receive a specific allomorph of the past tense suffix -te/-de, depending solely on the phonological form of the stem-final segment. Moreover, unlike the stem vowel alternation rules of the 4th conjugational class, the past tense suffix -te/-de is productive, and as such can be applied, for instance, to novel verbs or loanwords (e.g. blogga+de ‘blogged’).

All test sentences had the same structure: a single word subject, followed by the target verb and a sentence final adverb or object (one or two words), e.g. Tjejen skar frukterna ‘The girl sliced the fruits’ (see Appendix for a list of target verbs with sentence examples). In addition, 40 filler sentences of identical structure, but with different lexical verbs, were included to decrease the predictability of the stimulus material. Filler items were not considered during the analysis. In a manner analogous to test item construction, verbs in filler sentences appeared in
regular correct/incorrect and irregular correct/incorrect past tense forms, 10 times per form. However, it was not the case that each lexical verb appeared once in a correct and once in an incorrect version, but rather the verb forms followed a different unpredictable pattern, so that participants could not use the strategy of expecting an incorrect/correct verb after already having seen the correct/incorrect version of the same verb stem during the experiment.

2.3 Procedure

Sentences were presented at the centre of a computer screen, in white font against a dark blue background, one phrase at a time: i.e. the subject, the target verb and the final phrase of the sentence were presented separately in succession. Stimulus onset asynchrony was 450 ms including 50 ms of interstimulus interval of blank screen. A fixation cross displayed for 1000 ms preceded each trial. The grammaticality judgement question (“Correct or Incorrect”) was shown immediately after each sentence, for 2000 ms. Participants responded by pressing a left- or a right-hand key. Hand-response associations were counterbalanced within participants. Sentences were presented in a randomized order, distributed across two blocks of 80 sentences each.

2.4 Behavioral data

Response times (RTs) were calculated from the verb presentation onset. RTs of incorrect answers (8.6% of all test items), and items that were not responded to before the offset of the question presentation (i.e. within 2900 ms following the verb, 7.8%) were excluded from the analysis. Repeated measures ANOVAs with the factor Verb correctness (correct, incorrect) were carried out on reaction time and accuracy values for irregular and regular verbs separately. Accuracy values were expressed as the proportion of correct responses to the grammaticality judgement questions, which participants were required to respond to after each sentence presentation.

2.5 EEG recording

EEG data was recorded from 32 electrodes (Braincap-MR from Easycap) using BRAINAMP MR PLUS Amplifier and Brainvision Recorder (BrainProducts), at a sampling rate of 250 Hz. Online, a high-pass filter set at 0.1 Hz and a low-pass filter with a cut-off frequency of 70 Hz were used, and impedances were kept below 5 kΩ. Recording reference was FCz. Offline, reference was re-calculated to the average of the mastoid electrodes and a 30 Hz low-pass filter was applied. A low-pass filter of 16 Hz was used for presentation only.
2.6 ERP analysis

Epochs of 1000 ms were extracted following verb presentation onset, and a prestimulus period of 200 ms was used as a baseline. Epochs where signal amplitude exceeded ±100 μV after correction for ocular artifacts using Independent Component Analysis (ICA) were discarded before averaging (3.55% of trials). There was no significant difference, \( F(3, 66) = .714, p = .547 \), between the remaining number of epochs per condition.

For statistical analysis, two time windows were created based on previous literature and visual inspection of the data: 350-500 ms for the LAN and 600-1000 ms for the P600. Mean amplitude values were submitted to repeated measures ANOVAs in each time window, separately for regular and irregular verbs and for lateral and midline sites. The factors were Correctness (correct, incorrect), Anteriority (frontal, central, posterior) and Laterality (left, right) for the analysis on lateral sites. Lateral electrodes were accordingly grouped into six regions of interest (ROIs): left anterior (F7, F3), right anterior (F4, F8), left central (T7, C3), right central (C4, T8), left posterior (P7, P3), right posterior (P4, P8). For midline analysis, the topographical factor was Electrode (Fz, Cz and Pz). Greenhouse-Geisser correction was applied where relevant.

ERP waveforms for the target verbs in all four conditions are displayed in Fig. 3 for a visual representation of the observed tendencies. A direct comparison of ERP responses elicited by regular and irregular verbs in the same statistical test is, however, somewhat problematic since these conditions involved different lexical items as target words.

2.7 Follow-up phase synchrony analysis

Phase synchrony analysis was carried out in Matlab, using Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Following correction for ocular artifacts using ICA, segments were extracted from the EEG with large padding around the event of interest (-1.5 sec to 2.5 sec relative to verb onset), and epochs were baseline corrected (-200 to 0 ms). After rejecting epochs due to artifacts based on the criteria described above for ERP analysis, a small number of randomly selected trials were further excluded (0.6% of remaining trials) to ensure an equal number of trials per condition for the PLV (Phase Locking Values) analysis. To obtain amplitude and phase estimates of the oscillations, data was filtered with a Morlet wavelet decomposition (width of 7 cycles), from 2-30 Hz, in frequency steps of 1 Hz and time steps of 12 ms from -480 to 1500 ms. Four left-hemispheric frontal-posterior electrode pairs were defined by pairing each of the two frontal lateral electrodes (F3 and F7) with each of the two posterior lateral electrodes (P3 and P7). Corresponding electrode pairs were constructed for the right hemisphere (F4-P4, F8-P4, F4-P8, F8-P8). The Fieldtrip function
‘ft_connectivityanalysis’ was used to obtain PLV values. For each time-frequency data point in the time windows (350-500 ms and 600-1000 ms) and frequency bands (4-7 Hz and 8-12 Hz) of interest, the difference between the phase angles of two signals, recorded from the two electrodes in a pair, was calculated for each trial and signal pair. These magnitude normalized complex values were used to quantify the consistency of the two signals’ phase difference across trials, in terms of a PLV index ranging from 0 to 1, where 0 signifies absence of synchrony due to completely random distribution of phase difference across trials and 1 stands for perfect synchrony with uniform phase differences. PLV values were then normalized through Fisher z-transform and averaged across each frequency band and time window per participant. For statistical analysis, the resulting PLV values were submitted to repeated measures ANOVAs, with the factors Regularity (regular, irregular), Correctness and Hemisphere (left, right).

3. Results

3.1 Behavioural results

3.1.1 Irregular verbs

The ANOVA carried out on reaction time (RT) values detected no difference between correct ($M = 1251$ ms, $SD = 217$) and incorrect (overregularized) irregular verbs ($M = 1255$ ms, $SD = 234$). The analysis on accuracy values (the proportion of correct responses to the grammaticality judgement questions) found a significant effect of Verb correctness, $F(1, 22) = 15.706, p = .001, \eta^2_p = .417$, as irregular correct items ($M = 0.92, SD = 0.07$) received more accurate responses than irregular incorrect ones ($M = 0.86, SD = 0.10$).

3.1.2 Regular verbs

The ANOVA conducted on RT values for regular verbs indicated that incorrect (irregularized) items ($M = 1210$ ms, $SD = 228$) elicited significantly faster responses, $F(1, 22) = 12.867, p = .002, \eta^2_p = .369$, than correct ones ($M = 1284$ ms, $SD = 221$). In the analysis focusing on response accuracy, the difference between correct items ($M = 0.95, SD = 0.06$) and incorrect items ($M = 0.93, SD = 0.07$) did not reach significance.
3.2 ERP

3.2.1 350-500 ms time window

3.2.1.1 Irregular verbs
Incorrect irregular (overregularized) verbs elicited an increased left-lateralized negativity relative to correct irregular verbs between 350 and 500 ms (see Fig. 1). The ANOVA conducted in this time window indicated a Correctness × Laterality interaction, $F(1, 22) = 7.764, p = .011, \eta^2_p = .261$, due to a significant effect of Correctness for irregular items over left hemisphere sites only, $F(1, 22) = 6.309, p = .020, \eta^2_p = .223$. For midline electrodes, no significant effects were obtained.

![Fig. 1](image)

Left: ERPs at 9 selected electrodes for the target verb in the Irregular correct (solid line) and the Irregular incorrect (overregularized) (dashed line) conditions. Irregular incorrect verbs elicited a LAN between 350 and 500 ms and a P600 between 600 and 1000 ms. Right: topographic maps showing the scalp distribution of the LAN (top) and the P600 effects (bottom) for irregular incorrect verbs.

3.2.1.2 Regular verbs
For regular items, no significant effects or interactions involving the factor Correctness were found in the 350-500 ms time window (see Fig. 2).
3.2.2 600-1000 ms time window

3.2.2.1 Irregular conditions

In the 600-1000 ms time window, incorrect irregular items displayed a widely distributed positivity (see Fig. 1). This yielded a Correctness × Anteriority interaction, $F(2, 44) = 16.642, p < .001, \eta^2_p = .431$, and a significant effect of Correctness over bilateral posterior sites, $F(1, 22) = 22.175, p < .001, \eta^2_p = .502$. A Correctness × Laterality interaction was also found, $F(1, 22) = 6.001, p = .023, \eta^2_p = .214$, reflecting increased positivity for the Incorrect condition across right lateral sites, $F(1, 22) = 15.118, p = .001, \eta^2_p = .407$. The analysis focusing on bilateral anterior ROIs indicated the opposite effect, where incorrect items showed an increased negativity $F(1, 22) = 6.076, p = .022, \eta^2_p = .216$.

The increased positivity associated with incorrect items was also present over midline electrodes, where follow-up analysis to a Correctness × Electrode interaction, $F(2, 44) = 23.364, p < .001, \eta^2_p = .515$, indicated enhanced positivities over the central, $F(1, 22) = 14.450, p = .001, \eta^2_p = .396$, and the posterior midline electrode, $F(1, 22) = 31.149, p < .001, \eta^2_p = .586$.

3.2.2.2 Regular conditions

Incorrect regular items showed an enhanced positivity between 600 and 1000 ms (see Fig. 2), which was reflected in a Correctness × Anteriority interaction, $F(2, 44) = 6.647, p = .014, \eta^2_p = .232$, due to a significant effect of Correctness over bilateral central $F(1, 22) = 11.253, p = .003, \eta^2_p = .338$ and posterior regions, $F(1, 22) = 12.673, p = .002, \eta^2_p = .365$.

The analysis on midline electrodes detected increased positivity for the incorrect regular verbs over the central $F(1, 22) = 8.034, p = .010, \eta^2_p = .267$, and the posterior midline electrode $F(1, 22) = 8.625, p = .008, \eta^2_p = .282$, leading to a Correctness × Electrode interaction, $F(2, 44) = 4.518, p = .038, \eta^2_p = .170$. 


Fig. 2
Left: ERPs at 9 selected electrodes for the target verb in the Regular correct (solid line) and the Regular incorrect (irregularized) (dashed line) conditions. Regular incorrect verbs elicited a P600 between 600 and 1000 ms. Right: topographic map showing the scalp distribution of the P600 effect for regular incorrect verbs.

Fig. 3
ERPs elicited by the target verb in the four experimental conditions at 9 selected electrodes. There appears to be an enhanced negativity in the LAN time window (350-500 ms) most prominent at the left frontal site F3 for Regular correct (grey solid line), Regular incorrect (irregularized) (grey dashed line) and Irregular incorrect (overregularized) (black dashed line) conditions relative to Irregular correct items (black solid line).
3.3 Oscillatory phase synchrony

The follow-up analysis focusing on phase synchrony in the theta band (4-7 Hz) in the LAN time window (350-500 ms) revealed significant differences between the Irregular conditions in the left hemisphere, where irregular correct items showed increased synchrony relative to irregular incorrect (overregularized) items (see Fig. 4). The fact that the effect was confined to the left hemisphere was seen in a Regularity × Correctness × Hemisphere interaction, $F(1, 22) = 5.607, p = .027, \eta^2_p = .203$, and a Regularity × Correctness interaction present over left hemisphere electrode pairs only, $F(1, 22) = 7.092, p = .014, \eta^2_p = .244$, where an effect of Correctness reached significance for irregular items, $F(1, 22) = 6.493, p = .018, \eta^2_p = .228$. No significant effects were obtained over the right hemisphere.

In the P600 time window (600-1000 ms), increased phase synchrony for the Irregular correct condition relative to the Irregular incorrect condition was present bilaterally, $F(1, 22) = 18.577, p < .001, \eta^2_p = .458$, yielding a Regularity × Correctness interaction across left and right hemisphere electrode pairs, $F(1, 22) = 9.324, p = .006, \eta^2_p = .298$ (see Fig. 4). No significant or marginal differences were obtained involving Regular conditions.

Analysis conducted on phase synchrony values in the alpha band (8-12 Hz) indicated no significant effects involving any of the factors Regularity or Correctness, in the LAN or the P600 time window.
4. Discussion

The present study investigated Swedish speakers’ ERP responses to correct regular and irregular verbs as well as irregularized and overregularized verb forms. The obtained results were largely in line with findings of previous studies on regular/irregular morphology processing in a number of other languages (Gross et al. 1998; Morris & Holcomb, 2005; Penke et al., 1997; Rodriguez-Fornells et al., 2001; Weyerts et al., 1997). Attaching a regular suffix to an irregular verb stem yielded an increased left-lateralized negativity relative to the correct irregular form, between 350-500 ms (Fig. 1). This effect, most prominent at left anterior and central sites, is interpreted as a LAN, which has commonly been seen as a signal of morphological anomaly detection in the regular/irregular word processing paradigm. It has been argued that overregularized word forms elicit a
LAN as a result of the misapplication of a suffixation rule to a stem that normally blocks the default morphological mechanism due to the presence of stored whole word representations of irregular inflected forms in the mental lexicon (Penke et al., 1997; Pinker & Ullman, 2002; Rodriguez-Fornells et al., 2001). The LAN observed in the present case thus implicates that regular Swedish verbs might undergo morphological parsing based on the default past tense formation rule. No effect was obtained for irregularized regular verbs relative to correct regular items in the early time window (Fig. 2), indicating that misapplication of the irregular stem formation pattern to a regular verb was apparently not processed as a morphological rule violation. This interpretation would suggest that the vowel alternations characterizing the irregular past tense verbs in Swedish, even if these are largely predictable, might not be treated as productive morphological regularities, and that existing irregular forms are rather accessed as whole words. The findings are thus in general consistent with dual system models, assuming two distinct mechanisms for the processing of morphologically complex forms: irregular words are directly retrieved from the mental lexicon in their inflected form, whereas regular words might be decomposed into stem and suffix parts (Clahsen, 1999; Pinker, 1999). Finally, as expected, incorrect verbs elicited a late positivity between 600-1000 ms, a P600 (Fig. 1 and Fig. 2), possibly reflecting reanalysis of sentence structure or increased processing efforts during integration of the incorrect verb forms (Friederici, 2002; Kaan et al., 2000). The anterior negativity observed in the P600 time window for incorrect irregular items, but not for incorrect regular ones, might potentially represent the continuation of the LAN obtained in the irregular condition, or alternatively, the dipole inversion of the posterior P600 effect.

Visual inspection of waveforms of all four conditions presented in a single plot (Fig. 3) raises the possibility of an alternative interpretation of the effect obtained in the LAN time window. Not only irregular incorrect (overregularized) verbs but also regular correct and regular incorrect (irregularized) items appear to display an increased negativity relative to irregular correct verbs as shown in Fig. 3. Observation of a left-lateralized negativity even for correct words would be consistent with the proposed re-interpretation of the LAN as an effect signalling morphological rule application in general (Krott & Lebib, 2013). The presence of a regular inflection could be thought to have triggered rule-based decompositional processing of regular and overregularized verbs. Considering the fact that it was possible for test participants to restore an existing verb stem from the irregularized incorrect verb forms, by changing the stem based on vowel alternations that tend to occur in irregular verbs (e.g. *löft > lyfte), morphological processing could conceivably have taken place in that condition as well. An increased N400, an effect previously observed for non-words, has been reported for irregularized incorrect items (for applying irregular plural suffixes to regular German nouns in Weyerts et al., 1997). The lack of an N400 effect in the present case is consistent
with the proposal that irregularized verb forms were not simply treated as pseudowords, inaccessible for decomposition. At the same time, it is reasonable to assume that the limited set of relatively frequent Swedish irregular verbs is stored as whole words in the mental lexicon for fast and effective access, obviating the need for morphological decomposition, which would explain the absence of a negativity in the Irregular correct condition.

To further investigate the proposed alternative interpretation of the results, i.e. that the processing of all tested verb forms, except for correct irregular verbs, involved morphological decomposition; we conducted a follow-up oscillatory phase synchrony analysis. As a starting point, we considered the largely accepted view that language comprehension engages two parallel streams, a ventral and a dorsal one, projecting from posterior language-relevant regions to frontal areas (e.g. Hickok & Poeppel, 2007; Saur et al., 2008; Weiller, Bormann, Saur, Musso, & Rijntjes, 2011). While different proposals vary as regards the specific details, the ventral stream is generally assumed to play an important role in the semantic processing of language (Hagoort, 2013; Hickok & Poeppel, 2007; Saur et al., 2008). The dorsal stream has been associated with a variety of functions such as sensory-motor mapping (Hickok-Poeppel, 2007; Saur et al., 2008) and syntactic processing (Hagoort, 2013), and has been reported to underlie decompositional analysis related to inflectional morphemes (Bozic, Fonteneau, Su, & Marslen-Wilson, 2015; Marslen-Wilson & Tyler, 2007; Rolheiser, Stamatakis, & Tyler, 2011). Thus, language processing involves a flow of information across the Sylvian fissure, and it can be hypothesized that well-formed irregular verbs engage large-scale language networks that are partially distinct from those underlying comprehension of the other verb forms tested, due to the relative difference in demands placed on functional nodes crucially involved in whole word semantic processing versus morphological analysis. Neural oscillations can provide an insight into functional network formation. Neuronal populations generate oscillating electrophysiological activity, displaying a rhythmic variation in neuronal excitability, which creates specific windows of effective communication (Fries, 2005). Therefore, synchronization between distinct neuronal groups through oscillatory phase locking could support the transmission of information across anatomically connected neuronal populations (Fries, 2005). Low-frequency oscillations have been specifically reported to play a role in long-range interactions (e.g. Sauseng, Klimesch, Schabus, & Doppelmayr, 2005; von Stein & Sarnthein, 2000), and there is evidence that implicates theta and alpha synchronization in organizing simultaneous activities of several distinct populations into an extended network (von Stein & Sarnthein, 2000).

