Studies in Second Language Acquisition (2025), 1–28
doi:10.1017/S0272263125000026

RESEARCH ARTICLE

Challenges in inflected word processing for L2 speakers

The role of stem allomorphy

Rosa Salmela[1,2,4], Minna Lehtonen[2,3], Seppo Vainio[2] and Raymond Bertram[2]

[1]Department of Psychology, Åbo Akademi University, Finland; [2]Department of Psychology and Speech Language Pathology, University of Turku, Finland; [3]MultiLing Center for Multilingualism in Society across the Lifespan, University of Oslo, Oslo, Norway and [4]Turku Research Institute for Learning Analytics, University of Turku, Finland
Corresponding author: Rosa Salmela; Email: rnsalm@utu.fi.

(Received 10 November 2023; Revised 06 September 2024; Accepted 17 November 2024)

Abstract
Morphological knowledge refers to the ability to recognize and use morphemes correctly in syntactic contexts and word formation. This is crucial for learning a morphologically rich language like Finnish, which features both agglutinative and fusional morphology. In Finnish, agglutination occurs in forms like *aamu: aamu+lla* ('morning: in the morning'), where a suffix is transparently added. Fusional features, as seen in *ilta: illa+lla* ('evening: in the evening'), involve allomorphic stem changes that reduce transparency. We investigated the challenges posed by stem allomorphy for word recognition in isolation and in context for L2 learners and L1 speakers of Finnish. In a lexical decision task, L2 speakers had longer response times and higher error rates for semitransparent inflections, while L1 speakers showed longer response times for both transparent and semitransparent inflection types. In sentence reading, L2 speakers exhibited longer fixation times for semitransparent forms, whereas L1 speakers showed no significant effects. The results suggest that the challenges in L2 inflectional processing are more related to fusional than agglutinative features of the Finnish language.

Keywords: eye tracking; lexical decision; morphological processing; second language; stem allomorphy

Introduction

Morphological knowledge, i.e., the ability to recognize and use morphemes correctly in syntactic contexts and the formation of words, is an important aspect of effective second language acquisition (SLA) for many languages. Morphological knowledge predicts spelling and reading skills in a second language (L2) (Kieffer, & Lesaux, 2008; Wang, Cheng, & Chen, 2006; Zhao, Joshi, Dixon, & Chen, 2017). The role of morphological acquisition is emphasized when learning morphologically rich languages, where grammatical relations between words cannot be understood without the efficient use of case affixes. While there is a considerable body of research on morphological acquisition in the field of SLA, it is unclear what type of morphological challenges are encountered in L2 processing when the target language is morphologically rich.

In Finnish, grammatical relations are typically expressed through the attachment of morphemes. This process can be either purely agglutinative, e.g., *puisto: puisto+ssa* ('a park: in a park'), or it may include fusional features, when phonemes in the final syllable of a word stem change when occurring in certain phonemic environments, e.g., *vesi: vede+ssä* ('water: in water'). The altered stems can be qualified as bound stems, as they cannot appear by themselves and do not correspond to any dictionary form of a given word. As the stem alternation depends on the presence of specific phonemes in word stems, and can exhibit inconsistencies throughout morphological paradigms, it is reasonable to assume that the stem alternation system in Finnish is not immediately evident to L2 speakers. The primary focus of the current study is on the impact of stem alternation on inflectional processing in L2 speakers.

Currently, there is limited empirical research on morphological processing in L2 speakers, particularly with fusional morpheme variants. Moreover, most studies have examined the processing of morphologically complex words in isolation, while only a few studies have investigated complex word processing in a sentence context. The current study addresses these issues by investigating how L2 speakers of Finnish process morphologically complex word forms with and without stem alternation (i.e., fusional and agglutinative inflections) in comparison to monomorphemic words using both a single word and a sentence context paradigm.

Morphological acquisition in L2

Morphological acquisition can be roughly divided into the mastery of derivation, compounding, and inflection. Derivational morphology and compounding are part of lexical morphology, referring to the processes by which new words are created from existing ones through affixation (e.g., *happy* + *ness* = *happiness*) or the concatenation of two existing words (*note* + *book* = *notebook*). Inflectional morphology, the focus of this study, involves affixes that convey information about grammatical relationships between words, as in the sentence "He plays football with his friends," where the affix "-s" on the verb "play" is determined by the third–person singular subject. Inflectional morphology can also be determined by extralinguistic factors; for example, the plural "-s" in the word "friends" is determined by the number of its extralinguistic referents (Hippisley & Stump, 2016). Similarly, the Finnish suffix *-ssa* 'in' as in *puistossa* 'in the park,' determines the spatial relation between two entities. Understanding inflectional morphology is essential for learning Finnish as an L2, as Finnish is a morphologically rich language. Nouns can be inflected for number and 15 cases. Without mastering the case system, syntactic relationships between sentence elements may become obscure (Martin, 1995). For example, in the sentence "*Koira jahtaa kissa**a**,*" 'the dog chases the cat,' the object of the verb is indicated by the partitive marker *-a*; if the marker were attached to the word *koira*, 'dog,' the meaning of the sentence would change to indicate that it was the dog that was being chased by the cat ("*Koira**a** jahtaa kissa*").

Morphological acquisition has proven particularly challenging for late L2 learners (DeKeyser, 2005). It is influenced by several factors. First, the typology of the first language (L1) impacts the learning of L2 morphology. For example, speakers of the highly isolating Chinese language have more difficulty acquiring English morphology than speakers of agglutinative–fusional Turkish (Wu & Juffs, 2022). Morphological constituent frequency, perceptual salience of morphological constituents, semantic complexity, morphophonological regularity, and syntactic category also affect how quickly morphologically complex words are learned (Goldschneider & DeKeyser, 2001). Age of acquisition is a significant participant–related factor that predicts morphological ability. Early bilinguals tend to achieve near–native proficiency, while a later age of exposure is associated with slower learning and poorer mastery of morphology. In behavioral studies, this is observed through slower reaction times and a higher number of errors for complex word recognition among late bilinguals compared to early bilinguals (Brysbaert, Lagrou, & Stevens, 2017; Ransdell & Fischler, 1987). In a similar vein, neurocognitive studies have shown that late L2 speakers have lower sensitivity when identifying ungrammatical inflectional forms and derivations than early L2 speakers (Kimppa et al., 2019).

For theories in SLA regarding the representation of morphologically complex words in the bilingual mental lexicon, a central question related to this study is whether morphologically complex words are stored and processed differently in bilinguals compared to monolinguals, and if this potential difference in processing is mediated by the degree of morphophonological transparency in Finnish. The decompositional models posit that complex words undergo morphological parsing, and lexical access occurs via the words' constituent morphemes (cf. Taft & Forster, 1975, 1976). This relates to the idea of morphemes being the fundamental units of language, as also proposed in the theory of distributed morphology (Halle & Marantz, 1994; Harley & Noyer, 1999; Marantz, 2013). Holistic models, on the other hand, argue that morphologically complex words are stored in their full–form lexical representations, and lexical access takes place without morpheme–level parsing (e.g., Giraudo & Grainger, 2001). At a linguistic level, this is reflected by the construction model of morphology, which presupposes a word–based approach to the analysis of the morphological structure and proposes that lexical representations are distributed across networks rather than being based on discrete morphemes (Booij, 2010). In what follows, we will present studies on morphological processing in both L1 and L2 to shed light on the differences observed in lexical access in L1 and L2. We will first present L1 results, followed by L2 results, using L1 findings as a baseline due to their more established nature compared to L2.

Studies in morphological processing in L1 and L2

In word recognition studies on L1 Finnish, monomorphemic words have often been pitted against inflected words (e.g., Bertram, Laine, Karvinen, 1999; Hyönä, Vainio, & Laine, 2002; Laine, Vainio, & Hyönä, 1999; Niemi, Laine, & Tuominen, 1994). These single–word paradigm studies have typically shown that low–to–medium–frequency inflections elicit longer reaction times and lower accuracy rates than monomorphemic words matched on variables like word length and frequency. An exception to this pattern is observed with high–frequency inflections, which are processed equally fast as matched monomorphemic words (Lehtonen & Laine, 2003; Soveri, Lehtonen, & Laine, 2007). The processing cost for low–to–medium–frequency inflections has been taken to indicate that they are decomposed into stem and affix before lexical access. Priming studies in several languages including English and French report patterns of morphological priming that also support automatic decomposition prior to lexical access in derived and inflected words in isolation (e.g., Rastle & Davis, 2008; Meunier & Longtin, 2007). Decomposition is also a feature of models that suggest multiple pathways to morphologically complex words. These models propose that the orthographic input is mapped simultaneously onto a singular whole–word representation

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

alongside other representations, such as morphemes (Kuperman, Schreuder, Bertram, & Baayen, 2009) or embedded words, as in the Word and Affix model (Beyersmann & Grainger, 2023).

The evidence concerning the processing of morphologically complex words in late L2 speakers is inconclusive. Some studies have indicated that, unlike native speakers, L2 speakers exhibit limited or no sensitivity to morphological structure, as exposed by a lack of morphological priming effects for regularly inflected complex words (e.g., Babcock et al., 2012; Basnight-Brown, Chen, Hua, Kostić, & Feldman, 2007; Bowden, Gelfand, Sanz, & Ullman, 2010; Clahsen, Balkhair, Schutter, & Cunnings, 2013; Neubauer & Clahsen, 2009; Silva & Clahsen 2008). However, several studies have demonstrated the opposite, i.e., suggesting that L2 speakers can access complex words through decomposition (e.g., Coughlin & Tremblay, 2013; Diependaele, Duñabeitia, Morris, & Keuleers, 2011; Feldman, Kostić, Basnight-Brown, Đurđević, & Pastizzo, 2010; Foote, 2015; Gor & Cook, 2010; Lehtonen, Niska, Wande, Niemi, & Laine, 2006). The conflicting results may be related to differences across studies, for instance, the L1 background of the L2 speakers or the proficiency level of the L2 speakers; L2 speakers who share a morphologically more similar L1 background to L2 (Portin et al., 2008; Vainio, Pajunen, & Hyönä, 2014) and more proficient L2 speakers tend to exhibit more sensitivity to morphological complexity (e.g., Babcock et al., 2012; Basnight-Brown et al., 2007; Bowden et al., 2010; Kimppa et al., 2019).

Allomorphy effects in L1 and L2

Allomorphy is a linguistic phenomenon in which a single morpheme has different realizations (i.e., allomorphs) depending on surrounding phonemes in a word. This can also be described as the relationship between a theme and an exponent, where the basic form of the word is considered the theme, and its inflectional variations are the exponents (Bresnan, Asudeh, Toivonen, & Wechsler, 2015). Allomorphy is particularly prominent in fusional languages. In Finnish, about 57% of singular nouns have one or more bound stem variants, meaning that the nominative form often undergoes stem alternation and has one or more allomorphic forms within the inflectional paradigm (e.g., *vesi: vede + ssä*, 'water: in water'; *vesi: vete + en*, 'water: 'to water') (Martin, 1995).

Experimental evidence for the processing of stem allomorphy is limited in both L1 and L2. For L1, there is evidence that native speakers process regular inflections (inflections without stem alternation) similarly to irregular inflections (inflections with stem alternation). For instance, Orsolini and Marslen-Wilson (1997) found that in Italian both regular and irregular inflected forms produced a similar priming effect in L1 speakers. They suggested that as Italian is a heavily inflectional language, decomposition takes place even for forms with complex stem allomorphy. Similarly, Gor and Jackson (2013) found robust auditory priming effects for high– and low–frequency inflected verbs with different levels of regularity for Russian L1 speakers. Järvikivi & Niemi (2002a) showed that for Finnish L1 speakers bound stem allomorphs—which are pseudowords when presented without suffixation—required more time to be rejected in a visual lexical decision (VLD), a task in which participants have to decide for several letter strings whether they represent real words or not) than pseudowords with minimal orthographic manipulations but without morphological status. They also showed significant priming of stem allomorphs compared to control conditions, supporting the existence of form–based representations for allomorphs in the mental lexicon. A similar finding was reported in some earlier case studies by Niemi et al.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

(1994) and Laine et al. (1994). Two studies by Nikolaev, Lehtonen, Higby, Hyun, and Ashaie (2018) and Nikolaev et al. (2014), found faster response latencies to nominative uninflected word forms with rich stem allomorphy (e.g., *vesi* 'water') than uninflected forms with more limited stem allomorphy (e.g., *savi* 'clay') in native speakers. The authors hypothesized that this result indicates multiple parallel stem allomorph activation at the lemma level with rich stem allomorphy involving a larger neural network leading to faster responses. This hypothesis received further support in a later study by Hedlund, Wikman, Hut, and Leminen (2021), which reported stronger blood–oxygen–level–dependent activity in the frontal, subcortical, and cerebellar regions for stems with more allomorphs compared to those with fewer.