The follow-up analysis compared the experimental conditions in relation to neural synchrony (estimated through phase locking values) between posterior and frontal electrode sites, in the theta and alpha frequency bands, which could be assumed to potentially tap into long-range information flow along the language
processing streams. Results indicated effects in the theta band only, where irregular correct words displayed increased coherence relative to irregular incorrect (overregularized) words. The increase in coherence was left-lateralized in the LAN time window and bilaterally present in the P600 time range (Fig. 4). Due to the poor spatial resolution of EEG, it is not possible to identify the exact brain regions involved in the synchronization patterns observed here and, therefore, the interpretation of the results can only be highly tentative. With this limitation in mind, we suggest that the significant difference between the Irregular conditions could be taken to indicate relative differences in the language networks engaged. As irregular correct words are assumed to be directly accessed in the lexicon, the observed increase in synchronization could be interpreted as a greater reliance on the ventral, semantic stream. The bilateral distribution of the effect in the later time window supports this assumption, since the ventral stream is generally considered to be less left-lateralized (Hickok & Poeppel, 2007). Overregularized (irregular incorrect) verbs might therefore potentially engage the dorsal processing stream relatively more due to inflectional processing triggered by the regular suffix (Bozic et al., 2015; Marslen-Wilson & Tyler, 2007). The fact that no significant differences were found between regular correct versus regular incorrect (irregularized) verbs might suggest that they are more or less similarly processed along the language streams. From this perspective, the findings are less consistent with the assumption that irregularized verbs would be treated as unfamiliar words, inaccessible for internal analysis, and only correct regular verbs would be expected to undergo morphological parsing. Instead, considering the possibility that incorrect items might be interpreted through the application of stem formation regularities, processing could conceivably involve a certain degree of reliance on the dorsal stream for both regular and irregularized cases. Presumably, there is additional significant parallel engagement of the ventral route, for instance, during retrieval of the extracted verb stem from the lexicon. Generally, this proposed interpretation of the oscillatory synchrony patterns is consistent with a view of the ERP results, where increased morphological processing demands associated with all experimental conditions except for irregular correct words resulted in LAN-like effects.

In sum, overregularized verbs in Swedish elicited a LAN effect, which, in line with previous studies, was interpreted to suggest detection of morphological violation, and in turn rule-based processing related to the regular inflection. Subsequent analysis of long-range oscillatory synchronization patterns yielded results consistent with an alternative interpretation of the ERP patterns observed, raising the possibility that morphological analysis in general, assumed not only for the processing of overregularized words, but also for the processing of regular correct as well as irregularized verbs, could be reflected in enhanced LANs. Under both interpretations, the results provide support for the assumption that both whole
word access as well as decompositional processing are available during the comprehension of morphologically complex items in Swedish.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (grant number 2014.0139); Marcus and Amalia Wallenberg Foundation (grant number 2014.0039); and the Swedish Research Council (grant number 2011-2284).
References


Münte, T. F., Say, T., Clahsen, H., Schiltz, K., & Kutas, M. (1999). Decomposition of morphologically complex words in English: Evidence from event-related brain...


## Appendix

### List of verbs and sentences in the four experimental conditions

<table>
<thead>
<tr>
<th>Target verb</th>
<th>Correct infinitive</th>
<th>Sentence example</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>blåste</td>
<td>blåsa</td>
<td>Kvinnan blåste i bilen</td>
<td>Regular correct</td>
</tr>
<tr>
<td>brände sig</td>
<td>bränna</td>
<td>Tjejen brände sig på morgnarna</td>
<td>Regular correct</td>
</tr>
<tr>
<td>byggde</td>
<td>bygga</td>
<td>Kvinnan byggde på morgnarna</td>
<td>Regular correct</td>
</tr>
<tr>
<td>fräste</td>
<td>fräsa</td>
<td>Chefen fräste varje dag</td>
<td>Regular correct</td>
</tr>
<tr>
<td>gifte sig</td>
<td>gifta</td>
<td>Lärarna gifte sig i parken</td>
<td>Regular correct</td>
</tr>
<tr>
<td>grävde</td>
<td>gräva</td>
<td>Barnet grävde varje helg</td>
<td>Regular correct</td>
</tr>
<tr>
<td>hjälpte</td>
<td>hjälpa</td>
<td>Lärarna hjälpte varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>klippte</td>
<td>klippa</td>
<td>Paret klippte varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>kysste</td>
<td>kyssa</td>
<td>Lärarna kysste varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>kände</td>
<td>känna</td>
<td>Killen kände mannen</td>
<td>Regular correct</td>
</tr>
<tr>
<td>lyfte</td>
<td>lyfta</td>
<td>Killen lyfta stolen</td>
<td>Regular correct</td>
</tr>
<tr>
<td>läste</td>
<td>läsa</td>
<td>Killen läste dörren</td>
<td>Regular correct</td>
</tr>
<tr>
<td>läste</td>
<td>läsa</td>
<td>Barnet läste varje dag</td>
<td>Regular correct</td>
</tr>
<tr>
<td>miste</td>
<td>mista</td>
<td>Tjejen miste telefonen</td>
<td>Regular correct</td>
</tr>
<tr>
<td>nämnde</td>
<td>nämna</td>
<td>Kollegan nämnde problemet</td>
<td>Regular correct</td>
</tr>
<tr>
<td>ringde</td>
<td>ringa</td>
<td>Killen ringde på morgnarna</td>
<td>Regular correct</td>
</tr>
<tr>
<td>rymde</td>
<td>rymma</td>
<td>Läraren rymde på kvällarna</td>
<td>Regular correct</td>
</tr>
<tr>
<td>skrämde</td>
<td>skrämma</td>
<td>Barnen skrämde varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>slängde</td>
<td>slänga</td>
<td>Paret slängde soporna</td>
<td>Regular correct</td>
</tr>
<tr>
<td>sträckte</td>
<td>sträcka</td>
<td>Läraren sträckte på benen</td>
<td>Regular correct</td>
</tr>
<tr>
<td>stämde</td>
<td>stämma</td>
<td>Lärarna stämde varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>stängde</td>
<td>stänga</td>
<td>Tjejen stängde dörren</td>
<td>Regular correct</td>
</tr>
<tr>
<td>stärkte</td>
<td>stärka</td>
<td>Paret stärkte varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>svängde</td>
<td>svänga</td>
<td>Chefen svängde ibland</td>
<td>Regular correct</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-----------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>tydde</td>
<td>tyda</td>
<td>Läraren tydde texten</td>
<td>Regular correct</td>
</tr>
<tr>
<td>tåldé</td>
<td>tåla</td>
<td>Barnen tålte varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>tänkte</td>
<td>tänka</td>
<td>Chefen tänkte på stan</td>
<td>Regular correct</td>
</tr>
<tr>
<td>väckte</td>
<td>väcka</td>
<td>Barnen väckte varandra</td>
<td>Regular correct</td>
</tr>
<tr>
<td>vägde sig</td>
<td>väga</td>
<td>Chefen vägde sig varje helg</td>
<td>Regular correct</td>
</tr>
<tr>
<td>värmdé sig</td>
<td>värma</td>
<td>Killen värmdé sig på tåget</td>
<td>Regular correct</td>
</tr>
<tr>
<td>bläs</td>
<td>blåsa</td>
<td>Kvinnan bläs i bilen</td>
<td>Regular correct</td>
</tr>
<tr>
<td>brann sig</td>
<td>bränna</td>
<td>Tjejen brann sig på morgnarna</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>bögg</td>
<td>bygga</td>
<td>Kvinnan bögg på morgnarna</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>fras</td>
<td>fräsa</td>
<td>Chefen fras varje dag</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>gjaft sig</td>
<td>gifta</td>
<td>Lärarna gjaft sig i parken</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>grav</td>
<td>gräva</td>
<td>Barnet grav varje helg</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>hjalp</td>
<td>hjälpa</td>
<td>Lärarna hjalp varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>klapp</td>
<td>klippa</td>
<td>Paret klapp varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>köss</td>
<td>kyssa</td>
<td>Lärarna köss varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>kjann</td>
<td>känta</td>
<td>Killen kjann mannen</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>löft</td>
<td>lyfta</td>
<td>Killen löft stolen</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>läs</td>
<td>låsa</td>
<td>Killen läs dörren</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>las</td>
<td>läsa</td>
<td>Barnet las varje dag</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>mast</td>
<td>mista</td>
<td>Tjejen mast telefonen</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>namn</td>
<td>nämna</td>
<td>Kollegan nämna problemet</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>rang</td>
<td>ringa</td>
<td>Killen rang på morgnarna</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>röm</td>
<td>rymma</td>
<td>Läraren röm på kvällarna</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>skram</td>
<td>skrämma</td>
<td>Barnen skram varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>slang</td>
<td>slånga</td>
<td>Paret slang soporna</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>strack</td>
<td>sträcka</td>
<td>Läraren strack på benen</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>stam</td>
<td>stämma</td>
<td>Lärarna stam varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>stang</td>
<td>stänga</td>
<td>Tjejen stang dörren</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>stark</td>
<td>stärka</td>
<td>Paret stark varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>svang</td>
<td>svänga</td>
<td>Chefen svang ibland</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>tödd</td>
<td>tyda</td>
<td>Läraren tödd texten</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>täl</td>
<td>tåla</td>
<td>Barnen täl varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>tank</td>
<td>tänka</td>
<td>Chefen tank på stan</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>vack</td>
<td>väcka</td>
<td>Barnen vack varandra</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>vag sig</td>
<td>väga</td>
<td>Chefen vag sig varje helg</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>varm sig</td>
<td>värma</td>
<td>Killen varm sig på tåget</td>
<td>Regular incorrect</td>
</tr>
<tr>
<td>band</td>
<td>binda</td>
<td>Killen band snöret</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>bjöd</td>
<td>bjuda</td>
<td>Killen bjöd på maten</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>bar</td>
<td>bära</td>
<td>Paret bar hunden</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>drack</td>
<td>dricka</td>
<td>Chefen drack på morgnarna</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>dög</td>
<td>duga</td>
<td>Tjejen dög på jobbet</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>föll</td>
<td>falla</td>
<td>Kvinnan föll på tåget</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>fann</td>
<td>finna</td>
<td>Barnen fann varandra</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>flög</td>
<td>flyga</td>
<td>Mannen flög ibland</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>hann</td>
<td>hinna</td>
<td>Läraren hann på kvällarna</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>höll</td>
<td>hålla</td>
<td>Läraren höll telefonen</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>klev</td>
<td>kliva</td>
<td>Paret klev på morgnarna</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>knöt</td>
<td>knyta</td>
<td>Barnen knöt knuten</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>kröp</td>
<td>krypa</td>
<td>Chefen kröp på morgnarna</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>lág</td>
<td>ligga</td>
<td>Kollegan låg på soffan</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>ljög</td>
<td>ljuga</td>
<td>Mannen ljög på stan</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>sjöng</td>
<td>sjunga</td>
<td>Läraren sjöng varje dag</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>skrek</td>
<td>skrika</td>
<td>Barnet skrek på tåget</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>skrev</td>
<td>skriva</td>
<td>Kollegan skrev på fritiden</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>skar</td>
<td>skära</td>
<td>Tjejen skar frukterna</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>slapp</td>
<td>slippa</td>
<td>Mannen slapp mötet</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>smög</td>
<td>smyga</td>
<td>Kvinnan smög på tåget</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>sprang</td>
<td>springa</td>
<td>Paret sprang på tåget</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>----------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>stal</td>
<td>stjäla</td>
<td>Kvinnan stal på stan</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>söp</td>
<td>supa</td>
<td>Mannen söp i bilen</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>svalt</td>
<td>svälta</td>
<td>Kollegen svalt varje dag</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>svor</td>
<td>svära</td>
<td>Chefen svor på stan</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>teg</td>
<td>tiga</td>
<td>Tjejen teg på jobbet</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>vek</td>
<td>vika</td>
<td>Killen vek pappret</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>vann</td>
<td>vinna</td>
<td>Chefen vann varje helg</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>vred</td>
<td>vrida</td>
<td>Kvinnan vred på huvudet</td>
<td>Irregular correct</td>
</tr>
<tr>
<td>binde</td>
<td>binda</td>
<td>Killen binde snöret</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>bjudde</td>
<td>bjuda</td>
<td>Killen bjudde på maten</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>bärde</td>
<td>bära</td>
<td>Paret bärde hunden</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>drickte</td>
<td>dricka</td>
<td>Chefen drickte på morgnarna</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>dugde</td>
<td>duga</td>
<td>Tjejen dugde på jobbet</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>fallde</td>
<td>falla</td>
<td>Kvinnan fallde på tåget</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>finde</td>
<td>finna</td>
<td>Barnen finde varandra</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>flygde</td>
<td>flyga</td>
<td>Mannen flygde ibland</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>hinde</td>
<td>hinna</td>
<td>Läraren hinde på kvällarna</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>hållde</td>
<td>hålla</td>
<td>Läraren hållde telefonen</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>kliivde</td>
<td>kliva</td>
<td>Paret klivde på morgnarna</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>knytte</td>
<td>knyta</td>
<td>Barnen knytte knuten</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>krypte</td>
<td>krypa</td>
<td>Chefen krypta på morgnarna</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>liggde</td>
<td>ligga</td>
<td>Kollegan liggde på soffan</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>ljugde</td>
<td>ljuga</td>
<td>Mannen ljugde på stan</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>sjungde</td>
<td>sjunga</td>
<td>Läraren sjungde varje dag</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>skrike</td>
<td>skrika</td>
<td>Barnet skrike på tåget</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>skrivde</td>
<td>skriva</td>
<td>Kollegan skrivde på fritiden</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>skärde</td>
<td>skära</td>
<td>Tjejen skärde frukterna</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>slippte</td>
<td>slippa</td>
<td>Mannen slippte mötet</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>verb</td>
<td>verb</td>
<td>phrase</td>
<td>irregularity</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>----------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>smygde</td>
<td>smyga</td>
<td>Kvinnan smygde på tåget</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>springde</td>
<td>springa</td>
<td>Paret springde på tåget</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>stjälde</td>
<td>stjäla</td>
<td>Kvinnan stjälde på stan</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>supte</td>
<td>supa</td>
<td>Mannen supte i bilen</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>svälte</td>
<td>svälta</td>
<td>Kollegan svälte varje dag</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>svärde</td>
<td>svära</td>
<td>Chefen svärde på stan</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>tigde</td>
<td>tiga</td>
<td>Tjejen tigde på jobbet</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>vikte</td>
<td>vika</td>
<td>Killen vikte pappret</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>vinde</td>
<td>vinna</td>
<td>Chefen vinde varje helg</td>
<td>Irregular incorrect</td>
</tr>
<tr>
<td>vridde</td>
<td>vrida</td>
<td>Kvinnan vridde på huvudet</td>
<td>Irregular incorrect</td>
</tr>
</tbody>
</table>
Cortical thickness of planum temporale and pars opercularis in native language tone processing

Andrea Schremm, Mikael Novén, Merle Horne, Pelle Söderström, Danielle van Westen, Mikael Roll

*Department of Linguistics and Phonetics, Lund University, Box 201, 22100 Lund, Sweden
Department of Diagnostic Radiology, Clinical Sciences, Lund University, Box 201, 22100 Lund, Sweden

Abstract

The present study investigated the relationship between linguistic tone processing and cortical thickness of bilateral planum temporale (PT) and pars opercularis of the inferior frontal gyrus (IFGpo). Swedish tones on word stems function as cues to upcoming endings. Correlating structural brain imaging data with participants’ response time patterns for suffixes, we found that thicker cortex in the left PT was associated with greater reliance on tones to anticipate upcoming inflections on real words. On inflected pseudoword stems, however, the cortical thickness of left IFGpo was associated with tone-suffix processing. Thus cortical thickness of the left PT might play a role in processing tones as part of stored representations for familiar speech segments, most likely when inflected forms are accessed as whole words. In the absence of stored representations, listeners might need to rely on morphosyntactic rules specifying tone-suffix associations, potentially facilitated by greater cortical thickness of left IFGpo.

ARTICLE INFO

Keywords:
Cortical thickness
Linguistic tone
Planum temporale
Pars opercularis

1. Introduction

General cognitive abilities have been found to be related to the cortical thickness in associated brain areas (Karama et al., 2009). Similarly, the structure of primary auditory regions in the left hemisphere has been observed to affect non-native lexical tone-learning ability (Wong et al., 2008). To date, however, it is not known how individual differences in brain morphology are related to word tone processing in native speakers. It is difficult to quantify the use of lexical tones in languages such as Chinese, since tones change word meaning like any other phonemes: as a word pronounced with a certain tone pattern is either identified as a specific lexical item or not, variation in tone processing beyond word recognition success is not straightforward to observe. In Swedish and Norwegian, however, tones on word stems are associated with suffixes (Riad, 2014). For example, the word bil ‘car’ is pronounced with accent 1 (a low tone) if it appears with the singular suffix -en, but with accent 2 (a high tone) if it ends in the plural suffix -ar. Accordingly, tones can be used to speed up suffix processing, generating an increased pre-activation negativity (PrAN) in the electrical brain potentials (Roll et al., 2015; Söderström, Horne, Mannfolk, van Westen, & Roll, 2017). PrAN has been found to increase gradually as the number of possible word completions decreased, suggesting that it reflects predictive activation of memory traces, modulated by the certainty that an upcoming continuation will appear (Söderström, Horne, Frid, & Roll, 2016). Swedish tones are thus not associated with a categorical lexical distinction but with a graded process related to the degree of pre-activation, providing a unique opportunity to quantify the use of tones in native speech comprehension. In an experimental context, if the wrong tone-suffix association is presented, participants would be expected to take a longer time to recognize the suffix the more they rely on the tonal cue to process the ending. Thus, macrostructural differences in auditory association areas in the left planum temporale (PT) region, previously observed to be involved in processing tones on inflected Swedish word stems (Roll et al., 2015), as well as in Chinese and Thai tone processing (Xu et al., 2006), might be expected to be related to performance. Since Swedish speakers have been reported to process most inflected real words similar to monomorphic ones, i.e. by accessing full forms in the mental lexicon (Lehtonen, Niska, Wande, Niemi, & Laine, 2006), reliance on the tonal cue might depend on the degree to which the tone is incorporated in a whole word representation. In the absence of lexical content, however, participants can be assumed to rely on morphosyntactic rules for stem tone-suffix combinations in order to optimally process inflections, since whole-word representations are unavailable.
That Swedish tones also activate areas associated with the processing of word structure was particularly obvious in a study involving pseudo-words, where haemodynamic responses indicated a neural network partially distinct from that observed for real words (Roll et al., 2015), with activation centering around the pars opercularis of the left inferior frontal gyrus (IFGpo) (Söderström et al., 2017). The present study tested the relation between cortical thickness in bilateral PT and IFGpo and native speakers’ use of tones for suffix pre-activation in both real words and pseudowords.

Measures of cortical thickness can be obtained by calculating the shortest distance between the white matter surface to the pial surface, constituting the border between the grey matter and the cerebrospinal fluid (Fischl & Dale, 2000). As postulated by the radial unit hypothesis (Rakic, 1988), the human cortex is characterized by a columnar organization, established during embryonic development by migration of cells along radial glial guides towards their final location. Cortical thickness can thus be related to the number and size of neurons within a column (Rakic, 1988). The number and spacing of columns affect cortical surface area instead, implicating area as a distinct feature of cortical structure (Meyer, Liem, Hirsgser, Jancke, & Hänggi, 2014; Rakic, 1988).

Structural brain imaging data and response time values collected in two previous experiments with two different participant groups, focusing on processing Swedish tone-suffix associations in real words (Roll et al., 2015) as opposed to pseudowords (Söderström et al., 2017), were analysed. In both experiments, test items involved stimuli carrying a singular or plural suffix, which was either validly cued by its preceding associated tone or invalidly cued by a tone associated with a different suffix. A relative increase in response time from validly to invalidly cued suffixes can be assumed to reflect the degree to which the perceived ending disconfirmed the tone-induced prediction: the more the listener anticipated the validly cued number inflection based on the stem tone, the greater the expected processing cost for the invalidly cued suffix. In the present study, we investigated the association between response times and cortical thickness of bilateral PT and IFGpo. It was hypothesized that greater reliance on tonal cues on real words, when tonal patterns are incorporated into whole word representations, would be associated with variation in cortical thickness in left PT, in accordance with the central role of this area in tone processing in Swedish words (Roll et al., 2015). Pseudowords, however, do not have any existing word representations, and therefore their processing cannot rely on a tone-word form association. Since the experimental task involved a decision between singular versus plural means carried by inflectional suffixes but did not require interpretation of the meaning of target word stems, response times are not assumed to primarily reflect different semantic processing of real words and pseudowords due to the presence versus absence of familiar semantic content. Instead, pseudoword response time patterns would rather depend on the efficiency of tone-induced morphosyntactic rule application. Decompositional morphosyntactic processes underlying the extraction and interpretation of regular inflectional affixes seem to rely on an intact left inferior frontal gyrus (IFG) (Biotic, Fonteneau, Su, & Marslen-Wilson, 2015). Generally, the left IFG has been argued to function as a unification space, maintaining fragments of syntactic, semantic and phonological information retrieved from memory and assembling these into coherent representations. Unification processes are assumed to take place even below the phrasal level, by which word forms are composed from and decomposed into stem and affix parts (Hagoort, 2013). Previous studies indeed found areas of left IFG, specifically the pars opercularis or pars triangularis, to be crucially involved in inflectional morpheme processing (Biotic et al., 2015; Tyler, Stamatakis, Post, Randall, & Marslen-Wilson, 2005). The present study found that in the area that showed strong activation for processing inflected Swedish pseudowords (Söderström et al., 2017).

In order to develop an expectation for the direction of the relationship between tone processing and cortical structure, electrophysiological measurements obtained in Roll et al. (2015) as well as Söderström et al. (2017) were related to cortical thickness values (Supplementary material). As mentioned above, greater tone-induced suffix pre-activation was associated with an increased negativity in the electrical brain response in these studies. Electrophysiological measurements rely on synchronous firing of neurons; thus, increased electrophysiological signal might conceivably be generated by a larger number of neurons within a thicker cortex, or alternatively by more synchronized activity of neurons in a fine-tuned network of a thinner cortex (Liem, Zaelbe, Burkhard, Jäncke, & Meyer, 2012). The present analysis indicated that an increased pre-activation negativity for real words tended to be associated with thicker cortex in the left PT. We therefore predicted a positive association between tone processing performance and cortical thickness in the present study. As a complementary measurement to cortical thickness, surface area of bilateral PT and IFGpo was also related to tone processing performance in a supplementary analysis (see Supplementary material).

2. Results

2.1. Real word processing

Individual participants’ RT advantage for valid over invalid suffixes showed a significant positive correlation with average cortical thickness in the left PT ($r = .559, p = .030$), indicating that the thicker the cortex in left PT, the greater was the RT increase for invalidly cued suffixes (see Fig. 1). Cortical thickness in the left IFGpo did not correlate with RTs ($r = .134, p = .634$). No significant correlation was obtained between RTs and right PT ($r = .197, p = .481$) or right IFGpo ($r = -.097, p = .732$).

Participants responded significantly faster ($t(14) = 7.348, p < .001$) to validity ($M = 584$ ms, $SD = 172$) as compared to invalidly cued suffixes ($M = 656$ ms, $SD = 174$). As a follow-up analysis, participants were divided into two groups of equal size ($n = 7$) based on their mean cortical thickness in the left PT, resulting in a relatively thinner ($2.117\pm2.465$ mm) and a relatively thicker cortex group ($2.590\pm2.912$ mm). An independent-samples t-test conducted on average RT speed for validly cued suffixes indicated marginally faster RTs ($t(12) = 1.813, p = .054$) for the thinner PT cortex group ($M = 504$ ms, $SD = 89$) as compared to the thinner PT cortex group ($M = 664$ ms, $SD = 215$).

2.2. Pseudoword processing

There was a significant positive correlation between RT advantage for validly versus invalidly cued suffixes and average cortical thickness of the left IFGpo ($r = .492, p = .045$) (see Fig. 2). No significant correlation was obtained between RTs and left PT ($r = .071, p = .787$). There was no significant correlation between RTs and cortical thickness in right PT ($r = .135, p = .606$) or right IFGpo ($r = .303, p = .237$). RTs were significantly faster ($t(16) = 6.497, p < .001$) to validly cued suffixes ($M = 888$ ms, $SD = 258$) than to invalidly cued suffixes ($M = 979$ ms, $SD = 232$). Subsequently, a relatively thinner cortex group ($2.426\pm2.742$ mm, $n = 8$) and a relatively thicker cortex group ($2.758\pm2.992$ mm, $n = 8$) were created based on participants’ average cortical thickness in the left IFGpo. Results of an independent-samples t-test showed significantly faster RTs to validly cued suffixes ($t(14) = 1.926, p = .039$) in the thicker cortex group ($M = 763$ ms, $SD = 288$) relative to the thinner cortex group ($M = 997$ ms, $SD = 188$).

3. Discussion

Results indicated a relationship between linguistic tone processing in native speakers and cortical thickness of specific brain areas assumed to subserve these processes. Tones in Swedish are realized on word...
stems, and like lexical tones, are modulated by PT in the left hemisphere (Moen, 1993; Roll et al., 2015). Swedish tones function as cues to upcoming suffixes, and thus they engage left frontal cortical areas implicated in regular inflectional morpheme processing as well (Bozic et al., 2015; Tyler et al., 2005), with left IFG pars opercularis emerging as the major site of activation when tone-suffix connections were implemented on meaningless pseudowords (Söderström et al., 2017). In the present study, the degree to which listeners relied on tones to anticipate morphosyntactic structure was quantified by measuring the relative response time increase for invalid tone-suffix associations. Results showed that the thicker the cortex was in the left PT, the greater the RT increase was for invalidly cued suffixes on real words. No such correlation was found for inflected pseudowords, where larger RT increase for invalidly cued suffixes was instead associated with greater cortical thickness of left IFG pars opercularis.

These findings suggest that relatively greater cortical thickness of left PT is related to tonal cue processing, in ways specific to familiar lexical items. Furthermore, the same area showed functional activation for tone processing in the same subjects (Roll et al., 2015). Indeed, regions of the posterior superior temporal gyrus (STG) have been implicated in higher-level processes of acoustic analysis during speech perception (Chang et al., 2010; Xu et al., 2006) and left PT has been found to respond to phonemic changes in the native language (Jacquemot, Pallier, LeBihan, Dehaene, & Dupoux, 2003). The present findings indicate that the nature of the representations involved in processing related to the cortical thickness of left PT might correspond to linguistic units larger than phonemes, since the pseudowords tested consisted of Swedish phonemes, just like real word stimuli. For instance, one might assume that listeners develop memory traces for frequently occurring patterns such as syllables, supporting rapid analysis of native language input. This would constitute an important difference between the target words of the two experiments as none of the pseudoword stem syllables used in the stimulus material appeared in any real Swedish words. Thus, cortical thickness of left PT might play a role in decoding the speech signal in terms of chunks larger than phonemes – possibly syllables – incorporating tonal information in Swedish, when linguistically relevant tone patterns are processed as part of a left-lateralized network. From this perspective, the cortical thickness of left PT might be related to the efficiency with which tonal information is accessed and, in turn, tone-associated endings are subsequently pre-activated on real words, by supporting recognition of familiar sound patterns involving tones. In line with this assumption, cortical thickness of left PT was associated not only with the degree to which listeners used tonal information to activate real words but also, for the subgroup of participants with relatively thicker left PT cortex, showed a tendency to be related to faster suffix recognition as
compared to the subgroup with thinner left PT cortex.

Swedish word accents are realized over several segments, in terms of relative alternations in fundamental frequency, constituting slowly changing acoustic features as compared to consonant transitions. According to the AST hypothesis ( Poeppel, 2003 ), the duration of the integration windows used to chunk the incoming speech signal in non-primary auditory areas of the left hemisphere is around 20–50 ms, underlyng sensitivity to fast changing segmental information, such as consonant transitions. Slower, 150–250 ms time windows in the right hemisphere efficiently capture suprasegmental information in the speech signal, including tones and intonation. In line with this proposal, gradually decreasing the integrity of slowly changing acoustic information in the speech input was reported to shift lateralization in the PT to the right hemisphere ( Lieen, Hurschler, Jancke, & Meyer, 2014 ). Harasty, Seldon, Chan, Halliday, and Harding (2003) suggested that an expansion of the cortex in the left PT, possibly driven by an increase in the underlying white matter volume, could account for the well-established leftward macrostructural asymmetry of the PT. The expansion resulted in a larger surface area but thinner cortex on the left relative to smaller surface area but thicker cortex on the right ( Harasty et al., 2003 ).

Larger surface area of the left auditory-related region has been associated with the left-hemispheric advantage for rapid acoustic analysis: in the expanded cortex, the distance between neuronal columns might be greater and dendritic trees less overlapping. In such an organization, columns might be able to function more independently, performing finely differentiated processing of the incoming auditory signal ( Harasty et al., 2003 ; Meyer et al., 2014 ). Conversely, greater overlap among columns in the relatively thicker right hemisphere could result in more holistic and temporally coarse-grained analysis ( Harasty et al., 2003 ).

Thus, auditory-related regions of the right hemisphere might be fine-tuned for processing the kind of spectral information that differentiates Swedish word accents. Nevertheless, due to their strong integration in a left-lateralized morphosyntactic system, word accents appear to be processed predominantly on the left side ( Roll et al., 2015 ). The AST hypothesis conceptualizes hemispheric differences regarding temporal resolution as a relative phenomenon, since both hemispheres are assumed to contain neuronal populations underlying different, shorter or longer, integration windows ( Poeppel, 2003 ). Therefore, one might speculate that the prevalence of neural organization in the left PT tuned to slower acoustic cues – otherwise more typical of the right hemisphere homologue – is somewhat greater in those with thicker left PT cortex in the present study, resulting in more efficient left-hemispheric processing of tonal information, and, in turn, facilitated access to word forms with endings cued by the stem tone. Previous results reporting a positive association between cortical thickness of the right auditory areas and relative pitch task performance ( Foster & Zatorre, 2010 ) indicate that structural variation associated with the ability to analyze slower acoustic cues might be picked up by cortical thickness measurements.

Although pseudoword stems tested in the present study could not be processed in terms of familiar syllables, tones realized on such stems still pre-activated suffixes. This was reflected in an increase in RTs for invalidly cued endings relative to validly cued continuations. Moreover, tones on pseudowords were observed to generate an increased pre-activation negativity ( P3An ) in the electrical brain signal ( Soderstrom et al., 2017 ). The correlation found here between tone processing and cortical thickness of left IFGp, a region associated with morphosyntactic analysis among other functions ( Bozic et al., 2015 ; Tyler et al., 2005 ), might indicate the involvement of a morphosyntactic rule underlying the pre-activation process. Haemodynamic responses of the same participants also indicated strong functional activation of left IFGp during pseudoword processing ( Söderström et al., 2017 ). From this perspective, cortical thickness of left IFGp could be assumed to play a role in facilitated pre-activation of suffixes, through efficient application of the morphosyntactic rule specifying the relevant tone- suffix associations, subsequently speeding up the processing of an expected ending in the input. Also, strong pre-activation of the anticipated inflection might even support rapid morphological decomposition of pseudowords into stem and suffix parts, by providing reliable cues to stem boundaries in items for which no stored lexical representations exist. Such processes can be assumed to be essential for interpreting the pseudoword forms tested, since the task crucially depended on the ability to extract the number inflection from the otherwise meaningless string of segments. The subgroup with thicker cortex in left IFGp was indeed generally faster at responding to validly cued suffixes than the thinner cortex group.

As is the case with PT, leftward asymmetry of IFGp has been reported; however, its relation to functional language lateralization is far from clear ( Kehler, Crow, Foundas, Amunts, & Roberts, 2009 ), and it is difficult to speculate how thicker cortex in this area might support inflected pseudoword processing. Based on the discussion above, facilitated responses to word endings were presumably based on an abstracted connection between tone patterns and suffixes, implicating some form of higher level processing of the input. This might be related to previous results showing a positive connection between cortical thickness in association areas and performance of higher-level cognitive skills assumed to underlie general intelligence ( Karama et al., 2009 ).

No significant relation was found between cortical thickness in the left PT and tone processing on pseudowords. Nevertheless, smaller cortical surface area in the left PT was associated with greater response time increase from validly to invalidly cued suffixes on pseudowords, suggesting increased reliance on tonal cues to anticipate upcoming endings ( Supplementary material ). This was the only significant correlation obtained with surface area in the present study, which might be assumed to reflect in part the contribution of left PT to the prelexical processing of linguistically relevant tones ( Xu et al., 2006 ) since effective recognition and discrimination of tonal patterns necessarily precedes activation of the cued morphosyntactic information. This result is in line with the suggestion that a neural organization associated with a relatively smaller surface area, and in turn with greater overlap between cortical columns, might be especially adapt at decoding suprasegmental information ( Harasty et al., 2005 ). The fact that no significant relationship was observed with surface area in the real word experiment could be taken to indicate certain differences in the specific cortical features of the left PT that are associated with facilitated tone processing on real words versus pseudowords. Based on the discussion above, this difference could be tentatively related to the presence versus absence of stored memory representations for syllables or whole words, which include the stem tone.

The fact that no correlation was found between cortical thickness of left IFGp and tone processing on real words could be taken to indicate that morphosyntactic regularities and decompositional processing might play a smaller role in suffix pre-activation on familiar word stems. One possible explanation is that more than one route is available for the predictive processing of tones on inflected words. This would be in line with dual-system models distinguishing between two mechanisms for the production and comprehension of morphologically complex items ( Pinker & Ullman, 2002 ): through decompositional route into stem and affix parts based on the application of productive morphosyntactic rules, which combine morphemes into complex words, or by direct access to whole-word representations of inflected forms in the mental lexicon. The decompositional route is argued to be implemented in a fronto-striatal brain network whereas whole-word access has been associated with a largely temporal system ( Ullman, 2004 ), which is consistent with the frontal versus temporal distinction found in the present study.

Since different individuals were tested in the real word and the pseudoword experiment, one cannot exclude the possibility that the present results might in fact to some degree reflect anatomical variations between the groups. Also, these findings do not enable us to determine the source of the observed variation in cortical thickness across
participants, which might reflect genetic, environmental and/or experiential influences (e.g. Chiarello, Vazquez, Felton, & McDowell, 2016; Panizzon et al., 2009), including adaption to the demands of processing linguistic tones in the native language. Longitudinal studies involving focused training in word accent processing and tracking associated changes in cortical thickness and surface area could, therefore, further clarify the role the cortical structure of PT and IFGpo plays in linguistic tone processing.

In conclusion, the present results indicate that individual differences in the cortical structure of left PT as well as left IFGpo might be related to word tone processing in native speakers. The role of these brain regions seems to be modulated by the presence versus absence of lexical content in incoming speech, suggesting that the cortical thickness of left PT might influence processing when tones are accessed as part of stored representations for familiar speech sound sequences, potentially corresponding to whole-word forms, whereas the cortical thickness of left IFGpo might play a greater role during rule-based processing of regularly inflected items.

4. Method

4.1. Participants

As the research question of the present study focused on the relationship between the degree of predictive tone processing and cortical structure, participants who did not show a response time advantage for validly cued suffixes, indicating that they did not use tones predictively during the experimental task, were excluded from the present analysis (n = 3 in the real word experiment and n = 2 in the pseudoword experiment). Thus, data obtained with 15 native speakers of Central Swedish (11 females, mean age: 24.9 years, SD = 5.2) participating in the real word experiment, as well as data from 17 native speakers of Central Swedish (11 females, mean age: 24.9 years, SD = 4.0) participating in the pseudoword experiment were analysed.

4.2. Stimuli and procedure

There were two experimental conditions, valid and invalid tone-suffix association. In the valid condition, word accents on the stem validly cued their associated suffix, i.e. accent 1 was followed by the singular suffix and accent 2 was followed by the plural suffix. In the invalid condition, the stem tone-suffix combinations were invalid, i.e. accent 2 on the stem preceded the singular suffix and accent 1 preceded the plural suffix. The word stem was always a monosyllabic Swedish word in the real word experiment (e.g. Valid: hus[accent 1]+en, ‘hat+sg’; Invalid: hus[accent 2] + ar, ‘hat+pl’), and a meaningless, but phonotactically legal, pseudoword (pseudoword stem+suffix) in the pseudoword experiment (e.g. Valid: kvat[accent 1] + en, ‘kvat+sg’; Invalid: kvat[accent 2] + ar, ‘kvat+pl’). Target words were placed in carrier sentences with identical structure, e.g. Kurt fick [target word] till jul, ‘Kurt got [target word] for Christmas’. Stimulus material preparation involved the same recording and splicing procedures for both experiments. Sentences were presented auditorily and participants’ task was to decide, as quickly as possible, if the person mentioned in the sentence received one or many things, by pressing a button. Response times were measured from suffix onset. In the real word experiment, there were 60 different target words (30 different lexical words presented once in singular, once in plural) in each of the two experimental conditions. In the pseudoword experiment, each of the experimental conditions consisted of 80 different target words (40 lexical words presented once in singular, once in plural). For more details, see Roll et al. (2015) and Söderström et al. (2017).

4.3. MRI

T1-weighted MPRAGE MRI scans were collected on Siemens 3T Skyra (real word experiment) and Prisma (pseudoword experiment) systems (TR/TE/TI 1900/2.54/900 ms, 1 mm3 isovoxel, 256 × 256 matrix, 176 slices, flip angle 9°). Cortical reconstruction and volumetric segmentation were performed with the Freesurfer image analysis suite (Dale, Fischl, & Sereno, 1999). Masks of PT and IFGpo were taken from the probabilistic Harvard-Oxford Cortical Structure Atlas in FMRIB Software Library thresholded at 20 and 15, respectively. The value in a point in the atlas corresponds to the probability of that point being included in the anatomical structure of interest based on expert raters in studies involving cortical parcellation (Desikan et al., 2006). Mask thresholds were chosen generously to account for individual differences in patterns of cerebral gyri (see Figs. 1 and 2). Mean cortical thickness of the PT and IFGpo masks were extracted using Freesurfer tool mri_segsstats.