Taken together, the results suggest that allomorphy does not pose a processing problem for proficient L1 speakers, at least in morphologically rich languages like Russian, Italian, and Finnish. For L2 speakers, the results are not as clear. Gor and Jackson (2013) conducted the same auditory priming experiment with American L2 learners of Russian, stratified into three proficiency levels. Similar priming patterns to those of L1 speakers were observed for high–frequency verbs for all L2 learners, but priming for low–frequency semiregular verbs was observed only for the two higher levels of proficiency, and priming for irregular verbs was only observed for the highest level of L2 proficiency. The authors took this to suggest that also late L2 learners are sensitive to morphological structure, but that the decomposition of less productive and more intricate stem allomorphy only emerges as proficiency grows. Other studies suggest that morphophonological transparency facilitates word processing in the L2 (e.g., Basnight-Brown et al., 2007; DeKeyser, 2005; Goldschneider & DeKeyser, 2001; Hahne, Mueller, & Clahsen, 2006; Kempe & Brooks, 2008; Piccinin, Dal Maso, & Giraudo, 2018). Several other studies have reported the absence of priming effects for irregularly inflected forms in L2 speakers. This absence has been taken to indicate that these forms are stored in lexical memory as full–form representations (Babcock et al., 2012; Jacob, Fleischhauer, & Clahsen, 2013; Pinker, 1999).

As far as we know, only one study has investigated the role of allomorphy in processing morphologically complex words in Finnish L2 speakers (Vainio et al., 2014). This study investigated whether typological differences concerning morphological complexity in participants' L1 affect the processing of transparent and semitransparent inflectional forms in L2 Finnish. Russian L2 speakers of Finnish patterned with Finnish natives and showed a typical inflectional processing cost with uninflected nominative forms (e.g., *koulu*; 'school') being faster than transparent inflections (e.g., *tuoli: tuoli+a;* 'chair' + partitive ending), whereas for Chinese L2 speakers of Finnish there was no clear difference between nominative and transparent inflections. However, unlike Finnish natives, both the Chinese L2 speakers and the Russian L2 speakers tended to have longer reaction times for semitransparent inflections than transparent inflections. As the statistical power was relatively low, with 12–18 participants per language group and 18 items per condition, some of the results were only marginally significant. The current study is designed to investigate morphological transparency effects in word recognition in L2 speakers in more detail, using a larger number of items and participants than in the Vainio et al. study, while also studying whether the transparency effects hold when inflected words are presented in sentence context.

Taken together, earlier studies indicate that irregularly inflected forms do not pose significant difficulties for L1 speakers, but for L2 speakers it does, whereby the processing of these forms may be moderated by typological differences of the speakers' L1. In general, the processing of morphologically complex words in Finnish L2 speakers is relatively underexplored.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Morphological processing in single–word vs. sentence–context paradigms

It is worth noting that the majority of studies on inflected word processing have employed single–word recognition paradigms. Nonetheless, it is conceivable that the semantic and syntactic roles of inflectional suffixes may function differently when inflected forms are presented within a sentence context. In support of this notion, Hyönä et al. (2002) found in L1 Finnish a processing cost for inflected words in VLD, but this effect disappeared when the same inflected words were embedded in sentence context. They argued that the processing cost in single–word paradigms may be partly related to a lack of semantic context.

On the other hand, Mousikou and Schroeder (2019), in their study of German derivations, found evidence for early embedded stem processing in both single–word priming and fast–priming eye movement sentence context experiments. Similarly, Schmidtke, Matsuki, and Kuperman (2017) found practically no difference with respect to the impact of morphological variables in derived word processing in VLD or sentence context. On the basis of their findings, both Mousikou and Schroeder (2019) and Schmidtke et al. (2017) concluded that the same factors that underpin morphological processing in single–word reading also operate in sentence reading (Mousikou & Schroeder, 2019). However, it is worth noting that there is a syntactic difference between derivation and inflection. Inflectional affixes carry morphosemantic features imposed by the syntactic relationships between the words contained in a sentence, whereas derivational affixes mainly determine word classes (nouns, verbs, and adjectives). Derivations tend also to be more opaque than inflections. This notion is supported by the review of Bertram and Hyönä (2023) who conclude that differences between the role of morphology in complex word processing are more likely to emerge in the realm of inflectional morphology than derivational morphology, the former being more context-dependent than the latter. The target words in the current study encompass genitive, partitive, and locative cases, placing them within the domain of inflectional morphology.

The present study

In the present study, we investigated whether morphological complexity and particularly stem allomorphy affect processing words in adult Finnish L2 speakers in comparison to L1 speakers. We used two experimental paradigms to determine if the processing patterns differ with and without linguistic context: VLD (Experiment 1) and Eye Tracking (ET, Experiment 2). In VLD, the target words were presented in isolation, and in ET, they were embedded in a sentence context. The target stimulus set consisted of three conditions: monomorphemic nouns, transparent inflected nouns (no stem change), and semitransparent inflected nouns (inflections with a stem change).

Our hypotheses were as follows: Among L2 speakers, we expected a processing cost for transparent inflections in comparison to monomorphemic nouns (Basnight-Brown et al., 2007; Gor & Cook, 2010; Kempe & Brooks, 2008; Lehtonen et al., 2006; Portin et al., 2008). We also expected a larger processing cost for semitransparent inflections in comparison to transparent inflections (Vainio et al., 2014; Basnight-Brown et al., 2007). For the L1 speakers, we expected a processing cost for transparent inflections in comparison to monomorphemic nouns; however, it is also possible that there is only a minimal or no difference due to the high frequency of the target words (Bertram et al., 1999; Niemi et al., 1994; Laine et al., 1999; Soveri et al., 2007). We did not expect an L1 processing cost for semitransparent inflections in comparison to transparent inflections, in line with the studies reported above (Gor and Jackson, 2013; Järvikivi & Niemi, 2002a; Järvikivi & Niemi, 2002b; Niemi et al., 1994; Orsolini & Marslen-Wilson, 1997). In

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table 1. Characteristics of participants in Experiment 1

	L2 speakers (n = 39)	L1 speakers (n = 52)
Age	*M* = 29.63, *SD* = 7.43	*M* = 25.20, *SD* = 5.90
Gender	59% females	85% females
Native language	Various languages (*N* = 26)	Finnish
Education level[1]	87% higher secondary	98% higher secondary
Age of acquisition of Finnish	*M* = 23.4, *SD* = 8.73	*M* = 0, *SD* = 0
Self-rating in Finnish[2]	*M* = 4.6, *SD* = 1.9	*M* = 8.6, *SD* = .9
Finnish skill level[3]	A1–B1	Native
Exposure to Finnish[4]	*M* = 6.41, *SD* = 7.10	*M* = 24.2, *SD* = 5.90
Recruited from	University of Turku Language Center, Turku Adult Education Center, University of Turku e-mail lists	University of Turku

Note:
[1]Lowest completed degree.
[2]Self-rating on a scale 1–10.
[3]Recommended skill level of the course during the recruitment process.
[4]Years spent in Finland.

addition, we expected that the sentence context reduces the inflectional processing costs, at least for the L1 speakers (Bertram et al., 2000; Hyönä et al., 2002).