4.4. Response times

Correlation analyses were performed relating individual participants’ mean cortical thickness values in the left PT as well as in the left IFGpo masks to their response time (RT) differences between validly versus invalidly cued suffixes. Results were statistically evaluated using one-tailed t-tests with Bonferroni-corrected p-values for multiple comparisons. Invalid minus valid RT difference was also correlated with the right PT and right IFGpo in a separate analysis.

Statement of significance

We found a relationship between the cortical thickness of specific language-related brain areas and the performance of certain associated native language processing skills. Whole word access versus decompositional processing seems to modulate involvement of the left planum temporale and IFG pars opercularis, in line with dual-route models of morphosyntactic processing.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (grant number 2014.0139); Marcus and Amalia Wallenberg Foundation (grant number 2014.0039); and the Swedish Research Council (grant number 2011-2284). We would like to thank Jimmy Litt for his help with data analysis.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bandl.2017.12.001.

References

Study 3
Implicit acquisition of tone-suffix connections in L2 learners of Swedish

Andrea Schremm, Pelle Söderström, Merle Horne, and Mikael Roll
Lund University

Swedish native speakers (NSs) unconsciously use tones realized on word stems to predict upcoming suffixes during speech comprehension. The present response time study investigated whether relatively proficient second language (L2) learners of Swedish have acquired the underlying association between tones and suffixes without explicit instruction, internalizing a feature that is specific to their L2. Learners listened to sentences in which the tone on the verb stem either validly or invalidly cued the following present or past tense inflection. Invalidly cued suffixes led to increased decision latencies in a verb tense identification task, suggesting that learners pre-activated suffixes associated with stem tones in a manner similar to NSs. Thus, L2 learners seemed to have acquired the tone-suffix connections through implicit mechanisms. Correctly cued suffixes were associated with a smaller processing advantage in the L2 group relative to NSs performing the same task; nevertheless, results suggest a tendency for increasingly native-like tone processing with cumulative language experience. The way suffix type affected response times also indicates exposure-related effects.

Keywords: second language acquisition, implicit learning, morphology, word accents, response times

Research using the artificial grammar-learning paradigm (Reber, 1967) has repeatedly demonstrated the tendency people have to unconsciously pick up regularities in the environment, without being aware of the knowledge they internalize (e.g. Dienes, Altmann, Kwan, & Goode, 1995; Dulany, Carlson, & Dewey, 1984; Reber, 1967; Reber & Allen, 1978). Such implicit learning mechanisms have been argued to be involved in a range of real-world situations including the acquisition of one’s native language (Reber, 1993). For instance, it has been suggested
that implicit knowledge of the native sound system emerges through an inher-
et process mapping probability distributions of the phonemes of the language
(Pierrehumbert, 2001). Indeed, learning effects observed as a result of exposure
to linguistic regularities in non-native systems indicate that these mechanisms
are likely to play a role even in second language (L2) acquisition (e.g. Dell, Reed,
Adams, & Meyer, 2000; Francis, Schmidt, Carr, & Clegg, 2009; Morgan-Short,
Steinhauer, Sanz, & Ullman, 2012; Rebuschat & Williams, 2012). However, evi-
dence for implicit learning in various L2 domains comes predominantly from
laboratory studies conducted on artificial or semi-artificial languages, and it is
unclear how these observations apply to more complex input materials character-
istic of naturalistic settings.

From this perspective, it is thus interesting to consider natural L2 learning
situations and to investigate to what extent the acquisition of untaught regulari-
ties can occur in these contexts. One such untaught regularity in Swedish is of
particular interest due to its pervasiveness in the language: the presence of asso-
ciations between grammatical suffixes and specific tones (word accents) realized
on word stems (Bruce, 1977; Riad, 2014). Native speakers (NSs) are unconscious
of the associations; nevertheless, NSs have been observed to use stem tones as
predictive cues to upcoming suffixes during speech comprehension (Roll, 2015;
Roll, Horne, & Lindgren, 2010; Roll, Söderström, & Horne, 2013; Roll et al., 2015;
Söderström, Roll, & Horne, 2012). This connection between stem tone and suf-
fix is rare among the world’s languages, largely restricted to Swedish and other
Scandinavian languages with similar prosody-suffix associations (Clausen &
Kristensen, 2015), and it is not explicitly taught to L2 learners. This provides an
opportunity, therefore, for examining the implicit natural L2 acquisition of an
unfamiliar language feature.

Thus, the present study investigated whether adult L2 learners of Swedish
acquire the native-like ability to activate tone-suffix associations during speech
processing, without having received any explicit instruction or information about
the underlying L2 regularity. To examine the extent to which L2 learners use
Swedish stem tones to predict suffixes, we conducted a response time experiment
focusing on the processing of inflected verbs in which the suffix was either validly
or invalidly cued by the word accent on the stem. We also tested the correla-
tion between response times to (in)validly cued suffixes and estimated length of
exposure to the target language to see if the use of word accents increased with
cumulative language experience.
Swedish word accent processing in NSs

In Swedish, words are pronounced with one of the two lexical tonal patterns of the language referred to as Accent 1 and Accent 2. The same word stem may appear with different word accents depending on the particular suffix attached to it. For example, the verb stem läk- ‘heal-’ receives a low tone (Accent 1) if it is followed by the present tense suffix -er, and a high tone (Accent 2) if it appears with the past tense suffix -te (Bruce, 1977; Riad, 2014). In NS speech the same word accent is always associated with a given word stem-suffix combination, i.e. a verb such as läker ‘heals’, inflected with the present tense, always appears with Accent 1, with the possible exception of production errors. Even though the meaning of läker ‘heals’ pronounced with the mismatching Accent 2 stem tone would still be understood as the present tense of läka ‘to heal’, words with such invalidly cued suffixes have been rated by NSs as significantly less acceptable than corresponding pronunciations where the suffix follows its correctly associated stem tone. Declensionally incorrect suffixes, however, appear to decrease acceptability ratings to an even greater extent (Roll et al., 2010).

Evidence from behavioural and brain-imaging studies suggests that Swedish NSs rely on word accent-suffix associations to facilitate rapid speech comprehension by predicting upcoming grammatical suffixes based on the tonal pattern of the word stem. For instance, endings that are inconsistent with the preceding stem tone, such as the past tense suffix -te following Accent 1 take longer to process than correctly cued suffixes (Söderström et al., 2012). Invalidly cued suffixes have also been found to yield an electrical brain response associated with reanalysis processes (a so-called P600), suggesting a mismatch between a pre-activated continuation and the information actually encountered in the input (Roll et al., 2010, 2013). Furthermore, correlations between hemodynamic and electrophysiological brain responses have suggested that word accent recognition in temporal areas of the brain immediately leads to activations in the left inferior frontal gyrus commonly associated with morphological processing and representations (Roll et al., 2015).

Evidence for similar speech-processing mechanisms has been reported in non-Scandinavian languages as well. For example, studies investigating the processing of onset embedded words in English and Dutch (i.e. short words that constitute a phonemic sequence included in the initial part of longer words, e.g. can and candy) showed that listeners use information in the speech signal such as sequence duration and fundamental frequency to distinguish longer words from shorter ones already before a phonemic disambiguation point (Davis, Marslen-Wilson, & Gaskell, 2002) and in order to assess the probability of an imminent prosodic boundary (Salverda, Dahan, & McQueen, 2003). Furthermore, similar
to the proposed predictive function of Swedish word accents, Kemps, Ernestus, Schreuder and Baayen (2005) found that native speakers of Dutch rely on cues such as duration and intonation to facilitate early disambiguation between regularly inflected and uninflected forms of the same noun. These results are thus further indicative of mechanisms in the speech-recognition system that modulate activations of competing word forms by taking advantage of subphonemic acoustic information to cue immediately upcoming material.

**Implicit learning and L2 acquisition**

In experimental psychology, implicit learning has been typically explored using variants of the artificial grammar-learning (AG) paradigm, first introduced by Reber (1967). During the AG task, participants are exposed to symbol strings usually generated with a finite-state grammar. Even though participants are not informed about the existence of any underlying regularities in the input during the exposure phase, their performance on a subsequent grammatically judgement test has been repeatedly shown to be above chance level (e.g. Dienes et al., 1995; Dulany et al., 1984; Reber, 1967; Reber & Allen, 1978). It has been argued that the outcome of the AG learning process can in most cases be explained by assuming that participants have absorbed the statistical regularities in the input, such as the transitional probabilities between smaller chunks of symbol sequences (Perruchet, 2008). Importantly for the purpose of the present study, surface properties of spoken Swedish provide useful statistical cues to specific word accent-suffix associations, for instance in the form of probability distributions of particular tone-suffix sequences occurring in adjacent syllables in disyllabic words (e.g. $läk_{Accent1}$+er, ‘heal+s’ vs. $läk_{Accent2}$+te, ‘heal+ed’). Nevertheless, even though the AG studies are suggestive of pervasive domain-general unconscious learning mechanisms, it is not clear whether the mere presence of statistical regularities in the L2 input, which may be attended to and interpreted at various levels simultaneously due to its inherent complexity, leads to implicit learning.

Laboratory studies have reported learning effects under incidental conditions for a range of L2 features, including syntax (Francis et al., 2009; Morgan-Short et al., 2012; Rebuschat & Williams, 2012), morphosyntax (De Diego-Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Levi, 2007), form-meaning mappings (Leung & Williams, 2011), phonotactic constraints (Dell et al., 2000; Onishi, Chambers, & Fisher, 2002) and lexical stress rules (Chan & Leung, 2014), suggesting that implicit processes may successfully absorb various aspects of the L2 structure. At the same time, there are well-known examples in natural L2 acquisition when a regularity is abundant in the input, but where learning tends to be incomplete even
after extended periods of time, such as in the case of tense inflections and the plural suffix in English (Jiang, 2007; Lardiere, 1998). Factors that have been proposed to contribute to the observed difficulty of these features include their redundancy as regards interpretation of the meaning of the utterance (e.g. subject-verb agreement in the obligatory presence of the subject), as well as the potential complexity of the form-meaning mappings involved (DeKeyser, 2005; Ellis, 2006). From this perspective, Swedish tone-suffix associations might be expected to be particularly difficult to acquire. First, there is no simple one-to-one relationship between the predictive cue and its grammatical target due to the fact that word accents are associated with different sets of affixes with a range of grammatical functions. Furthermore, since Swedish tones do not have a heavy functional (distinctive) load (Riad, 2014) nor are they necessary for recovering the meaning of the associated suffix, it is uncertain if, or to what extent, L2 learners attend to word accents in the input.

Failure to notice tone variations at the word level might have important implications for the learning process, as it has been argued that the registration of a stimulus in focal attention is required for L2 acquisition to take place (Schmidt, 2001). Similarly, more recent findings in the field of domain-general learning suggest that implicit mechanisms are influenced by attentional effects (Perruchet & Pacton, 2006), and processes forming mental associations between covarying representations might not be engaged unless at least the predictive cue is attended to (Hoffmann & Sebald, 2005). Attention might also need to be directed to the specific aspect of the stimulus material that carries predictive value (Jiménez & Méndez, 1999). For instance, focusing listeners’ attention to pitch changes in an artificial speech stream was reported to prevent unconscious extraction of word segmentation regularities embedded in transitional probabilities between syllables (Toro, Sinnett, & Soto-Faraco, 2005).

L2 learners of Swedish might therefore fail to acquire the predictive association between tones and suffixes as long as they selectively attend to those aspects of L2 speech that are crucial for understanding the message and possibly to other features their attention has been specifically drawn to through explicit instruction. Extensive experience with using tonal variation to make lexical distinctions, characteristic of NSs of languages with lexical tones, might be expected to facilitate the perception of Swedish word accents. Speakers of non-tone first languages (L1s) are, however, not likely to have any predisposition to focus on tonal patterns at the word level, and these learners might require instruction or training that explicitly promotes detection of the predictive associations (Schmidt, 2001). Another possibility is that L2 comprehension becomes less demanding on attentional resources with increasing proficiency. Thus with cumulative language exposure, learners might start to attend to features of the language input they previously
ignored, including word accents in the case of speakers of non-tone L1s. This could in turn form a basis for the operation of statistical learning mechanisms, establishing increasingly stronger tone-suffix associations. Exploring the extent to which relatively proficient L2 learners with a non-tone L1 background anticipate upcoming suffixes on the basis of word accents could thus shed light on these issues, providing insights into the implicit acquisition of a novel L2 regularity that has a low functional load but a high predictive value in speech comprehension.

The present study

The present response time study investigated whether L2 learners of Swedish at an intermediate level of proficiency have acquired the predictive use of Swedish word accents without explicit instruction. Specifically, we tested whether learners made use of stem tones on verbs to anticipate upcoming tense inflections during online speech comprehension, in the same way as NSs of Swedish do (Söderström et al., 2012). Since NSs of tone versus non-tone languages have been found to differ as regards the acquisition and perception of tonal contrasts in an L2 (Hao, 2012; Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004), the learner group included only individuals with a non-tone L1 background. Short sentences consisting of a subject pronoun and an inflected verb form (e.g. Han läk+er, ‘He heal+s’) were presented auditorily. The verb had either a present tense or a past tense suffix, which was correctly cued by its associated word accent in half of the utterances and invalidly cued by the word accent in the rest of the sentences. Thus, a verb stem such as läk- ‘heal-’ pronounced with Accent 2 was either followed by the Accent 2-associated past tense suffix -te (läk₂+te, ‘heal+ed’) or by the Accent 1-associated present tense suffix -er (*läk₂+er, ‘heal+s’). The time it took for participants to decide (from the end of the stem) whether the utterance was in the present or the past tense was measured.

If L2 learners used word accents to anticipate the upcoming suffix, the meaning of tense inflections that were correctly cued by their associated word accent would be expected to be processed faster, leading to reduced response latencies for utterances with valid tone-suffix combinations. Invalidly cued suffixes would be expected to increase response times. Alternatively, if L2 learners did not take advantage of the predictive value of word accents, the presence of stem tones validly versus invalidly cuing the following suffix would not be expected to affect the processing of suffix meaning. It is nevertheless possible that one of the inflection types would be processed faster due to prosody-independent factors such as greater frequency of occurrence in the L2 environment. For instance, the present tense form, being the only way in Swedish to inflectionally mark present tense
reference, might be assumed to be relatively more frequent in the input than the tested past tense form, which, in contrast, constitutes only one of several options to grammatically express past tense in the language. Such frequency differences could be thought to influence response time patterns, potentially speeding up the processing of the more frequent verb form. However, having been exposed to the L2 for a relatively short time, learners might not display clear frequency-related effects, and factors such as processing complexity of present tense versus past tense verb forms could also be expected to mediate response time results. Response time patterns of L2 learners were compared to results obtained with Swedish NSs in a previous study (Söderström et al., 2012) conducted using the same experimental materials and procedures.

Method

Participants
Twenty L2 learners of Swedish (13 females) participated in the experiment. Mean age was 25.4 years, $SD = 5.32$. All learners had a non-tone L1 background (Russian, Greek, Spanish, German, Italian, Dutch, Arabic, Finnish, French, Slovakian, English and Belorussian), and none of them had studied any other Scandinavian language (Norwegian or Danish) than Swedish. At the time of the experiment, the L2 learners attended one of two Swedish language courses at Lund University aimed at the same proficiency level. Based on the entry requirements and the syllabus of the courses, learners’ proficiency level in Swedish corresponded to level B1 – B2 of the Common European Framework of Reference (intermediate to upper intermediate). The participants had not received any instruction in the predictive use of Swedish word accents and were assumed to be unaware of the existence of associations between stem tones and suffixes. This assumption was based on the fact that the investigated phenomenon is not taught at Swedish as a foreign language courses at Lund University, where several of the participants received instruction in the L2. Moreover, associations between stem tones and suffixes are highly unlikely to be part of the curriculum of Swedish language courses abroad either, as can be judged from the absence of explicit discussions of the regularity in Swedish language learning materials the authors have encountered. In addition, NSs are unaware of tone-suffix associations and are not introduced to the regularity as part of L1 grammar teaching during compulsory education in Sweden, making it unlikely that NSs would point out the phenomenon to L2 learners. Word accents in general (such as the production and perception of Accent 1 and Accent 2) do not receive much emphasis in L2 teaching, since their incorrect use does not affect word meaning in the vast majority of cases (Bannert, 2004). Also,
having received an explanation of the investigated phenomenon after the pre-
sent experiment, participants expressed that they had no conscious knowledge of
word accent-suffix associations.

All participants in the L2 group were multilingual reporting good proficiency
in at least one non-native language in addition to Swedish. Based on participants’
self-evaluation, the most proficient L2 was English, with the exception of one par-
ticipant who was an English native speaker and had good knowledge of Spanish.
Several participants indicated some knowledge of three (n = 11) or more (n = 4)
non-native languages. In addition to Swedish and English, these were mostly
Romance (French, n = 9; Spanish, n = 5) and Slavic languages (Russian, n = 2;
Polish, n = 1; Belarusian, n = 1), with two participants reporting low to intermedian proficiency in German and Chinese, respectively. Before taking the B1 – B2
language course at Lund University, all participants had received formal educa-
tion in Swedish. Ten of them had attended courses exclusively in Sweden: these
were mostly organized by Lund University, but also included municipal adult
education or extramural courses at other Swedish universities. Four participants
took Lund University courses in Sweden after having obtained some knowledge
in their home countries, through self-study and formal courses (n = 2) or univer-
sity courses (n = 2). The remaining six participants had attended courses in their
home countries (university courses: n = 5, high school: n = 1).

The NS control group originally consisted of 21 speakers of Central Swedish.
Data collected with one randomly selected NS was excluded from the analysis in
order to have an equal number of participants in each group. The remaining 20
NS participants (9 females) had a mean age of 37.1 years (SD = 14.53).

Materials
The stimulus material and the experimental procedure were the same as in
Söderström et al. (2012). Each of the four test conditions (Validly cued-present,
Invalidly cued-present, Validly cued-past and Invalidly cued-past) involved twenty
short sentences (see Table 1 for example sentences) consisting of the subject pro-
noun han ‘he’ and a verb form inflected for either present tense (in the Validly cued-
present and Invalidly cued-present conditions) or the past tense (in the Validly
cued-past and Invalidly cued-past conditions). In the conditions with correctly
cued suffixes (Validly cued-present and Validly cued-past), the suffix was preceded
by its associated word accent realized on the verb stem, i.e. the present tense suffix
-er appeared with Accent 1 and the past tense suffix -te with Accent 2. The Invalidly
cued-present and Invalidly cued-past conditions involved invalid stem tone-suffix
combinations, i.e. Accent 2 + the present tense suffix -er and Accent 1 + the past
tense suffix -te. Twenty different verb stems were used to construct the stimulus
sentences, with each verb stem appearing once in each experimental condition
Implicit acquisition of tone-suffix connections

(For a list of verbs used, see Appendix in Söderström et al., 2012). Only verbs with stem-final voiceless plosive were used to ensure the presence of an unambiguous cutting point (10 ms before the release of the plosive) during subsequent editing of the recorded stimulus material. Log word frequencies calculated on the basis of the PAROLE corpus showed no significant differences between present ($M = 2.26$) and past tense ($M = 2.24$) verb forms, $t(19) = 0.38, p = .71$.

Table 1. The four experimental conditions in terms of stem tone and suffix combinations with example sentences.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Stem tone + suffix</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validly cued-present</td>
<td>Accent 1 + present tense</td>
<td>Han läk$_1$ + er. ‘He heals.’</td>
</tr>
<tr>
<td>*Invalidly cued-present</td>
<td>Accent 2 + present tense</td>
<td>*Han läk$_2$ + er. ‘He heals.’</td>
</tr>
<tr>
<td>Validly cued-past</td>
<td>Accent 2 + past tense</td>
<td>Han läk$_3$ + te. ‘He healed.’</td>
</tr>
<tr>
<td>*Invalidly cued-past</td>
<td>Accent 1 + past tense</td>
<td>*Han läk$_4$ + te. ‘He healed.’</td>
</tr>
</tbody>
</table>

Twenty utterances with present tense verbs (e.g. *Han läker*, ‘He heals’), and twenty utterances with past tense verbs (e.g. *Han läkte*, ‘He healed’) were recorded in an anechoic chamber by a male speaker of Central Swedish. As prosodic focus was on the initial subject pronoun (*Han*, ‘he’), the verb always appeared in a non-prominent sentence position. In the original recordings, each inflected verb form was pronounced with a word accent that validly cued the suffix. Utterances were subsequently edited so as to minimize variations across corresponding items in different conditions independent of the experimental manipulation. For a more detailed description of the stimulus material, see Söderström et al. (2012).

Using the speech editing software Praat, each recorded sentence was divided into two segments: a stem part stretching from the beginning of the utterance to the release of the stem-final voiceless plosive (*Han läk-*, ‘He heal-’) and a suffix part (-er or -te). The duration of the stem-final plosive’s closure phase was adjusted so as to ensure that both a following vowel (such as in the present tense suffix -er) and a consonant (such as in the past tense suffix -te) would sound natural. Stem parts were then spliced together with suffix parts in a way that each stem as well as each suffix appeared in test items with both validly and invalidly cued suffixes (see Figure 1). In the resulting stimulus material, items with identical lexical content but with validly versus invalidly cued suffixes of the same tense shared the same suffix part (Validly cued-present and Invalidly cued-present; Validly cued-past and Invalidly cued-past), and the stem part was identical across pairs carrying the...
same word accent (Validly cued-present and Invalidly cued-past; Validly cued-past and Invalidly cued-present). Conditions involving the same stem tone on the verb (Validly cued-present and Invalidly cued-past, as well as Validly cued-past and Invalidly cued-present) had identical stem part durations. The mean difference between minimum and maximum pitch values was 12.5 semitones in Accent 1 words and 11.0 semitones in Accent 2 words.

Figure 1. Illustration of splicing method for stimuli. For a given condition, identical stem fragments were concatenated with present and past endings. The disambiguation point was at the explosion of the stem-final stop consonant. Reprinted from “Processing morphologically conditioned word accents,” by P. Söderström, M. Roll, M. Horne, 2012, The Mental Lexicon, 7(1), p. 82. Copyright 2012 by John Benjamins Publishing Company.
Procedure
Participants were seated in front of a computer screen and stimulus presentation was controlled using E-Prime software. Following 10 training trials at the beginning of the experiment, the 80 test sentences were presented auditorily via loudspeakers, distributed over 4 blocks with short breaks in between. Participants’ task was to decide as quickly as possible if the utterance was in the present tense or the past tense, by pressing 1 or 2 on the keyboard. For half of the participants, key 1 represented ‘present tense’ whereas the same key stood for ‘past tense’ for the other half of the participants. Response times were measured from the point at which items with the same stem in different conditions started to differ, i.e. at the release of the final plosive of the verb stem. After the presentation of each utterance, a fixation cross appeared in the centre of the screen. The disappearance of the cross signalled the impending presentation of the next sentence.

Data analysis
For the analysis of the response time (RT) data, repeated measures ANOVAs were performed by subjects (F1) and by items (F2), with Validity (levels: validly cued suffix, invalidly cued suffix) and Suffix (levels: present tense, past tense) as within-subjects factors and Group (NSs vs. L2 learners) as a between-subjects factor. RT values below 200 ms (0.06% of all items in the NS group and 0.25% in the L2 group) and above 3000 ms (0.06% in the NS group and 0.13% in the L2 group) were excluded from the analysis.

Results
Accuracy
Overall accuracy on the verb tense decision task was very high for both participant groups, with NSs responding correctly to 98.19% of the items and L2 learners to 99.56% of the items. An ANOVA with the within-subjects factors Validity and Suffix as well as the between-subjects factor Group resulted in a main effect of Group ($F(1, 38) = 7.99, p = .007, \eta^2_p = .174$), due to generally higher accuracy scores for the L2 group relative to the NSs. As the distribution of the accuracy data was characterized by a negative skew, a Mann-Whitney’s U test was also run comparing average accuracy score per participant across the two groups. Similarly to the ANOVA, the results indicated significant differences between the L2 and the NS groups, $U = 107, p = .006$ two-tailed (the mean ranks of the NS group and the L2 group were 15.88 and 25.13, respectively).
Response times
The ANOVA comparing response time data from the NS and the L2 learner groups found a main effect of Validity ($F(1, 38) = 61.69, p < .001, \eta^2_p = .619$; $F(2, 38) = 35.82, p < .001, \eta^2_p = .485$), a main effect of Suffix ($F(1, 38) = 4.60, p = .038, \eta^2_p = .108$; $F(2, 38) = 11.80, p = .001, \eta^2_p = .237$) and a main effect of Group present only in the items analysis ($F(1, 38) = .423, p = .520, \eta^2_p = .011$; $F(2, 38) = 6.84, p = .013, \eta^2_p = .153$), reflecting the fact that NSs had longer overall RTs ($M = 687$ ms, $SD = 176$) than L2 learners ($M = 661$ ms, $SD = 143$) (Figure 2). Additionally, there was an interaction between Validity and Group in the subjects analysis ($F(1, 38) = 4.89, p = .033, \eta^2_p = .114$; $F(2, 38) = 1.90, p = .176, \eta^2_p = .048$), and a Suffix $\times$ Group interaction reaching significance in the items analysis ($F(1, 38) = 1.62, p = .211, \eta^2_p = .041$; $F(2, 38) = 4.37, p = .043, \eta^2_p = .103$). As these results suggested differences between NSs and L2 learners concerning the way item validity and suffix type affected processing, further analyses were conducted separately for the two participant groups.

NSs
The ANOVA carried out on the NS data revealed a main effect of Validity ($F(1, 19) = 49.29, p < .001, \eta^2_p = .722$; $F(2, 19) = 19.67, p < .001, \eta^2_p = .509$), indicating that invalidly cued suffixes ($M = 715$ ms, $SD = 176$) took longer to process than correctly cued suffixes ($M = 661$ ms, $SD = 177$) (Figure 2 and 3). There was also a main effect of Suffix ($F(1, 19) = 4.87, p = .040, \eta^2_p = .204$; $F(2, 19) = 13.20, p = .002, \eta^2_p = .410$), reflecting shorter response latencies to present tense ($M = 666$ ms, $SD = 190$) relative to past tense suffixes ($M = 709$ ms, $SD = 170$) (Figure 3).
Implicit acquisition of tone-suffix connections

**Figure 3.** Mean response times in the native speaker group for Validly vs. Invalidly cued suffixes in the present tense suffix (circle) and past tense suffix (square) conditions.

**L2 learners**

In a manner similar to NSs, L2 learners responded faster to validly cued suffixes ($M = 639$ ms, $SD = 138$) as compared to invalidly cued suffixes ($M = 669$ ms, $SD = 149$), resulting in a main effect of Validity ($F(1, 19) = 16.39$, $p = .001$, $\eta^2_p = .463$; $F(1, 19) = 17.05$, $p = .001$, $\eta^2_p = .473$) in the ANOVA focusing on the L2 learner group. Validity had nevertheless a smaller effect size than in the NS analysis, indicating smaller differences between validly versus invalidly cued suffixes in the L2 learner group (Figure 2 and 4). Unlike the NS group, no main effect of Suffix was obtained for the L2 group, suggesting similar RTs for present tense ($M = 649$ ms, $SD = 140$) and past tense suffixes ($M = 660$ ms, $SD = 154$) (Figure 4).

**Figure 4.** Mean response times in the L2 learner group for Validly vs. Invalidly cued suffixes in the present tense suffix (circle) and past tense suffix (square) conditions.
Correlation analyses were performed relating the number of months individual L2 learners had spent in Sweden to their average response time differences between validly versus invalidly cued suffixes and present tense verbs versus past tense verbs. Marginally significant correlations suggested a tendency for greater response time differences with increasing amount of time spent in the target language environment, between validly versus invalidly cued suffixes ($r = .358$, $p = .066$) and present tense verbs versus past tense verbs ($r = .329$, $p = .085$).

Discussion

Previous studies have found that NSs of Swedish rely on stem tones to anticipate suffixes (Roll et al., 2010, 2013, 2015; Söderström et al., 2012), but it has been unclear whether L2 learners acquire the underlying tone-suffix association through exposure to the target language, without explicit instruction or focused training. The results of the present response time experiment suggest that L2 learners at an intermediate proficiency level do in fact unconsciously use word accents as predictive cues to upcoming grammatical suffixes in a way similar to NSs, albeit less extensively. Thus, suffixes invalidly cued by the word accent on the verb stem were generally more difficult to process than correctly cued suffixes in both the NS and the L2 learner groups. It seems that listeners pre-activated the specific tense inflection associated with the word accent of the stem, which led to increased processing time when the upcoming suffix was inconsistent with the prediction.

L2 learners nevertheless displayed a relatively smaller processing advantage for target verbs with validly cued suffixes, as shown by a greater effect size for the Validity factor in the NS group. This difference might have been modulated by the amount of L2 exposure: a marginally significant correlation obtained between the time spent in the target language environment and the validly versus invalidly cued suffix difference suggests that learners might gradually become more native-like in their predictive processing of word accents with increasing L2 experience. From this perspective, the comparably smaller influence that invalidly cued inflections had on learners’ speech comprehension might reflect relatively weaker associations between tones and suffixes, probably due to less extensive exposure to the relevant covariations in the input. Weaker associations in turn could be thought to lead to somewhat reduced pre-activations presumably inhibiting alternative continuations to a smaller extent, which would explain why learners found it less demanding to process invalidly cued suffixes. Certainly, a number of other factors are also likely to have played a role accounting for the observed difference. For instance, L2 learners can be assumed to have greater difficulties in discriminating
tones, which could have occasionally interfered with the activation of rapid tone-based predictions to be confirmed or disconfirmed by the upcoming suffix. The non-focal variants of the word accents used in the test stimuli can perhaps also be expected to have been more difficult to perceive and discriminate than focal word accents had they been used instead.

For participants in the present study, the use of tone variation as a predictive cue to grammatical suffixes is a feature that is specific to the L2 language being acquired, i.e. Swedish; therefore, transfer of L1 speech comprehension strategies cannot account for the near native-like word accent processing patterns observed in the L2 group investigated. Apparently, learners implicitly acquired the underlying regularity, possibly through domain-general statistical learning mechanisms absorbing transitional probabilities between adjacent stem tones and suffixes. Even though the language feature examined is not necessary for accessing the meaning of utterances and it was therefore not obvious whether L2 learners would attend to the suffix-related word accent variation without explicit instruction, the obtained results are indicative of learning effects. From the perspective of approaches maintaining that acquisition does not take place without attention to relevant dimensions of the input (Hoffmann & Sebald, 2005; Jiménez & Méndez, 1999; Perruchet & Pacton, 2006; Schmidt, 2001), the present findings suggest that at least after extended periods of exposure to the target language, L2 learners might begin to focus on aspects of the L2 that are not attended to in the L1, including even features that lack a high functional load in the L2.

It is important to mention here some group-specific factors that could have contributed to the observed results. First, all participants in the L2 group were multilinguals, having some knowledge of at least two non-native languages. While foreign language learning experience is most likely to have facilitated the learning of Swedish in general, it would not account for the incidental acquisition of word accent-suffix associations, as this feature is not present in the other L2s of the participants. Nevertheless, assuming that attention to the predictive cue is necessary for acquisition, enough attentional resources to take in aspects of the L2 such as word accents might have become available faster for the participants compared to what would be expected with less experienced learners. Also, the most proficient L2 for the majority of the learners was English. Due to the large number of similarities between Swedish and English, both of them being Germanic languages, it is likely that the learners relied on their English skills to support the acquisition of Swedish. Such a shared language might partly explain participants’ relatively uniform response time patterns despite the variety of L1s and additional non-native languages involved. For instance, it has been shown that learners of the same L2 with different L1s might differ as regards the extent to which they use decompositional versus whole-word processing of inflected words in the L2, in ways
that suggest the influence of L1 speech comprehension strategies. Thus, speakers of inflectionally rich L1s might show a greater tendency to analyse morphologically complex words into their stem and suffix parts during speech comprehension, instead of accessing full form representations of inflected words (Portin, Lehtonen, Harrer, Wande, Niemi, & Laine, 2008). The common English language background in the present experiment could have moderated such variations in underlying morphological processing mechanisms in Swedish L2.

There is a range of further factors that could have modulated individual participants’ acquisition of the investigated feature, as well as Swedish in general. For instance, there is some evidence for the role of aptitude and individual differences in implicit learning ability (Skehan, 2015), in ways that have consequences for sequence learning (Karpicke & Pisoni, 2004) and L2 acquisition (Granena, 2013). Therefore, one might expect that some variation in the learning outcome (the degree to which learners used word accents predictively) could be attributed to pre-existing differences among participants concerning the ease with which they could infer patterns from the input. Moreover, exposure to relevant input is intuitively important for incidental acquisition, which, in addition to factors such as length of residence in the L2 country, could have been modulated by various learner-internal variables affecting the degree to which the participants actively searched for opportunities to interact with Swedish NSs. Relevant individual differences include general “willingness to communicate” (MacIntyre, Clément, Dörnyei, & Noels, 1998), attitudes to the L2 as well as learning styles and strategies (Dörnyei, 2005). For instance, learners displaying a strong integrative orientation (Gardner, 1985) might have been highly motivated to participate in conversations with NSs, and individuals with a more auditory approach to language learning could be thought to have had a relatively greater exposure to L2 speech than the more visual learners.

Interestingly, L2 learners were generally faster and more accurate in their responses than NSs, even though accuracy was close to 100% in both groups. The response time difference is likely to be related to the fact that NSs on average were significantly older than the learners. Also, learners might have been more focused on the task, which they could have regarded as a test of their L2 knowledge, expecting it to be potentially demanding to perform. An additional factor that could have contributed to the observed accuracy difference was the fact that a larger number of trials were rejected due to exceedingly quick or slow responses in the L2 group.

A further difference between the participant groups concerns the way suffix type influenced verb processing. While NSs were generally faster to respond to the present tense than to the past tense inflection, no such difference reached significance in the L2 learner group. The suffix effect found in NSs can be assumed to
be related to the way tense is expressed in Swedish morphosyntax. Like English, Swedish has several ways to grammatically mark reference to past events, such as the simple past (*preterite*) used in the current study, the past perfect (*pluperfect*) and, in certain contexts, the present perfect. However, there is only one grammatical tense for expressing present time in Swedish, the simple present, which, in contrast to English, has no progressive form. NSs would therefore be expected to process the suffix that exclusively and straightforwardly expresses reference to the present tense faster than grammatical forms expressing the more complex past time. Furthermore, these differences are likely to be reflected in the frequency distribution of the various inflections in the L2 input, further contributing to the present tense advantage observed in the NS group.

As for the L2 learners, their more limited exposure to the target language might explain the lack of clear native-like frequency-related effects. In addition, the past tense (*preterite*) tested in the experiment seems to be typically the first type of past tense form taught in language courses. It can also be assumed to be the easiest to process, since the tense feature is simply expressed on the main verb itself, without involving any auxiliary verbs (Pienemann & Håkansson, 1999). Therefore, learners might initially have a tendency to treat the simple past as the default and dominant form for expressing references to the past, very much like in the case of the present tense. As their tense system in the L2 gradually develops, however, their speech processing would also be expected to begin to reflect the complexity difference across the grammatical tenses. Consistent with this assumption is the marginally significant correlation obtained between learners’ estimated exposure time to the L2 and the suffix type difference, such that longer L2 exposure in the target language country was associated with greater response time increase for the past tense inflection. The absence of stronger correlations in analyses that related estimated L2 experience to word accent processing might be due to the use of length of stay in Sweden as a variable, which provided only a rough estimate of exposure to the L2 since some learners had received instruction in Swedish in their home countries as well.

In conclusion, results of the present study suggest that relatively proficient L2 learners of Swedish process word accents in a manner similar to NSs, by unconsciously anticipating upcoming suffixes based on the tonal pattern of the word

---

1. The typical order in which the Swedish preterite is introduced to L2 learners relative to other verb forms was identified based on the inspection of a selection of course books and exercise books that are used in Swedish as a foreign language teaching in Sweden or abroad. The following books introduce preterite as the first past tense form: Levy Scherrer and Lindemalm (2007); Nyborg, Pettersson and Holm (2001); Paulsson (2006); Rehnqvist (2010); Thunberg (2008). One inspected book (Göransson & Parada, 1997) introduces all verb tense forms at the same time.
stem. Smaller processing difficulties associated with invalidly cued suffixes in the L2 group indicate that learners relied on stem tones less extensively than NSs. Nevertheless, the predictive use of word accents appeared to increase with longer exposure time to the L2 in an immersion context, and learners also showed a tendency to become more native-like in their responses to different tense inflections with increasing scores on the same measure of L2 experience. Thus, through exposure to Swedish and in the absence of explicit instruction, learners seem to have acquired a language feature specific to the L2, i.e. an internalized connection between word accents and grammatical inflections, which can be activated as a predictive device in speech processing.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (Grant Number 2014.0139), and Marcus and Amalia Wallenberg Foundation (Grant Number 2014.0039). We would like to thank Anna Smålander for collecting the data and two anonymous reviewers for their helpful comments which have led to a much improved manuscript.