Experiment 1

Method

Participants. Thirty-nine L2 speakers and fifty-two L1 speakers of Finnish participated in the experiment. L1 speakers were university students who participated as a part of their study curriculum. L2 speakers were recruited from Finnish language classes in adult education centers in the Turku area, Finland. The skill level of the courses was A1–B1 in the Common European Framework of Reference for Languages (CEFR). L2 speakers received a movie ticket as a reward for participation. The group exhibited linguistic diversity with 26 different languages. The most prevalent were Russian and English, each represented by four members, while others had between one and two participants. Proficiency was assessed by self-ratings and exposure. On a scale of 1–10, the average Finnish language self-rating for L2 speakers was 4.6 (SD = 1.9), and for Finnish natives 8.6 (SD = .9). The mean exposure to the Finnish language was 6 years for the L2 speakers (SD = 7.1) and 25 years for the L1 speakers (SD = 5.9). Exposure to the Finnish language and self-ratings correlated positively for L2 speakers (r = .62, CI = .62, .65). All the subject characteristics are listed in Table 1.

Materials. There were 144 Finnish nouns used as target items. The target items were divided into three conditions: 1) monomorphemic nouns, 2) bimorphemic inflections without a stem change, and 3) bimorphemic inflections with a stem change due to consonant gradation (CG). The monomorphemic case is nominative, which is represented by a zero morpheme in Finnish (*lääkäri*, 'doctor'). The inflectional suffixes used in conditions 2 and 3 were adessive (*aamu: aamu+lla*; 'morning' + adessive ending) and genitive (*isä: isä+n*; 'father' + genitive ending).[1] All target cases are frequent in Finnish:

[1]It is worth noting that one of the challenges with Finnish is that on many occasions, one cannot easily infer the correct nominative form on the basis of the inflected form, as the inflected form may allow more than one nominative form. For this reason, some of the transparent target words may have been incorrectly perceived as semitransparent and vice versa (i.e., sivulla, allowing both the correct nominative form 'sivu', but

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

in a running text, 30.7% of nouns appear in the nominative, 17.3% in the genitive, and 5.1% in the adessive case (Nikolaev & Bermel, 2023). CG is a frequent phenomenon in the Finnish language. It involves the stem plosives /*k, p, t*/, which vary with their alternating pairs when an inflectional suffix is added to the stem. This alternation may make the stem less transparent. The plosives occur either in a strong or a weak form: the weak form typically occurs in front of a closed syllable (ha**t**un, ha**t**ul-la, ha**t**us-sa) and the strong form occurs before an open syllable (*ha**tt**u, ha**tt**u-a, ha**tt**u-na*). The alternation may be quantitative (*ha**tt**u*: *ha**t**u+ssa*; 'hat' + inessive ending) or qualitative (*ko**t**i: ko**d**issa*; 'home' + inessive ending). Both qualitative and quantitative CG are prominent features in the Finnish language; however, quantitative CG is slightly more productive than qualitative CG. Quantitative CG applies even to new loanwords with few exceptions; qualitative CG may be occasionally absent, especially in proper nouns and newer loan words. Altogether, CG affects approximately 21% of Finnish words (Karlsson, 1982). All three cases, as well as CG, are part of essential grammar rules of L2 Finnish and are typically taught during the early stages of the L2 curriculum. For reasons of simplicity, these conditions are referred to from now on as 1) monomorphemic nouns, 2) transparent inflections, and 3) semitransparent inflections. The full item list is available at https://osf.io/hwtd8/.

There were 48 target items per condition. The conditions were matched for their logarithmic lemma (i.e., base) and surface frequency, and length in characters, as these characteristics are found to be fundamental in word processing and language comprehension (Barton, Hanif, Björnström, & Hills, 2014; Chetail, 2015; Rayner, 2009; Taft, 1979; Whaley 1978). Due to the relatively low proficiency level of the L2 participants, most items were frequent in the Finnish language (all lemma frequencies > 14 words per million). There were no significant differences between the conditions in length or any of the frequency variables (all p-values > .10). All frequency information was obtained by the WordMill search program (Laine & Virtanen, 1999) utilizing an unpublished Finnish morphologically parsed Turun Sanomat newspaper corpus of 22.7 million words.[2] A summary of the lexical characteristics can be found in Table 2.

The 144 target nouns were presented among 128 pseudowords and 112 filler items. All pseudowords followed the phonotactic rules of Finnish. Forty of the pseudowords appeared in the adessive form, forty in the genitive form, and forty-eight in the nominative form. The average length of the pseudowords was 6.5 characters (SD = 1.14) and the average bigram frequency was 7.6 (SD = 2.6). The fillers included real words (derivations, compounds, and inflectional cases, n = 56), and pseudowords (n = 56) of the same constellation as the real word fillers.

Apparatus and procedure. The VLD experiment was performed in E-Prime 2.1 on a desktop PC. Stimuli appeared one by one in the middle of the computer screen. A fixation point appeared before each stimulus for 500 ms. The stimulus remained until response or until the item timed out (4,000 ms). The lexical decisions were made

also an incorrectly parsed strong grade *sipu; or illalla, allowing both the correct nominative form 'ilta', but also an incorrectly parsed nominative form *illa) by the L2 speakers. However, as this misinterpretation is likely to occur in both directions, it should not bias our current results in any meaningful way.

[2]Subsequently, the lexical frequencies were double-checked with a larger corpus, which comprises over 5 billion tokens of dependency-parsed Finnish language data crawled from the Internet (Luotolahti, Kanerva, Laippala, Pyysalo, & Ginter, 2015). This corpus was accessed by the lexical search program LASTU (Itkonen, Häikiö, Vainio, & Lehtonen, 2024). No significant differences between conditions were found when the items were checked against the larger internet-based corpus (p > .13).