References


All rights reserved


All rights reserved


All rights reserved
Implicit acquisition of tone-suffix connections


**Corresponding address**

Andrea Schremm
Department of Linguistics and Phonetics
Lund University, Box 201
221 00 Lund
Sweden
andrea.schremm@ling.lu.se
Study 4
Training predictive L2 processing with a digital game:
Prototype promotes acquisition of anticipatory use of tone-suffix associations

Andrea Schremm*, Anna Hed, Merle Horne, Mikael Roll
Department of Linguistics and Phonetics, Lund University, Box 201, 22100 Lund, Sweden

Abstract
The present article introduces the concept of an educational game application aimed at providing training in predictive second language (L2) processing. The prototype of the game, focusing on Swedish tone-suffix associations, was tested during a two-week-period, with L2 learners whose native language lacked the targeted anticipatory linguistic cue. Results indicated that the game successfully promoted the learning of a novel L2 predictive strategy, as reflected in a general increase in accuracy throughout the test period and a gradually faster performance of the predictive task. More time spent on the highest level of the game was associated with greater accuracy gains. Furthermore, results suggest that perceptual training provided by the prototype even leads to improved production of the tonal cue. Implementation of the presented game concept in the form of a platform game is also discussed.

1. Introduction
Active and continuous predictive processing have been suggested to constitute a core function of the human brain (Bar, 2009; Clark, 2013), which has been linked to learning in general and language acquisition in particular (e.g. Dell & Chang, 2014; Phillips & Ehrenhofer, 2015). There is indeed considerable empirical evidence that native speakers (NSs) are able to make use of a wide range of cues to generate expectations about upcoming information at several different levels of representation during language processing (e.g. Altmann & Kamide, 1999; DeLong, Urbach, & Kutas, 2005; Dikker, Rabagliati, Farmer, & Pyllkänen, 2010; Kamide, Altmann, & Haywood, 2003; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004; see Kuperberg & Jaeger, 2016 for a review), and even second language (L2) learners in general seem to be capable of making linguistic predictions while processing their L2 (Kaan, 2014). Intuitively, predictive processing skills could benefit L2 learners in a variety of ways, being powerful mechanisms assumed to aid in the comprehension of noisy and ambiguous input (Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2007), rapid speech processing (Lau, Holcomb, & Kuperberg, 2013), as well as potentially facilitating L2 acquisition itself due to their proposed essential role in learning. Nevertheless, learners appear to be more restricted in their anticipatory processing in the L2 relative to NSs, especially if they are at lower proficiency levels, the predictive cue is absent in the native language (L1), or the predictive...
information is more complex (Dussias, Valdes Kroff, Guzzardo Tamargo, & Gerfen, 2013; Mitsugi & MacWhinney, 2015). Focused training in specific L2 predictive processing skills might therefore be advantageous for a wide range of L2 learners and it could contribute to more native-like processing of the L2. The present paper introduces a game application currently under development, which aims at providing such training for L2 learners. It also reports on testing the prototype of the game.

The game mechanics have been designed to facilitate learning of specific predictive language processing strategies, by requiring the player to form expectations about upcoming speech input based on a cue and providing immediate feedback on the accuracy of this prediction in the form of actually experienced input. The key features of the game mechanics were developed by drawing on results from previous research on NS and L2 predictive language processing. In its current implementation, the game focuses on Swedish tone patterns realized on word stems, which constitute predictive cues due to their associations with specific immediately following grammatical suffixes. The game trains L2 learners to use such stem tone patterns to anticipate upcoming suffixes during online speech comprehension, i.e. to apply a predictive strategy similar to that observed in NSs of Swedish (Roll, 2015; Roll, Horne, & Lindgren, 2010; Roll, Soderstrom, & Horne, 2013; Roll et al., 2015; Soderstrom, Horne, & Roll, 2016; Soderstrom, Roll, & Horne, 2012).

In an experiment, we investigated if playing a prototype implementing the discussed game mechanics would promote acquisition of the trained L2 predictive strategy in low-proficient learners whose native language lacked the specific language feature (tone) with predictive value. Therefore, the investigation aimed at determining whether L2 learners showed improvements in carrying out the task constituting the essence of the game after a two-week-period of play, as reflected in their speed and accuracy of performance. Furthermore, the study tested whether such perceptual training would lead to gains even in L2 speech production, manifested in more native-like production of the language feature constituting the predictive cue. The experiment and its results are presented below. A more complete version of the tested prototype, implementing the game mechanics in the context of a platform game, will also be introduced.

1.1. Predictive processing

In cognitive neuroscience, there is an increasingly growing emphasis on predictive processing as an essential computational mechanism of the brain, shaping and supporting the perceptual system and even underlying action (Bubic, von Cramon & Schubotz, 2010; Clark, 2013). Within this framework, sensory processing is often seen as being realized in a hierarchically organized system, with higher, more abstract levels constantly generating predictions about the probability of activities at lower levels, based on hypotheses drawn from previous knowledge as well as the current model of the context. Subsequent matching of the prediction to the actually experienced input may result in the detection of discrepancies, constituting a prediction error. Information from such an error signal can then be used at higher levels to adjust the functioning of the system in order to minimize the discrepancy with subsequent predictions (e.g. Clark, 2013; Friston, 2005). The result is the modification of processing mechanisms and/or behavior and as such, this prediction-feedback cycle has been suggested to constitute an essential learning mechanism (e.g. Schultz & Dickinson, 2000). Ongoing predictive functioning of the brain would provide the further cognitive benefit of freeing up resources from processing what is predictable, which would enable the individual to concentrate on the unexpected, discovering new things to learn. Also, in cases where prediction of stimuli facilitates task completion, this mechanism could guide directed allocation of attentional resources (Bar, 2009).

Along these lines, several models and accounts of language processing involve prediction or expectation as a core mechanism (e.g. Chang, Dell, & Bock, 2006; Levy, 2008; Pickering & Garrod, 2007), and there is now compelling evidence that NSs anticipate, at least to some degree, upcoming information prior to encountering the actual input (Kuperberg & Jaeger, 2016). For instance, contextual visual information in combination with the semantics of verbs and nouns has been found to facilitate anticipatory selection of likely upcoming arguments (e.g. Altmann & Kamide, 1999; Kamide et al., 2003). Comprehenders were also reported to anticipate certain specific properties of expected input such as gender class of upcoming nouns, based on inflections on preceding adjectives (Van Berkum et al., 2005) or gender-marked articles (Wicha et al., 2004). Speech comprehension might even involve the generation of more concrete sensory predictions concerning, for instance, the phonological form of expected nouns (DeLong et al., 2005) and orthographic forms cued by syntactic structure of the preceding context (Dikker et al., 2010). NSs have also been observed to use prosodic cues to predict syntactic structure (Roll & Horne, 2011), upcoming argument (Weber, Grice, & Crocke, 2006), to generate pragmatic inferences (Kurumada, Brown, Bibyk, Pontillo, & Tanenhau, 2014) and to anticipate upcoming inflections (Roll, 2015; Roll et al., 2010, 2013, 2015; Soderstrom et al., 2012, 2016).

In addition to the nature of the predictive cue and anticipated information, available time also appears to be an important factor modulating anticipatory behavior in NSs (Wlotko & Federmeier, 2015). Temporal constraints seem to affect the degree to which more complex cues can be used for generating predictions (Phillips & Ehrenhofer, 2015), as well as the extent to which information even at lower levels of representation is pre-activated (cf. Kuperberg & Jaeger, 2016).

1.2. Predictive processing in an L2

Applying predictive strategies during the comprehendion of an L2 could be thought to provide various benefits for the learner. For instance, generating expectations about upcoming information and linguistic structures based on what has just been processed, and then subsequently evaluating the correctness of the prediction by matching it against the incoming input provides a way of testing one’s knowledge about the L2 and learning about various dependencies in the language (Phillips &
Ehrenhofer, 2015). The way predictive processing can have a scaffolding function during the comprehension of noisy or otherwise impoverished input might be especially advantageous for learners, who have greater difficulties with processing speech in noise in their L2 than in their L1 (Shi, 2010).

Non-native anticipatory processing has been extensively studied in languages such as Spanish by comparing L2 learners’ and NSs’ use of morphologically marked gender-agreement on determiners as anticipatory cues to the following noun that the determiner agrees with. For instance, in a looking-while-listening test, L2 learners of Spanish were found to rely on the predictive value of gender information in the article in a native-like way only at a high proficiency level. Learners with low proficiency displayed no or only limited anticipatory behavior depending on the absence vs. presence of the corresponding feature in the L1 (Dussias et al., 2013). There is, however, some indication that even highly proficient learners use gender information less consistently than NSs. Grüter, Lew-Williams, and Fernald (2012) observed that gender cues had a reduced facilitating effect on L2 learners’ processing of familiar nouns relative to the NS controls. This has been argued to be related to relatively weak associations between lexical items and gender markers, as a result of limited reliance on distributional information in typical adult L2 acquisition. Similarly, Hopp (2013) found that the degree to which advanced L2 learners of German evidenced native-like gender information processing was related to their mastery of the gender system of the language as a whole, and was presumably dependent on the presence of well-established links between nouns and their gender classes supporting online comprehension. Interestingly, the learners in Grüter et al.’s (2012) study displayed native-like predictive processing with novel words acquired during teaching trials where learning conditions simulated characteristics of native language acquisition.

Studies on syntactic processing also highlight the fact that learners might be limited in their ability to make use of complex predictive cues, even if, in principle, they seem to be capable of acquiring strategies for anticipating upcoming syntactic structures which do not have any correspondence in their L1 (Lee, Lu, & Garnsey, 2013). Thus, Mitsugi and MacWhinney (2015) reported that L2 learners of Japanese did not appear to make use of their attested knowledge of case markers in order to anticipate upcoming arguments in a native-like way. The high complexity of the predictive cue, requiring the integration of lexical and grammatical information from more than one constituent, and the short time-window for generating the prediction, the span of a single word, were suggested to have contributed to the lack of predictive processing.

I.3. Implementing training in L2 predictive processing

As discussed above, empirical evidence suggests that predictive language processing as such is available to L2 learners. At the same time, learners might approximate the degree to which NSs use anticipatory cues only at a higher proficiency level, and even then they might be more limited in their ability to exploit predictive features of the L2. In certain cases, as the findings of Grüter et al. (2012) indicate, adult learners might not even be able to make use of the full potential of cues to facilitate language comprehension without learning the relevant co-variabilities under specially designed conditions. One reason might be the absence of strong enough associations between cues and the associated upcoming information, the target, in the mental representations of learners (Grüter et al., 2012; Hopp, 2013). Therefore, training should ideally involve fostering the underlying links between the predictive cue and its target. Learning of such co-variabilities has often been associated with implicit learning mechanisms, which are assumed to underlie people’s apparent ability to unconsciously pick up statistical regularities present in the input they are exposed to (e.g. Perruchet, 2008; Reber, 1967). Simple passive exposure to cue-target dependencies might, however, not be sufficient as more recent results on domain-general implicit learning mechanisms suggest that attention to the input material might be necessary for acquisition (Perruchet & Pacton, 2006), and covariation-learning mechanisms might only be engaged once the predictive cue has been attended to (Hoffmann & Sebald, 2005). Task-relevance of the predictive cue could effectively direct attention to it, indicating that carrying out language tasks for which co-occurrence relations between cue and target are crucial could be thought to promote the emergence of associations underlying anticipatory processing. Such a task might be one that explicitly requires the learner to perform the actual predictions based on the cue. In addition, as incomplete overall knowledge of the relevant dependencies might also interfere with effective predictive processing (e.g. Hopp, 2013), exposure to varied, carefully designed material is also essential.

Furthermore, as previous research has indicated, the lack of corresponding features with predictive significance in the L1 seems to modulate the success or speed with which anticipatory mechanisms are acquired (Dussias et al., 2013). This is not surprising considering the fact that learners in these cases need to acquire a new online language processing strategy. In training, the emergence of such a new strategy could be thought to be promoted by repeated practice of the skills involved, such as recognition of the predictive cue and generation of the associated expectations.

Learners might also not have the necessary processing speed or automaticity for integrating a range of information constituting more complex predictive cues. Available time for generating useful predictions has been seen to be an important variable even in native language processing (Phillips & Ehrenhofer, 2015; Wloto & Fodermeier, 2015), and considering the presumably relatively slower processing speed of inexperienced L2 learners (e.g. McDonald, 2000; McElree, Jia, & Litvak, 2000), it would not be surprising if temporal constraints had an even more profound effect. As gradual speed-up and eventual automatization of L2 skills have been argued to be expected as a result of a great deal of practice (DeKeyser & Criado, 2012), effective training should preferably involve extensive performance of the skills underlying the predictive mechanism. One possible strategy that could be thought to be consistent with promoting the rapid generation of expectations is the introduction of time constraints on the predictive task. Making speed a central aspect of the task and rewarding rapid responses might be assumed to motivate learner efforts to make faster responses.
1.4. Digital games and training in L2 predictive processing

Digital games can be argued to provide an especially suitable medium for implementing training in predictive processing. To begin with, video gaming is pervasive in society, even among adults (Entertainment Software Association (ESA), 2016), suggesting that digital games have the potential to appeal to a large learner group. Also, the inherent ability of well-designed games to sustain motivation and engagement (Plass, Homer, & Kinzer, 2015; Whitten, 2011) could be crucial considering the large number of repetitions presumably needed to establish strong enough cue-target associations as well as to speed up and eventually automate anticipatory processing. Well-designed games have been argued to have characteristics that contribute to maintaining motivation and engagement in an educational context, including the ways the goal-structure, feedback and challenges in the game are constructed. First, the goals within the game should be clear and appropriately organized so that they allow for gradual progression (Garris, Ahlers, & Driskell, 2002). Immediate feedback is also important (e.g. Kili, 2005), in a form that is relevant to achieving the goals of the game. Well-designed games also maintain an optimal level of challenge throughout the gameplay (Malone & Lepper, 1987). Challenges that are dynamically adopted to the player’s current skill level in a way that they are perceived neither too simple nor impossible to achieve have been associated with the experience of “flow”, resulting in an intrinsically rewarding state of optimal performance during complete immersion in an activity (Csikszentmihalyi, 1990; Hamari et al., 2016). Furthermore, games constitute their own virtual world, allowing the user to take risks, experiment and potentially fail without having to face real-world consequences (Plass et al., 2015). This creates a learning environment that is very much different from the one experienced in traditional classroom settings and might be especially important for learners to gain confidence in their L2 skills.

In what follows, the outline of the game mechanics will be introduced. The game has been constructed so that it promotes the emergence of predictive processing, the core of the educational tool presented in this paper. In general terms, the game is made up of cycles of stimulus presentation, followed by response from the player and subsequent automatic feedback. First, the player is presented with a sentence context leading up to a predictive cue. In the current form of the game, the sentence context is provided auditorily, while the corresponding text is also shown to the player. If the cue is auditory in nature, such as a segmental or prosodic feature, the audio recording of the sentence is played up until after the cue, the disambiguation point. The sentence context is presented in written form as soon as the audio starts. For instance, the alternatives could be represented by different platforms, i.e. suspended surfaces on which the targets are shown as text, and the player would indicate his or her choice by selecting one of them. In the present form of the game, the choice is always binary and only one of the continuations results in an utterance that would be judged by NSs as well-formed including features involved in the cue-target relation. The player is instructed to make a choice as quickly as possible once the disambiguation point has been reached, i.e. the player has arrived at the point where the next step would be to move on to one of the possible continuations, for instance by clicking on one of the alternative choices. Decision latency is calculated from the disambiguation point up until the player’s selection of one of the displayed alternatives. As soon as the selection of a continuation has been made, feedback is given on the correctness of the choice. Feedback is partly visual, showing the correct complete sentence, with the visual representation form depending on whether the choice was correct or not; and partly auditory as the recording of the correct continuation is played up after the game, the sentence context is provided auditorily, while the corresponding text is also shown to the player. If the cue is auditory in nature, such as a segmental or prosodic feature, the audio recording of the sentence is played up until after the cue, the disambiguation point. The sentence context is presented in written form as soon as the audio starts. For instance, the alternatives could be represented by different platforms, i.e. suspended surfaces on which the targets are shown as text, and the player would indicate his or her choice by selecting one of them. In the present form of the game, the choice is always binary and only one of the continuations results in an utterance that would be judged by NSs as well-formed including features involved in the cue-target relation. The player is instructed to make a choice as quickly as possible once the disambiguation point has been reached, i.e. the player has arrived at the point where the next step would be to move on to one of the possible continuations, for instance by clicking on one of the alternative choices. Decision latency is calculated from the disambiguation point up until the player’s selection of one of the displayed alternatives. As soon as the selection of a continuation has been made, feedback is given on the correctness of the choice. Feedback is partly visual, showing the correct complete sentence, with the visual representation form depending on whether the choice was correct or not; and partly auditory as the recording of the correct continuation is played up. After selecting a continuation and receiving feedback, the player can choose to replay the complete and correct auditory recording of the last item. Players then receive scores, which are calculated by taking into account, not only the accuracy of the response, but also the speed of the decision.

1.5. The Swedish word accent-suffix association

The game mechanics are implemented in an educational software currently under development, which promotes the acquisition of predictive processing of Swedish word stem tones. In Swedish, each prosodic word is pronounced with one of the two word accents of the language, either accent 1 or accent 2. As the tonal pattern (accent 1 or accent 2) associated with the same stem may vary depending on the grammatical suffix attached to it (Bruce, 1977; Riad, 2014), Swedish word accents can function as cues to the upcoming suffix. For example, the word stem *bil ‘car* is pronounced with a low tone (accent 1) if it is followed by the singular definite article *-en* attached to the stem as in *bil:accent1-en* ‘car-sg’. The same stem, however, receives a high tone (accent 2) if the word ends in the plural suffix *-ar*, i.e. *bil:accent2-ar* ‘car-pl’. NSs have been shown to rely on stem tones to anticipate how words end (Roll, 2015; Roll et al., 2010, 2013, 2015; Söderström et al., 2012, 2016) and neuroimaging data suggests that the recognition of the stem tone is immediately followed by the pre-activation of associated suffixes (Roll et al., 2015). The fact that previous research has consistently demonstrated native speakers’ tendency to anticipate upcoming suffixes based on the tone might be taken to suggest that such a strategy effectively supports speech processing. Tones are important features of the Swedish language that increase the predictability of upcoming speech material, thereby playing a role in reducing processing effort and freeing up resources to focus on other aspects of language and communication. Indeed, in the context of non-compound nouns and verbs, word accents on stems are highly predictive, greatly restricting the possible set of upcoming suffixes. For instance, accent 1 on a word stem such as *bil ‘car* excludes the possibility of any continuation other than the singular definite suffix *-en*. Also, the word accent on the verb stem distinguishes between upcoming simple present and past tense suffixes in the conjugational class of verbs that take regular *-er/-te* endings (*rök:accent1-*er* ‘smoke-pres./rök:accent2-*te* ‘smoke-past*). Results obtained with stimulus material designed to test listeners’ ability to predict an upcoming ending relying solely on the
word stem tone support this assumption. In a study where the suffix on pseudoword stems was completely masked by a cough, and the only cue to the singular versus plural ending was the stem tone (e.g. tvuk\_accent1-COUGH), Swedish native speakers were still able to recover the correct ending (tvuk\_en) with above-chance accuracy, 87.8% for accent 1 and 72% for accent 2 (Söderström et al., 2016).