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table 2. Item characteristics of the target items in Experiment 1

		Lemma Fr.	Surface Fr.	Length	Bigram Fr.
Condition	N	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Monomorphemic nouns	48	185 (259)	36 (43)	6.5 (1.2)	7.0 (3.3)
Transparent inflections	48	235 (315)	29(34)	6.5 (1.3)	8.3 (2.4)
Semitransparent inflections	48	234 (258)	30 (42)	6.6 (1.2)	6.5 (2.2)

Note: Fr. = Frequency; Lemma (i.e., base) and surface frequency are scaled to 1 million, bigram frequency to 1000.

by pressing the space button ("no") or the enter button ("yes") with left and right index fingers. Ten practice trials (five words and five pseudowords) preceded the actual experiment. Participants were tested individually in a quiet room. The stimuli were presented in two blocks. The order of the blocks was counterbalanced and the order of the items in each block was randomized. The font was black Courier New (font size 14) against a white background. Before the lexical decision experiment, participants answered several background questions on L1 background, gender, self-ratings of Finnish language skills, age, age of acquisition, and educational level. The whole experimental session took approximately 30 minutes for L1 speakers and 1 hour for L2 speakers.[3]

Statistical analyses. Data were analyzed with linear mixed effect models (LMM) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R statistical software (Version 4.21; R Core Team, 2022). A generalized linear mixed effect model was used for the VLD accuracy (1/0) and a linear mixed effect model for VLD RTs. For the global models, the interaction terms of condition (nominative, transparent inflection, semitransparent inflection) and the group were fitted as a treatment–coded fixed effect variable. A trial number was added to control possible time trends, except for VLD accuracy models, where it was left out due to a convergence issue. Log–transformed lemma frequency was added to control for the possible discrepancy of item frequencies between experiments (for details, see section *Materials* in Experiment 2). For condition, the transparent inflections were set as a baseline, as we were particularly interested in the comparison of transparent inflections to monomorphemic words and semitransparent inflections. For the language group, the baseline was L1. Subjects and items were added as random effects. They were treated as intercepts, as the models did not converge with more complex random structures. For the RT data, responses below 300 ms and incorrect responses were removed, and log transformations were made to normalize the data. After this, the exclusion of the RTs over 2.5 SD was conducted, as the comparison of nonfiltered to filtered models indicated a better R^2 value for the filtered model. Following these criteria, 4% of the data were excluded.

If the global model indicated a significant interaction between language group and condition, we ran separate models for L2 and L1. For L2 speakers, we added proficiency as a control variable, as there was some variation in proficiency among L2 participants. Proficiency was a centered composite variable based on the mean of the standardized values of the exposure and self–rating score in Finnish. To preserve the comparability between the models across experiments, nonsignificant predictors were not removed after running the models, as we aimed to understand the effects of the condition in the presence of the same control variables across measures (Babyak, 2004; Shmueli, 2010).

[3]Notice that the original procedure also included running a pilot version of the 10 min Lexize vocabulary knowledge test (Salmela et al., 2021), with the background questionnaire being a component of this test.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

For the sake of simplicity, only significant effects are reported in the text. The language group had a significant effect in all analyses (i.e., L2 speakers were slower and more error-prone than L1 speakers) and the finding is not repeated in the running text. Data and R scripts are available at https://osf.io/hwtd8/.

Degrees of freedom or *p*-values are not reported in the lmer analyses, as the exact *p*-values are difficult to determine for the *t*-statistics estimated by LMMs. However, as our data set comprises over 3,000 observations, the degrees of freedom become extremely large. Thus, for reasons of practicality, the *t*-distribution can be converged to the standard normal distribution and the statistical significance at the .05 level can be informally indicated by values of the $|t \text{ or } z| > 2.00$ (Baayen, Davidson, & Bates, 2008). We follow these guidelines in our interpretations of the statistical tests.

Results

In general, the word recognition of L2 speakers was slower and more error-prone than that of L1 speakers. In both groups, semitransparent conditions elicited more errors and longer processing times in comparison to monomorphemic nouns. The observed means of the L1 and L2 groups are listed in Table 3.

Reaction times. The global model showed that transparent inflections were processed more slowly than monomorphemic nouns. In addition, a statistically significant effect was observed for lemma frequency, indicating that response times decrease with increasing lemma frequency, and for trial number, indicating that response times decrease towards the end of the experiment. Most importantly, there was a significant interaction between language group and condition: in comparison to transparent inflections, semitransparent inflections produced a larger processing delay for the L2 group than for the L1 group (Table 4). The illustration of the results is presented in Figure 1.

The separate analyses for language groups revealed that for L2 speakers, there was an effect of proficiency (β = -.16, *SE* = .04, *CI* = -.24, -.08, *t* = -4.10), indicating that less proficient L2 speakers had longer response latencies across conditions. There was also an effect of condition: the semitransparent inflections were processed more slowly than transparent inflections (β = .08, *SE* = .03, *CI* = .03, .14, *t* = 3.11). However, there was no significant effect between monomorphemic nouns and transparent inflections. An increase in logarithmic lemma frequency was associated with faster reaction times (β = -.09, *SE* = .03, *CI* = -.14, -.04, *t* = -3.51) as was increasing trial number (β = -.00, *SE* = .00, *CI* = -.00, -.00, *t* = -2.16).

Separate analysis for L1 speakers showed that L1 speakers were faster in their responses to monomorphemic words than to transparent inflections (β = -.03, *SE* = .01, *CI* = -.06, -.01, *t* = -2.37); in addition, semitransparent inflections were slower to process than transparent inflections (β = .03, *SE* = .01, *CI* = .00, .06, *t* = 2.15). The increasing trial number was associated with shorter reaction times (β = -.00,

Table 3. Mean VLD RT (in ms) and accuracy (%) with SDs by language group and condition

		RT (ms)		Accuracy (%)	
Language Group	Condition	M	SD	M	SD
L2	Monomorphemic nouns	1017	549	.86	.35
	Transparent inflections	1035	534	.89	.32
	Semitransparent inflections	1133	574	.81	.40
L1	Monomorphemic nouns	606	159	.99	.11
	Transparent inflections	630	195	.99	.12
	Semitransparent inflections	648	187	.98	.15

Table 4. Global model for log RT in the VLD experiment

	log(RT)			
Predictors	Estimates	*std. Error*	*CI*	*t*
(Intercept)	6.63	.06	6.52, 6.75	109.41 ***
Condition [Monomorphemic]	-.04	.02	-.07, -.00	-2.11 *
Condition [Semitransparent]	.03	.02	–.00, .06	1.83
Language Group [L2]	.41	.04	.33, .48	10.05 ***
Lemma Frequency (log)	–.06	.02	–.09, –.03	–3.88 ***
Trial Number	–.00	.00	–.00, –.00	–4.17 ***
Condition [Monomorphemic] * Language Group [L2]	.00	.01	–.02, .02	.11
Condition [Semitransparent] * Language Group [L2]	.05	.01	.03, .08	4.76 ***

**p*<0.05; ***p*<0.01; ****p*<0.001.

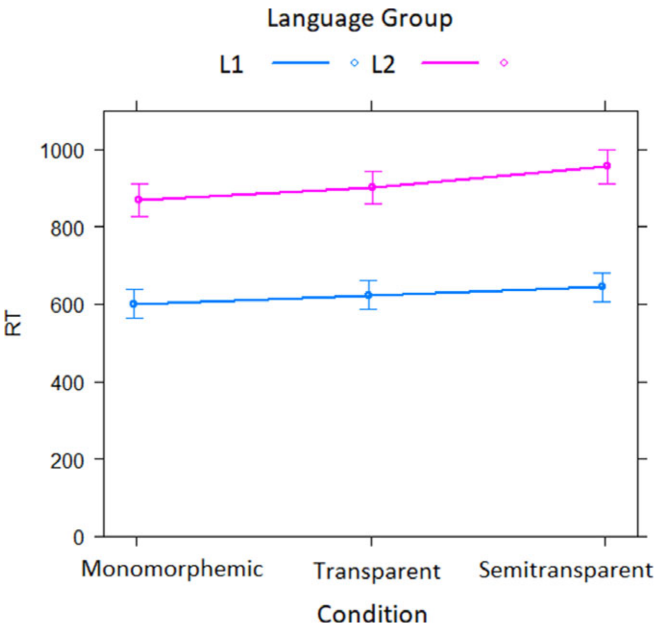

Figure 1. VLD RT (in ms) as a function of condition in L1 and L2.

SE = .00, *CI* = -.00, -.00, *t* = -4.36). The full models for L1 and L2 are presented in Appendix A (Table A1 and A2).

Accuracy. In the global model, there was a significant interaction between language group and condition: in comparison to transparent inflections, accuracy in monomorphemic nouns was lower in L2 than in L1 (Table 5). There was also a significant effect of the semitransparent condition: the semitransparency was associated with lower accuracy rates in both groups, although the L1 group was at ceiling level across conditions. In addition, a statistically significant effect was observed for logarithmic lemma frequency, indicating that more frequent words were processed more accurately in both groups.

Separate analyses per language group showed that for the L2 group, semitransparent inflections were more error-prone than transparent inflections (*OR* = .38, *95% CI* [.21,

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table 5. Global model for accuracy in the VLD Experiment.

	Accuracy			
Predictors	*Odds Ratios*	*std. Error*	*CI*	*Z*
(Intercept)	1.12	1.02	.19, 6.64	.12
Condition [Monomorphemic]	1.61	.60	.77, 3.36	1.28
Condition [Semitransparent]	.45	.15	.23, .86	–2.40 *
Language Group [L2]	.09	.03	.05, .16	–8.12 ***
Lemma Frequency (log)	4.58	1.15	2.80, 7.49	6.07 ***
Condition [Monomorphemic] * Language Group [L2]	.51	.15	.28, .92	–2.23 *
Condition [Semitransparent] * Language Group [L2]	.92	.23	.56, 1.52	–.33

**p<0.05*; ***p<0.01*; ****p<0.001*.

.70]). There was no significant difference between monomorphemic nouns and transparent inflections. There was a statistically significant effect of proficiency, indicating that the responses of L2 speakers of higher proficiency were more accurate than those of lower proficiency (*OR* = 2.43, *95% CI* [1.61, 3.67]). In addition, a statistically significant effect was found for logarithmic lemma frequency, indicating that more frequent words were responded to more accurately (*OR* = 6.31, *95% CI* [3.46, 11.48]). The detailed model is presented in Appendix A (Table A3).

For the L1 group, accuracy rates were equally high across conditions (98 to 99%) and did not show any differences across conditions. This reflects a ceiling effect, as the target words were frequent Finnish words and hence familiar to the native speakers.

Summary of Experiment 1

In both groups, monomorphemic nouns were processed the fastest. For the L2 speakers, the monomorphemic nouns were processed faster than semitransparent forms; however, there was no statistically significant difference in processing monomorphemic nouns and transparent inflections. The same results were observed in reaction times and accuracy. For the L1 speakers, the pattern was such that the monomorphemic nouns were processed fastest, followed by transparent forms, and further by semitransparent forms. There were no effects of condition in accuracy; there the L1 speakers performed at ceiling level. Together the results indicate that in isolation, semitransparent inflections are more challenging for L2 speakers than transparent forms. This may be related to the reduced salience of the allomorphic stem in the L2 mental lexicon, or to the nature of the task, i.e., the fact that the inflected words were presented without a sentence context. Experiment 2 was designed to examine if processing costs occur when the words are presented in a sentence context.

Experiment 2

Earlier studies with L1 speakers indicated that the role of morphology in complex word processing is not necessarily the same when words are processed in isolation as compared to when they are processed in context. Sentence context has been demonstrated to potentially facilitate morphological processing, particularly for inflected words (Hyönä et al., 2002). However, Mousikou and Schroeder (2019) found that context may not have an effect on morphological processing in the specific case of derivations. However, inflectional affixes, in contrast to derivational affixes, are more important in establishing syntactic relations between different parts of the sentence. This may contribute to stronger support for inflectional affixes by the sentence context. By using eye tracking to examine visual word processing during natural sentence reading, it is possible to not only tap into semantic processing but also syntactic processing. This is not the case when using a VLD, which measures access speed to single visual word representations. Moreover, in VLD, responses primarily involve discrimination processes between words and pseudowords, and answers can be based on the probability that an item is a word or not, without necessarily reaching complete activation of the lexical representation itself. Thus, in general, measuring natural reading with eye tracking can be argued to be a more ecologically valid method to study word recognition than lexical decisions. For Experiment 2, we included the majority of the words used in Experiment 1 and embedded them in sentence context. Participants were to read these sentences, while their gaze were being tracked. The goal of Experiment 2 was thus to investigate whether the processing cost identified in VLD for semitransparent forms persists in sentence context for both L2 and L1 speakers.

Method

Participants. Thirty-nine L1 speakers and thirty-seven L2 speakers participated in the experiment. Two L1 speakers were excluded due to diagnosed dyslexia and two L2 speakers due to poor performance in comprehension questions relevant to the main task (success rate close to chance level, i.e., < 60%)[4]. Following these exclusions, thirty-seven L1 speakers and thirty-five L2 speakers were included in the final analyses. None of the participants had taken part in Experiment 1. All participants had either normal uncorrected vision or corrected vision (via contact lenses or eyeglasses). The L1 speakers were university students who participated in the experiment as part of their study curriculum. The L2 speakers were recruited from Finnish language classes in adult education centers in the Turku area, Finland. The skill level of the courses was A1–B2 in the CEFR. L2 participants received a movie ticket as a reward for participation. The L2 group exhibited linguistic diversity with 18 different languages. The most prevalent was Russian, represented by 10 members, while others had between one and two participants. Characteristics of the participants of Experiment 2 are presented in Table 6. For both the L1 and L2 speakers, the age of acquisition and the level of Finnish were similar to the participants in Experiment 1.