L2 learners of Swedish often make mistakes in their use of word accents (Bannert, 2004), suggesting that it is an especially difficult feature of the L2 to acquire. Considering the fact that every prosodic word is pronounced with either accent 1 or accent 2, such difficulties can significantly interfere with communication with Swedish native speakers. On the one hand, if L2 learners fail to use word accents, native speakers cannot rely on the same pervasive cues facilitating speech comprehension as they are accustomed to. On the other hand, L2 learner speech might be characterized by frequent incorrect tone-suffix combinations, leading native speakers to generate tone-based predictions that frequently turn out to be wrong, hampering efficient speech processing and, in turn, mutual comprehension. Indeed, when suffixes were preceded by invalid tones in experimental stimuli, native speakers displayed not only significantly slower response times for deciding on suffix meaning, but also a P600 in their electrophysiological brain response, reflecting increased efforts to reprocess the perceived word forms (Roll, 2015; Roll et al., 2010, 2013, 2015; Söderström et al., 2012, 2016). Native speakers also rated such words with incorrect tone-suffix combinations significantly less acceptable than words in which endings were preceded by their correctly associated tones (Roll et al., 2010). Nevertheless, despite the importance of Swedish word accents for communication in the L2, the association between tones and suffixes is a feature that is not commonly taught or practiced in Swedish L2 courses (Schremm, Söderström, Horne, & Roll, 2016).

Training focusing on Swedish tone-suffix connections provides an interesting testing ground for the proposed game mechanics. First, word accents are a pervasive property of the language, constituting therefore a great potential for facilitating speech comprehension in L2 learners. At the same time, the predictive connection between the word accent and the upcoming suffix is a feature that is largely unique to Swedish, and related Norwegian. Also, the predictive cue itself, the word accent, is a prosodic property that is absent from the L1 of most L2 learners of Swedish. As a result, learning effects due to training would indicate the acquisition of a novel L2 predictive strategy, instead of simply the transfer of L1 speech processing mechanisms. Furthermore, previous results suggest that it is in principle possible for L2 learners to acquire tone-suffix associations, but apparently only at a higher proficiency level and after extensive exposure to the L2: whereas beginner learners of Swedish showed no indication of predictive use of word accents (Gosselke Berthelsen, Horne, Shtyrov, Brännström, & Roll, submitted), learners at an upper-intermediate level displayed facilitated processing of suffixes correctly cued by the preceding stem tone (Schremm et al., 2016). Nevertheless, these more proficient learners apparently made use of the anticipatory value of word accents to a more limited extent than NSs, as reflected in a relatively smaller disruption in suffix processing caused by incorrect tonal cues in this group. Possibly, learners still had not developed associations between the mental representation of cues and targets which were as strong as those of NSs, something which could have led to weaker pre-activations of predicted material, and therefore less inhibition of alternative suffix continuations. With increasing L2 exposure, however, word accent processing in the L2 group showed signs of becoming more native-like. These results thus point to the value of providing training in Swedish tone-suffix associations, especially for L2 learners at a lower proficiency level. Finally, since perceptual training of segmental and suprasegmental features of an L2 has previously been seen to lead to improvements in the production of the relevant features (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Huensch & Tremblay, 2015; Wang, Jongman, & Sereno, 2003), playing the game application might prove to be a way to promote the emergence of more native-like productive use of word accents by L2 learners of Swedish.

In addition, L2 research has shown that certain language features such as the English present tense third person singular -s inflection on the verb are notoriously difficult to acquire despite extensive exposure to relevant L2 input (R. Ellis, 2006). It has been argued that factors such as complex form-meaning mappings as well as the redundancy of these features in interpreting meaning might contribute to this difficulty (DeKeyser, 2005; N. C. Ellis, 2006). In several respects, Swedish word accents are characterized by similar properties: the same tone is associated with a variety of upcoming inflections with different morphological functions; thus, there is no straightforward one-to-one connection between a specific tonal cue and its grammatical target. Also, it is possible to extract the meaning of suffixes independently of the preceding word accent. Therefore, successful acquisition of Swedish tone-suffix connections with the help of the proposed game mechanics would have important implications for a training strategy that might be effective for teaching other anticipatory relations across different L2s that have proved to be especially challenging. For instance, one might consider training the English present tense -s inflection as a feature reflecting a predictive association between third person singular subjects and the verb ending.

2. Current study

For the present study, a prototype of the game was constructed in order to test the proposed game mechanics. For that reason, the prototype implemented the basic game mechanics but lacked any visual context and animations characteristic of digital games. Training focused on a selected subset of Swedish word accent-suffix connections, involving only stem tone variations related to the appearance of singular versus plural endings on nouns as well as present tense versus past tense endings on verbs. Participants played the prototype 15—60 min each day, for ten days during a period of two weeks. The game involved responding to sentence items such as Kungen bygger aldrig ‘The king never builds’. The task was to listen to the first
part of the sentence *Kungen bygg*- ‘The king never build’- and then choose one of two suffix alternatives (present tense suffix -er or past tense suffix -de) as quickly as possible, based on how the player thinks the sentence will continue (see Fig. 1). For instance, if the verb stem *bygg*- ‘build’ was pronounced with accent 1, the correct continuation would be –er; due to the association between accent 1 and the present tense suffix –er. Alternatively, if accent 2 had appeared on the verb stem, the correct choice would be the past tense suffix –de, the inflection associated with accent 2. The time it took for participants to decide on an ending after the presentation of the word stem tone was measured, and the accuracy of each choice as well as the amount of time participants spent playing the game were logged. The material to be presented was organized into three successive main levels. The first two main levels were further broken down into three sublevels each, which were assumed to consist of language material of increasing difficulty (see section 3.3 for more details). Main level 3 contained a mixture of material from all previous levels. Directly before and after the gaming period, participants were recorded while reading a set of Swedish sentences. This was done in order to be able to investigate if perception training would result in more native-like production of word accents.

As a result of training with the game, the hypothesis was that participants would begin to acquire the ability to use word accents in order to predict the upcoming suffix. Several sub-hypotheses can be derived from this main hypothesis (main hypothesis 1). First, learning effects would be seen in a general increase of accuracy by the end of the gaming period, relative to an initial accuracy close to chance level (hypothesis 1.1). As participants progressed through several levels of complexity throughout the game, accuracy was not expected to display a steadily increasing tendency, since at the beginning of each new level, participants had to get experience in applying the skill under training in a somewhat different context. Nevertheless, since the same general predictive strategy was required to perform the task throughout the whole duration of the game, improvement in predictive processing was expected to show up in increased final versus initial accuracy. Second, extensive training was also expected to result in a more rapid performance of the suffix prediction, which would be reflected in a general reduction of response times at later stages of the game compared to the beginning (hypothesis 1.2). From the point of view of the acquisition of tone-suffix associations, the relevant response time measure is the speed with which learners make correct suffix choices, which is expected to generally decrease from the initial to the final levels of the game, even if, as it has been argued for the accuracy measurement, a completely monotonous decrease is not expected. Observing reduced final versus initial response times for correct responses would be consistent with the assumption that using the game leads to a speeding up in performing the predictive L2 skill, not just to a general performance improvement in mechanistically making a button choice. Third, if training with the game promotes acquisition of tone-suffix associations, more time spent with the game might be expected to result in higher accuracy and faster response times in the final measurements relative to participants’ initial performance (hypothesis 1.3).

There are also a number of game-external factors that can be thought to influence the outcome of L2 training (main hypothesis 2). One possible factor is the length of residence in the target language country (hypothesis 2.1). Those learners who have spent more time in the target language environment might possibly benefit more from using the game due to greater previous exposure to L2 input, which could provide a more solid foundation to build on when developing the trained L2 skill. Positive associations between learners’ length of stay in Sweden and response time or accuracy improvement scores in the game would suggest such a relationship. Alternatively, those learners who have spent a shorter time in Sweden might show relatively greater improvements, due to having very little initial knowledge of the trained L2 features. For such learners, the present experiment might capture the complete developmental trajectory from zero to functional knowledge of Swedish word accents and their association with suffixes, manifested as a relatively greater improvement from the beginning to the end of the training period. Secondly, formal Swedish instruction might also play a role in learners’ performance on the game, given that they have received an adequate amount of instruction relevant to the trained L2 skill (hypothesis 2.2). Failing to find a positive relationship between length of Swedish instruction and response time or accuracy improvement scores would be consistent with the assumption that the game provides a form of L2 training not commonly implemented in Swedish L2 teaching.

Perceptual training with the game might be hypothesized to influence learners’ production of the targeted L2 features (main hypothesis 3). Based on previous studies reporting more native-like speech production following training in segmental or suprasegmental features of an L2 (Bradlow et al., 1997; Huensch & Tremblay, 2015; Wang et al., 2003), participants were expected to use word accents in a more native-like way in their L2 speech, as would be reflected in increased accuracy in the production data post-training relative to pre-training (hypothesis 3.1). Previous research has often failed to find a direct relationship between learners’ improvement on perceptual and production measures, suggesting that development of perceptual skills might not directly translate into increased L2 production accuracy (Bradlow et al., 1997; Huensch & Tremblay, 2015; Wang et al., 2003). Therefore, production accuracy improvement scores might not be associated with a corresponding increase in accuracy or decrease of response speed in the game (hypothesis 3.2).
3. Method

3.1. Participants

Nineteen L2 learners of Swedish (13 females) participated in the testing of the prototype, with an average age of 23 years ($SD = 1.8$). All of them were NSs of non-tone languages (see Table 1 in the Supplementary Material for participant background information). Their self-reported level of Swedish proficiency corresponded to A1 to B2 levels of the Common European Framework of Reference. On average, they had received instruction in Swedish for 22.2 months ($SD = 24.6$), abroad and/or in Sweden, and had lived in Sweden on an average of 5.8 months ($SD = 4.7$).

3.2. The prototype

In the prototype, sentence items were shown as text in dark font color, in the middle of a white presentation area. Each item was presented auditorily as well, and the start of the visual and auditory presentations were synchronized. Sentences were visually divided into three parts as shown in Fig. 1. The first part consisted of the beginning of the sentence up until the target word that was associated with the prosodic predictive cue. Next, the two possible alternative continuations of the target word were shown on two buttons with text representing the choices. The choice was always between two suffixes, one of which was cued by the word accent on the preceding target word stem. After the two buttons, the part of the sentence that followed the target word plus suffix choice was shown as text. The buttons were only enabled when the cue had been presented in the audio playback. When the player then clicked on one of the suffix choices, the audio of the rest of the sentence was played up, with the correct continuation regardless of the player choice, and feedback was displayed. The whole sentence was shown at that point as a single stretch of text, incorporating the appropriate continuation to the target word (Fig. 2). When the choice of suffix was correct, the text color switched to green; otherwise it remained in the default, dark text color. By associating a salient visual signal only to correct choices, emphasis was on providing positive feedback. The players also had the possibility of replaying the audio of the whole sentence after responding, as many times as they wished.

Sentences were presented in game rounds, consisting of 18–20 items. At the end of a round, statistics were shown summarizing performance at the last round as well as the previous rounds of the current level (except for level 3, where all the rounds played on the same day were displayed). At this point, the player could choose to stop playing or to continue with the next round. Returning after a break from the game, players could continue where they left off.

The prototype also implemented a simple, adaptive mechanism that advanced the player to the next level as soon as performance indicated learning the task of the current level. This performance limit was defined as 80% accuracy calculated over a whole round. The minimum amount of play required at a given level was two rounds. The prototype had three main levels, with three sublevels at both main level 1 and 2 (for a description of the levels, see section 3.3). In order to make sure that all participants got the opportunity to play at each level, players who had spent too much time on a single level, in the sense that they would not have been able to complete the whole game before the end of the experiment continuing at the same pace, were automatically advanced to the next level (a next sublevel under a main level or a new main level), even if they had not reached 80% accuracy. In order to determine whether such progression was necessary, the actual time that a participant had already spent playing was evaluated against expected playtime values per main level and sublevel set by the authors. Expected playtime was three complete days for main level 1 (one day for each sublevel) and six days for main level 2 (two days for each sublevel). In this way, a guaranteed one day of practice was left for the final main level 3, considering the requirement that participants were expected to play the game for at least ten days during a two-week period.

As mentioned above, one form of feedback the player received was always immediate, following choice of a suffix continuation after the predictive cue. Furthermore, summary statistics displayed after each round of play showed the percentage of correct choices, the average speed of responding, and response speed considering only correct choices (see Fig. 1 in the Supplementary Material for an example). Points were calculated taking into account the proportion of correct responses and the speed of these. The point system was formulated in a way that very fast but apparently random choices (around chance level) were not rewarded. This summary after each round served to give players feedback on their progression towards the goals within the game (cf. Kiili, 2005). By selecting a menu option on the website of the prototype, players could also view a similarly structured summary for every round they had played, presented in a chronological order, organized according to the levels of the game. Here, they could assess their progression towards reaching high enough accuracy to move to the next level. They could also compare their current performance level, both in terms of speed and accuracy, to those at earliest stages of the game. This gave players the opportunity to formulate their own future goals and evaluate their progression towards these.

Kungen byggde/aldrig.

Fig. 2. Display of the correct suffix continuation in the game prototype after the player made a choice of suffix. After a correct response, the sentence was shown in green instead.
Instruction texts about each main level were made available for players on the game website, both in English and in the target L2, Swedish. By providing detailed explanations and examples, these texts also served as a form of scaffolding. Key concepts such as accent 1 and accent 2, as well as focused and unfocused sentence positions, conditioning the way the intonational contour of word accents are realized (cf. Section 3.3), were defined and explained. The fact that word accents are associated with different suffixes was also mentioned, and associations to the singular suffix -en and the plural suffix -ar were exemplified. To facilitate the recognition and the discrimination of the predictive tonal cue, visual illustration of the tone curves of word accents were provided along with corresponding audio examples (see Fig. 2 in the Supplementary Material for the illustrations). Players could return to the instruction text at any point while using the game.

3.2.1. Technical description of the prototype

The prototype was imbedded in a website, developed using the ASP.NET MVC framework. The website ran on a server at Lund University and participants could access the game after authentication in the browser with pre-assigned credentials. The interactive part of the game was implemented using HTML5 and JQuery, and the public SoundJS API was used to manage audio playback. The website was optimized for viewing in Firefox web browser, and participants were requested to use that specific browser. In order to decrease variations related to playing the prototype using a wide range of devices, the game was designed so that it did not load on mobile devices.

At the beginning of each game round, the client-side requested the game material from the server for the complete next round to be presented for the specific authenticated user. The server-side application compiled the appropriate game material based on the current status of the user in the game. The game material as well as player-associated data was stored in a local database on the server. The presentation of the sentence items and the collection of response data were then managed by the client-side script. At the end of each round, response data was sent to the server, where it was saved in the local database, where scores as well as summary statistics were calculated taking into account the results of the most recent rounds.

The website provided functionality for each player to view accuracy and response time statistics per round. They could also track the amount of time they had spent playing the game and play time distribution throughout the test period. Participants had access only to data concerning their own results, whereas users with administrative rights, the experiment leaders, could monitor the activity of all players.

3.3. Materials

In the game prototype, each presented item was a single sentence. The target word after which the choice alternatives were displayed was either a monosyllabic verb or noun stem. The alternative continuations were always inflectional suffixes attaching to the target word stem. The noun sentences were constructed by selecting the 20 most frequent nouns from the Stockholm-Umeå corpus (Stockholm-Umeå Corpus SUC 1.0, 1996), placing each in 5 different carrier sentences. For the verb sentences, the 45 most frequent verbs in SUC were chosen to appear in 8 different carrier sentences. Not every verb could meaningfully occur in each carrier sentence due to differences in argument and thematic structure (transitive, intransitive and reflexive verbs). Therefore, there was a certain variation as regards the number of different sentence frames a verb appeared in. A set of 100 verb sentences was constructed in this way. Within the same carrier sentence, each target word stem appeared once with an accent 1 suffix and once with an accent 2 suffix. For the noun stems, the accent 1 version was a definite singular noun form, taking the inflectional suffix -en, and the accent 2 version constituted of the indefinite plural form, with one of the suffixes -ar and -or. For verb stems, the accent 1 version was in the present tense form, as indicated by the inflection -er, and the accent 2 version was in the past tense (preterite) form, taking one of the suffixes -te, -de, or -e.

The game was divided into three main levels (1, 2 and 3), and main levels 1 and 2 were further divided into three sublevels each. Sublevels are referred to as level 1.1, level 1.2 etc., where the first number identifies the main level and the second number indicates the sublevel under the main level. The difference between sublevels of the same main level concerned the type of target word. At the first two sublevels of both main level 1 and 2 (levels 1.1 and 1.2 as well as 2.1 and 2.2), all target words were of the same type, nouns and verbs, respectively. At the third sublevel (levels 1.3 and 2.3), both nouns and verbs were presented in a mixed manner. Level 3 had no further subdivision, corresponding therefore to a single sublevel, due to the fact that sentences with both noun and verb targets were randomly mixed throughout the entire level. Originally, each of the levels 1.1, 1.2 as well as 2.1 and 2.2 contained 200 sentences (100 sentence pairs, appearing with both accent 1 and accent 2 on the target word), but due to artifacts in the recordings, one sentence pair was removed from level 1.1, and two sentence pairs were removed from level 2.1, resulting in 198 sentences for level 1.1, 200 sentences for level 1.2, 196 sentences for level 2.1 and 200 sentences for level 2.2. The mixed levels 1.3 and 2.3 consisted of all the sentences that appeared on the previous two levels. Therefore, the material for level 1.3 was the combination of the 398 (198 noun + 200 verb) sentences associated with levels 1.1 and 1.2, whereas level 2.3 contained 396 items (196 noun + 200 verb sentences) from levels 2.1 and 2.2. The total number of unique sentences was thus 794. At main level 1, the target word always appeared in focused sentence position, i.e. it was pronounced with the prosodic characteristics of the most prominent phrase, constituting the most informative element in the sentence. At main level 2, targets were invariably in unfocused sentence position. Both accent 1 and accent 2 are characterized by different acoustic patterns depending on whether they are realized on words in focused or unfocused sentence positions (Bruce, 1977) (see Fig. 2 in the Supplementary Material), and due to the less prominent intonational contours of word accents in unfocused positions, the discrimination of word accent types was thought to be initially more
challenging here than at the previous main level. For the final stage, level 3, presented sentences were drawn from the complete stimulus set (794 sentences), randomly mixing noun and verb targets as well as focused and unfocused variants.

The stimulus sentences were recorded in an anechoic chamber by a female speaker of Central Swedish. The sentences were read as responses to context questions to ensure the placement of focus on the appropriate word. Each sentence was read twice. After the recordings, phonetically trained NSs of Swedish screened the stimulus material and excluded items with artifacts. Three sentence pairs were excluded in this way (from levels 1.1 and 2.1). Using the speech editing software Praat (Boersma & Weenink, 2014), sentences were subsequently divided into two parts, cutting after the target word stem syllable, at zero crossing points.

Stimulus selection and sequencing in the prototype game was designed so as to promote participants’ exposure to a varied and more complete range of stimulus material. When a player started a new level, a stimulus-sequence was generated for the specific player and divided into rounds so that all available sentences in the given level appeared exactly once. For instance, for level 1.1 where there were 198 stimulus sentences available, a sequence of 10 rounds were generated, with 9 rounds consisting of 20 sentence items and 1 round of 18 items. Sentences were randomly sequenced, while ensuring that each round had an equal number of accent 1 and accent 2 sentences. The sequencing algorithm also ensured that the same sentence item was not presented in both an accent 1 and accent 2 version within the same round. At levels of mixed target types, sentences with verbs and nouns were distributed equally within the round, as long as there were enough not-yet-presented sentences available for both target types. If the participant interrupted playing during a level, the game continued with the next round in the pre-generated sequence following the last round completed before the interruption. For the 3rd, final level, sentences were randomly selected from the complete stimulus material before each round, ensuring 50-50% presentation of accent 1 and accent 2 sentences.

Participants played the prototype for a period of 14 days, during which time they were requested to play at least on 10 different days, for a minimum of 15 min and a maximum of 60 min each day. Response time and accuracy of response for each presented sentence item as well as playtime measures were collected and stored in a database on the website’s server, associated with specific players. Response times were measured from the offset of the auditory presentation of the predictive cue, i.e. from the end of the target word stem, until the moment the player pressed one of the buttons representing an alternative continuation. Tracking of playtime was based on logging the exact date and time of day when the player started as well as finished playing each round.

3.4. Production test

During the production test before and after the training period, participants read aloud a list of 20 sentences. Participants read the same set of sentences both pre- and post-test. Learners were instructed to read the sentences as responses to context questions provided next to each item, in order to control for the placement of sentence focus. Similar to the materials in the game, each sentence had a target word, a noun or a verb, which in its correct pronunciation was associated with either accent 1 or accent 2 depending on the suffix appearing on the stem. For example, the test sentence Anna såg bilen på TV ‘Anna saw the car on TV’ was read as a response to the context question Var såg Anna bilen? ‘Where did Anna see the car?’ The most prominent phrase in the response carrying new information elicited by the question, and hence the constituent expected to receive intonational sentence focus, was the sentence-final på TV ‘on TV’. In native speaker speech, the target word (bilen ‘car’) preceding the final, focused, constituent would therefore always have been pronounced with the word accent pattern associated with unfocused sentence position. Ten of the sentences had noun targets, with either the singular definite suffix -en (associated with accent 1 on the stem) or with the plural indefinite suffix -ar (accent 2). Ten sentences had verb targets, appearing either with the present tense suffix -er (accent 1) or the past tense suffix -te or -de (accent 2). There was one carrier sentence for nouns and one for verbs. In order to see whether knowledge gained during playing the prototype would generalize to novel stimuli, the target words were selected so that half of them appeared in the game and half of them did not.

All participants were recorded using a portable recorder (TASCAM DR-07) in the same experiment room. Subsequent analysis was conducted in Praat, during which a NS of Central Swedish, a trained phonetician, performed audiovisual evaluation of the fundamental frequency (F0) curve of the target words. These were classified into five pre-defined categories, coded as (1) focused accent 1, (2) focused accent 2, (3) unfocused accent 1, (4) unfocused accent 2 or (5) no resemblance to either Central Swedish accent. One additional class, that of South Swedish accents, was introduced during the analysis since the F0 curves of some of the target words were found to resemble accent patterns characteristic of the dialect spoken in the area where the participants lived in Sweden during the test period.

3.5. Data analysis

Initial accuracy and initial response time averages were calculated based on the first two rounds played (40 items), which was at level 1.1 for all players. Final accuracy and final response time values were calculated based on the average of the last 10 rounds (200 items), which were played on the final level by all participants (level 3). A smaller number of rounds were used for the initial as compared to the final performance measurements, due to the fact that some players progressed to level 1.2 already by the third round played. Also, a significantly larger number of rounds were played at level 3 relative to level 1.1, 167 vs. 9 rounds on average.
For all calculations, items with response times (RTs) above 5000 ms were excluded (1.28% of all items) and RTs were calculated by including only correct responses. Data were collapsed over word accents, by considering both accent 1 and accent 2 target words for all accuracy and RT analyses. The time participants spent on a specific level or playing the game in total was measured by summing the time span difference between the beginning and the end of each round, on a specific level or for the whole gaming period. All correlational tests were two-tailed. In pairwise comparisons on accuracy and RT measures per level, p-values were adjusted for multiple comparisons using Bonferroni-correction.

On the production data, a repeated measures ANOVA was carried out with the within-subjects factors Pre-post (levels: pre-test, post-test), Accent (levels: accent 1, accent 2) and Game-presence (levels: present, absent). The factor Game-presence distinguished between target words that appeared in the game material and those that did not.

4. Results

4.1. Game data

Level and accuracy showed a correlation \( r = 0.923, \ p = 0.003 \) as accuracy increased throughout the levels of the game (see Table 1 and Fig. 3). Average accuracy for the first two rounds was 59.9% and 76.0% for the last ten rounds (hypothesis 1.1). Pairwise comparisons of accuracy values at each level also indicated a significant accuracy increase from the initial two to the final ten rounds \( p = 0.002 \). A significant accuracy increase was observed even from the first mixed level, level 1.3 to the final level 2.3.

<table>
<thead>
<tr>
<th>Level</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>59.9</td>
</tr>
<tr>
<td>1.2</td>
<td>66.2</td>
</tr>
<tr>
<td>1.3</td>
<td>66.4</td>
</tr>
<tr>
<td>2.1</td>
<td>73.5</td>
</tr>
<tr>
<td>2.2</td>
<td>76.4</td>
</tr>
<tr>
<td>2.3</td>
<td>75.1</td>
</tr>
<tr>
<td>Final</td>
<td>76.0</td>
</tr>
</tbody>
</table>

Note. Initial level value: the average of the first 2 rounds played at level 1.1. Final level value: the average of the last 10 rounds played at level 3.

Fig. 3. Average response accuracy increased throughout the levels in the game.
ten rounds ($p = 0.013$): at level 1.3 some learning was already expected to take place due to training; nevertheless, the sentence material was more similar to the last level due to mixing noun and verb target word types, providing a more homogenous comparison in that respect. For results of the pairwise comparisons, see Table 2 in the Supplementary Material.

There was also a negative correlation between level and RTs of correct responses averaged for level ($r = -.840, p = 0.018$) indicating that RTs generally decreased throughout the game (see Table 2 and Fig. 4). Average RT was 1787 ms for the first two rounds of the game and 748 ms for the last ten rounds (hypothesis 1.2). Pairwise comparisons of RT values at each level showed a significant RT reduction from the initial two to the final ten rounds ($p < 0.000$). Such significant RT reduction could be observed even from each mixed level, level 1.3 ($p = 0.007$) as well as level 2.3 ($p = 0.009$), to the last ten rounds. For results of the pairwise comparisons, see Table 3 in the Supplementary Material.

On average, players spent 9.10 ($SD = 8.42$) rounds (block of 20 sentences) on level 1.1 and 166.79 ($SD = 109.77$) on the last level (level 3). Levels in between took an average of 5.56 rounds to complete (level 1.2: $M = 6.84, SD = 6.81$; level 1.3: $M = 6.63, SD = 5.83$; level 2.1: $M = 5.84, SD = 7.54$; level 2.2: $M = 4.26, SD = 6.04$; level 2.3: $M = 4.21, SD = 3.79$). There was no correlation between total playtime and initial to final RT improvement ($r = 0.081, p = 0.743$) or initial to final accuracy increase ($r = 0.240, p = 0.322$) (hypothesis 1.3). Due to the proportionally greater amount of time players spent on level 3 training the complete game material, correlational analyses were conducted considering play time on that level. Results showed that the total time in hours players spent on the final level 3 correlated with the difference between final and initial accuracy ($r = 0.573, p = 0.010$), indicating that final accuracy increase relative to the initial performance was greater the more time spent on the last level (see Fig. 5). There was no correlation between initial to final RT improvement and time spent on level 3 ($r = 0.141, p = 0.564$).

<table>
<thead>
<tr>
<th>Level</th>
<th>RT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>1787</td>
</tr>
<tr>
<td>1.2</td>
<td>1253</td>
</tr>
<tr>
<td>1.3</td>
<td>1357</td>
</tr>
<tr>
<td>2.1</td>
<td>1102</td>
</tr>
<tr>
<td>2.2</td>
<td>1002</td>
</tr>
<tr>
<td>2.3</td>
<td>1233</td>
</tr>
<tr>
<td>Final</td>
<td>748</td>
</tr>
</tbody>
</table>

Note. Initial level value: the average of the first 2 rounds played at level 1.1. Final level value: the average of the last 10 rounds played at level 3.

Table 2
Average RTs for correct responses per level.

Fig. 4. There was a negative correlation between average RTs for correct responses and levels in the game.
The time (in months) participants reported to have had spent in Sweden showed a negative correlation with the difference between initial RT minus final RT of correct responses ($r = -0.467$, $p = 0.044$). Thus, less time spent in Sweden was associated with greater relative RT decrease (hypothesis 2.1). No correlation was found between accuracy improvement and reported length of stay in Sweden ($r = -0.028$, $p = 0.908$). There were no significant correlations with the reported total length of Swedish instruction (initial RT minus final RT of correct responses, $r = 0.071$, $p = 0.772$; final minus initial accuracy; $r = 0.235$, $p = 0.333$) (hypothesis 2.2).

4.2. Production data

The accuracy with which learners produced both word accents increased significantly from the pre-test to the post-test, as indicated by a main effect of Pre-post ($F(1,18) = 11.47$, $p = 0.003$) (hypothesis 3.1). There was no interaction between the factors Pre-post, Accent and Game-presence ($F(1,18) = 0.329$, $p = 0.573$). Accent 1 was generally associated with higher production accuracy relative to accent 2, reflected in a main effect for Accent ($F(1,18) = 22.86$, $p < 0.001$). The accuracy advantage for accent 1 remained stable across the test occasions based on the lack of interactions with the factor Accent.

Before training, mean accuracy was 68.75% ($SD = 34.79$) for accent 1 and 16.12% ($SD = 27.15$) for accent 2, which increased to 73.18% ($SD = 30.13$) and 26.01% ($SD = 30.36$) respectively, by post-training. There were no interactions involving the factor Game-presence either, suggesting that training-effects transferred even to novel words not included in the game material. For production accuracy values per participant and condition, see Table 4 in the Supplementary Material. Accuracy increase in production by post-test did not correlate with initial versus final accuracy increase ($r = 0.035$, $p = 0.885$) or decrease in RTs of correct responses ($r = 0.212$, $p = 0.383$) in the game (hypothesis 3.2).

5. Discussion

In the present study, we tested the prototype of an L2 learning game which has been developed to facilitate the acquisition of predictive language processing in the L2. The present implementation of the game mechanics focused on Swedish and trained L2 learners to use tone patterns on word stems as anticipatory cues to immediately upcoming suffixes. Learners with low to intermediate proficiency in Swedish used the game for a period of two weeks, while their activity and progress were continuously logged. Results of the experiment indicate that the prototype effectively promoted the learning of the trained language feature, which was seen in a gradually more accurate and faster performance of the predictive task in the game as well as in a correlation between accuracy improvement and time spent on playing at the last level. Furthermore, perceptual training with the prototype seems to have led to increased accuracy even in the production of the L2 tonal cue.

While accuracy in predicting upcoming language material based on the tonal cue was generally close to chance level (59.9%) at the initial stage of playing the game, the proportion of correct responses on average increased to 76.0% by the final
rounds of the test period. Taken together with the observed general decrease of response times throughout the game, the obtained results indicate that playing the prototype effectively promoted the acquisition of the trained tone-suffix co-variations as well as processing the anticipatory function of tones. Furthermore, more extensive play seemed to have led to relatively greater improvements in performing the task, as longer total playtime on the last level (level 3) was associated with greater differences between individual players’ final and initial accuracy. Findings related to play time at the final level can be considered especially informative about the effectiveness of the game, due to the fact that this was the level where all participants spent a significant amount of time, after a relatively rapid progression until this point in several cases. As no correlations were obtained between play time and relative RT decrease from the initial to the final values, it seems that more extensive practice did not straightforwardly lead to a continuous speeding up of performing the suffix prediction. Possibly, significant increase in speed might not be an issue of a simple quantitative change but might take place as a result of a more qualitative reorganization as the skill becomes automatized.

The tendency of gradually decreasing response latencies with increasing level throughout the game is only broken by level 1.3 and 2.3, where RTs temporally increased (Fig. 4 and Table 2). At these levels, the word accent cue appeared either on a verb or a noun in a randomly mixed fashion, while level 1.1 and 1.2 as well as level 2.1 and 2.2 focused on a single target word class. Mixing target word types thus seems to increase task difficulty. Arguably, learners needed to acquire not only the pre-activation of language features associated with a cue, but also the selective inhibition of predictions that are not relevant in the current context. For example, following accent 1, strong pre-activation of the singular definite suffix —en attaching to nouns is beneficial as long as the context of the word accent cue is a noun. However, pre-activation of the same feature might even slow down processing if the context is instead a verb, due to the emerging competition with the correct present tense inflection —er. The need to acquire the ability to appropriately inhibit temporally non-relevant predictions could have created additional difficulty at levels 1.3 and 2.3. Nevertheless, during the relatively long time most players spent at level 3 (similarly mixing noun and verb targets), results indicate they eventually seem to have absorbed such context-based modulation of pre-activations, leading to decreased final RT values.

Appropriate performance of the task in the game was crucially dependent on correct identification of word accent patterns constituting the predictive cue. The fact that participants generally became more accurate at producing word accents by the end of the gaming period suggests that this form of predictive speech processing training leads to gains even in the production of the relevant L2 cue. Increased accuracy was observed even in the pronunciation of novel target words not included in the training material. The finding that game performance data did not show a correlation with specific post-test production improvements is in accordance with the results of several previous studies that similarly failed to observe a direct one-to-one correspondence between increased production accuracy, following training in the identification of L2 segmental or suprasegmental features, and corresponding perceptual tests (e.g. Bradlow et al., 1997; Huensch & Tremblay, 2015; Wang et al., 2002). In general, these results are in line with proposals in the field of segmental phonology suggesting that the degree to which learners form accurate perceptual L2 categories limits the extent to which they approximate NSs in the production of the same sounds (Flege, 1995). While modifications in perceptual categories might be a prerequisite for attaining more accurate production, it is not assumed that any change in perception would automatically be reflected in learners’ speech production. For instance, learners might experience difficulties in acquiring the required articulatory commands to produce the L2 sound or they might be reluctant to sound more native-like due to psychosocial reasons (Flege, 1999). Furthermore, dissociation between production and perception improvements is consistent with the assumption that different neural networks are involved in transforming acoustic input into articulatory motor representations (dorsal stream) as opposed to mapping speech signals to their meaning during comprehension (ventral stream) (Hickok & Poeppel, 2007).

A negative correlation was obtained between RT decrease from the initial to the final stage in the game and the length of residence in Sweden. In more proficient learners, exposure to Swedish L2 in the target language country was previously found to be associated with gradual implicit acquisition of predictive tone processing (Schremm et al., 2016). It is therefore possible that those participants who had been in Sweden for a longer time had some basic implicit knowledge of the predictive significance of tones already at the outset of the study. From this perspective, the obtained correlation with length of residence in the current study would suggest that the game had greater benefit for those who presumably had less initial knowledge of the tone-suffix associations. Thus, it seems that training provided by the game is especially effective for speeding up the acquisition of predictive processing at early stages of L2 acquisition. Moreover, no association was observed between game performance and the length of instruction the participants received in the L2, which indicates that the game might be able to develop skills not targeted by traditional L2 Swedish instruction.

5.1. Platform game implementation of the prototype

Results of the study indicated that the game mechanics work as intended and that the game is an effective tool for promoting the acquisition of L2 predictive processing. Therefore, the prototype is currently being developed into a full-featured engaging platform game, the first version of which is presented through screen shots in Figs. 6, 7. Implementation as a platform game was motivated by a number of considerations. First, alternative linguistic targets following a predictive cue can be appropriately represented by two or more platforms, one of which the player is required to choose in order to be able to move on. Such visual presentation in the form of platforms makes it easy for the player to quickly recognize an upcoming decision point and get an overview of the possible choices. Also, the speed and fluency with which the player moves across the platforms can be important contributors to the game experience in this genre, which can be thought to promote attempts
at performing rapid choices between the target alternatives following a cue. Furthermore, linguistic co-variations underlying the predictive strategy to be learned can be made inherently task-relevant in this context, by ensuring that correct platform choices and subsequent rewards are dependent on the correct knowledge of the relevant co-variation relations. This might be thought to support the development of strong cue-target associations, the presence of which might be essential for the emergence of effective predictive processing strategies (Grüter et al., 2012; Hopp, 2013). Furthermore, the generally straightforward nature of the game-mechanics and rules ensures that players quickly develop an understanding of the game itself and are then able to focus on the learning material. Finally, platform games offer the possibility of relatively shorter periods of casual play that still create a sense of achievement and learning: it is inherent in the genre that players can frequently experience small achievements, such as each time when they have chosen a correct platform.

6. Conclusions

In conclusion, after only a two-week period of play with the game prototype presented in this paper, participants’ performance results showed learning effects concerning predictive L2 processing. These are assumed to involve the gradual acquisition of the underlying tonal cue-suffix associations in the Swedish L2 as well as the development of faster performance of the targeted anticipatory speech processing skill. Furthermore, purely perceptual training led to more accurate production of the tonal cue. The game, therefore, seems to be an appropriate tool for developing an essential predictive feature of Swedish in L2 learners’ speech, at the same time as it enables learners to acquire anticipatory strategies that can underlie

---

Fig. 6. Screenshot from the platform game implementation. Alternative choices following the auditory predictive cue are represented by different platforms, here displaying the present tense versus past tense suffixes -er and -te after the sentence fragment Läraren rökte: ‘The teacher smoke’.

Fig. 7. Screenshot from the platform game implementation, showing the moment when the player has selected a continuation to the target word stem by having the dinosaur figure jump on the -er ‘-s’ present tense suffix platform after processing the sentence fragment Läraren rökte: ‘The teacher smoke’.
more effective processing of the target L2, promoting efficient and successful communication. As the game idea introduced simulates mechanisms assumed to be related to general predictive linguistic processing, it has a broad range of possible implementations for different types of predictive co-variations in a variety of L2s. Importantly, such L2 skills might be especially difficult to train without specially designed educational tools, highlighting the value of harnessing the motivational power of a digital game that provides learners with extensive exposure to relevant co-variations, while also creating opportunities for the actual performance of the developing L2 predictive skills.

Acknowledgements

This work was supported by Knut and Alice Wallenberg Foundation (grant number 2014:0139) and Marcus and Amalia Wallenberg Foundation (grant number 2014:0039).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.compedu.2017.07.006.

References