Materials. All 144 words from the three conditions (48 monomorphemic nouns, 48 transparent inflections, and 48 semitransparent inflections) were organized in triplets and embedded in matching sentence frames. The well-formedness of the 144 sentences was assessed by eight raters, who read the sentences and evaluated them on a scale of 1–3 (1 = well-formed, 2 = somewhat well-formed, 3 = not well-formed). Sentence triplets with a clear mismatch and triplets that contained a sentence for which the mean was larger than 1.9 were excluded (n = 39). In addition, during the implementation phase, only two sentences of one triplet were included due to which we excluded the whole triplet before analyses. This led to a final set of 102 items (34 per condition). The item characteristics of Experiment 2 are presented in Table 7.

Each target word appeared in one sentence. The 34 triplets were divided over three blocks so that each of the triplet's sentences was presented in a separate block. The order of the blocks was counterbalanced. The stimulus order within one block was fixed.

[4]See section *Materials* for description of comprehension questions.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table 6. Participants of Experiment 2

	L2 speakers (*n* = 35)	L1 speakers (*n* = 37)
Age	*M* = 30.5, *SD* = 7.3	*M* = 25.0, *SD* = 6.3
Gender	69% females	86% females
Native language	Various languages (*n* = 18)	Finnish
Education level[1]	94% higher secondary	97% higher secondary
Age of acquisition of Finnish	*M* = 25.5, *SD* = 8.3	*M* = 0, *SD* = 0
Self-rating in Finnish[2]	*M* = 4.8, *SD* = 1.7	*M* = 8.8, *SD* = .8
Finnish skill level[3]	A1–B1	Native
Exposure to Finnish[4]	*M* = 5.0, *SD* = 5.1	*M* = 25.0, *SD* = 6.3
Recruited from	University of Turku Language Center, Turku Adult Education Center, University of Turku e-mail lists	University of Turku

Note:
[1]Lowest completed degree.
[2]Self-rating on a scale 1–10.
[3]Recommended skill level of the course during the recruitment process.
[4]Years spent in Finland.

Table 7. Item characteristics of Experiment 2

		Lemma Fr.	Surface Fr.	Length	Bigram Fr.
Condition	N	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Monomorphemic nouns	34	182 (230)	37 (47)	6.5 (1.1)	6.7 (2.8)
Transparent inflections	34	175 (158)	22 (24)	6.4 (1.3)	8.3 (2.8)
Semitransparent inflections	34	221 (234)	27 (35)	6.6 (1.2)	7.6 (2.5)

Note: Fr. = Frequency; Lemma and surface frequency are scaled to 1 million, bigram frequency to 1,000.

Within one triplet, sentence frames were identical up to the target word. The first word on the right side of the target was either identical or had the same initial letters. Target words appeared in a sentence position 2–5, mostly at the beginning of the sentence. The sentence position was controlled between conditions to exclude the possibility that any differences in fixation times would be related to this factor. The length, frequency, and bigram values of the target words were matched. However, due to the reduced number of items, the item frequencies did not completely match between Experiment 1 and Experiment 2. This was accounted for in the analyses by controlling the lemma frequency as a fixed effect in the models of both experiments. All sentences are listed at https://osf.io/hwtd8/. An example triplet is presented in Table 8.

To assess the predictability of the target word given the previous context, ten native speakers who did not participate in the experiment proper were presented with the text up to the target word (e.g., "Because…") and were asked to continue the sentence. None of the target words were produced by the participants in any of the sentences, so the target word was not immediately predictable on the basis of the previous context. The sentence beginnings up to the target word were semantically neutral, consisting of structures like "Because …," "Did you know that…," "I hope that…" or "They say that…."

In addition to the 102 target sentences, we included 47 filler sentences of approximately equal length and similar structure. In addition, we included 36 comprehension questions/statements related to the content of the target sentences. Answering these 'yes/no' statements required the subjects to understand the meaning of the target sentence. All sentences were presented against a light background and the text was

Table 8. An example of a sentence triplet. Target words are bolded for illustrative purposes

Condition	Example sentence
Monomorphemic noun	Koska **lääkäri** oli neuvonut minua lepäämään, päätin jäädä kotiin. 'Because the **doctor** had ordered me to rest, I decided to stay home.'
Transparent Inflection	Koska **aamulla** oli pilvistä, otin mukaani sateenvarjon. 'Because it was cloudy **in the morning**, I took an umbrella with me.'
Semitransparent inflection	Koska **illalla** oli huono ilma, en mennyt ulos. 'Because the weather was bad **in the evening**, I did not go out.'

presented in black Courier New font size 14. The length of the sentences ranged from four to twelve words. To accommodate the L2 speakers, the sentences were relatively simple syntactic constructions and contained for the most part frequent words and familiar events or descriptions.

Apparatus and procedure. Eye movement patterns were recorded monocularly with the Eyelink Portable Duo eye tracker (SR Research Ltd.). The tracker is an infrared video–based eye tracker with a sampling rate of 500 Hz. A chin rest was used to minimize head movements. The experimental sentences were presented on a laptop screen in a quiet room. Before the experiment proper, participants were instructed to read the sentences for comprehension and to signal the completion of a trial by pressing a space bar. Participants were occasionally presented with a statement that tested the comprehension of the previous sentence which they had to answer by pressing a space button ("no") or enter button ("yes"), depending on whether its content was consistent with that of the last read sentence. The size of the screen was 17 inches, the resolution was 1,920 x 1,080, and the screen type was Asus. During the experimental session, the participants were seated with their heads positioned on a chin rest 50 cm from the monitor. This led to the presentation of approximately four characters per 1-degree visual angle. For all participants, reading was binocular, but only the right eye was tracked.

Before the reading experiment, the eye tracker was calibrated by using a three–point calibration grid. The experiment started with thirteen practice trials (nine sentences and four yes/no statements). Before each trial, a fixation point appeared at the left side of the screen and the eye tracker automatically corrected the fixation for possible drifts in the original calibration.

The experiment consisted of three blocks and after each block, the participant could take a 1–3–minute break. After each break, the eye tracker was recalibrated. To control the familiarity of the target words in the L2 speakers, they were asked to fill in a questionnaire after the experiment proper. The questionnaire consisted of a list of the target words and the participants were instructed to cross the words they were not familiar with. Data from words that were reported as unknown were removed before the analyses. This yielded 18% of data loss in L2 speakers. Altogether, the experiment took approximately 30–45 minutes for L1 speakers and 45–90 minutes for L2 speakers.

Statistical analyses. Statistical analyses were identical to Experiment 1; however, the dependent variables we analyzed in this experiment were Gaze Duration (Gaze), Selective Regression Path Duration (SRPD), and Total Fixation Duration (ToFD). All measures focused on the target word only. The measures were log-transformed. Together these measures give a good insight into the time course of processing with Gaze tapping into online lexical access processes, SRPD tapping into how easily the

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

target word can be integrated within the unfolding sentence representation, and total fixation duration capturing later target word processing difficulties as well (Bertram, 2011; Liversedge, Paterson, & Pickering, 1998). Fixations shorter than 50 ms were removed from the data. This, together with the exclusion of fixation times 2.5 standard deviations above the average group mean, led to the additional exclusion of 2.9–3.2% of the data, depending on the eye–tracking measure.

Results

The results showed that for L2 speakers fixation durations were longer than for L1 speakers, and that longer fixation times were elicited by transparent and semitransparent forms than monomorphemic nouns. However, the differences between morphological conditions were relatively small for the L1 speakers. The observed fixation times for each measure per condition and language group are summarized in Table 9.

Gaze and SRPD. In the global model, the interaction between group and condition was not statistically significant for either Gaze or SRPD. Additionally, there was no effect of condition in these measures. In SRPD, there was an effect of trial number, suggesting that the selective regression path durations decreased during the course of the experiment (β = -.00, *SE* = .00, *CI* = -.00, -.00, t = -4.09). The full models are presented in Appendix A (Tables A4 and A5).

ToFD. In ToFD, there was a significant interaction between condition and language group: the difference in ToFD for semitransparent inflections in comparison to transparent inflections was larger in L2 speakers than in L1 speakers. There was also an effect of trial number, indicating that total fixation times decreased towards the end of the experiment (Table 10). Figure 2 depicts the interaction between the condition and language group.

In separate models, we found that L2 speakers had longer ToFDs for semitransparent inflections than transparent inflections (β = .13, *SE* = .05, *CI* = .02, .23, t = 2.40). However, the difference between monomorphemic nouns and transparent inflections was not statistically significant. There was also an effect of trial number, suggesting that ToFDs were shorter towards the end of the experiment (β = -.00, *SE* = .00, *CI* = .00, .00, t = -3.99). For the L1 group, none of the differences involving the condition were statistically significant. However, there was an effect of trial number (β = -.00, *SE* = .00, *CI* = .00, .00, t = -7.43). The full models for L1 and L2 are presented in Appendix A (Tables A6 and A7).

Table 9. Observed fixation times (in ms) for the dependent variables (means and SDs) in both language groups in Experiment 2

		L1		L2	
Measure	Condition	M	SD	M	SD
Gaze	Monomorphemic noun	231	91	494	379
	Transparent inflection	238	108	516	369
	Semitransparent inflection	240	101	546	409
SRPD	Monomorphemic noun	242	100	552	411
	Transparent inflection	248	117	593	419
	Semitransparent inflection	255	114	630	455
TOFD	Monomorphemic noun	333	205	958	836
	Transparent inflection	350	246	1051	891
	Semitransparent inflection	360	244	1176	945

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table 10. Global model for ToFD in the eye–tracking experiment

Predictors	log(ToFD)			
	Estimates	*std. Error*	*CI*	*t*
(Intercept)	5.78	.15	5.49–6.07	39.25 ***
Condition [Monomorphemic]	–.02	.04	–.10, .06	–.50
Condition [Semitransparent]	.04	.04	–.04, .12	1.06
Language Group [L2]	.96	.09	.79, 1.14	10.71 ***
Trial Number	–.00	.00	–.00, –.00	–8.02 ***
Lemma Frequency (log)	–.01	.04	–.09, .06	–.29
Condition [Monomorphemic] * Language Group [L2]	–.02	.03	–.07, .04	–.56
Condition [Semitransparent] * Language Group [L2]	.08	.03	.02, .14	2.69 **

*$p<0.05$; **$p<0.01$; ***$p<0.001$.

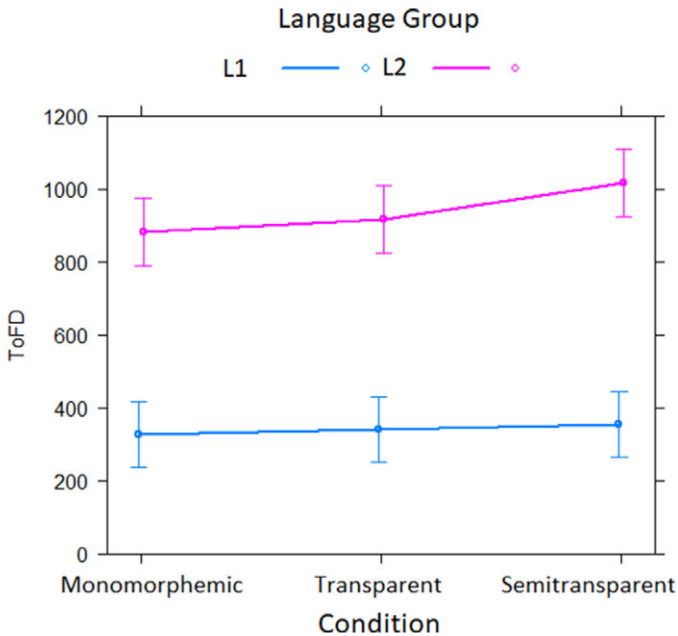

Figure 2. Total fixation duration (in ms) as a function of condition in L1 and L2.

Summary of Experiment 2

For L2 speakers, the results of Experiment 2 mirrored the results of Experiment 1: semitransparent inflections were fixated longer than transparent inflections, but there was no difference between monomorphemic nouns and transparent inflections. This effect was found for ToFD, but not in the two other measures. The results indicate that in comparison to transparent inflections, processing of semitransparent stem variants creates additional challenges for the L2 speakers, even if they are presented in sentence context. This effect was detected in ToFD, which considers cumulative fixations on a target word during the reading process. In Gaze and SRPD there was neither an effect of morphological condition nor an interaction between morphological condition and group. For the L1 speakers, there were no statistically significant differences in

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

processing times between conditions in any of the three measures. In other words, the processing cost found in VLD for transparent and semitransparent inflections in comparison to monomorphemic words was not observed when reading the words in sentences. Together the results suggest that during natural reading, L1 speakers deal more effectively with morphological complexity and morphosyntactic processing than L2 speakers.

General Discussion

The aim of this study was to investigate how morphological complexity and particularly stem allomorphy affect the processing of Finnish words in L2 speakers in comparison to native speakers. The findings reveal a difference between L1 and L2 speakers in the processing of semitransparent inflections. For the L2 speakers, we found a robust processing cost for stem allomorphy: in both experiments, semitransparent inflections took longer time to process than transparent inflections, while there was no difference between monomorphemic nouns and transparent inflections. This has implications for theories on second language acquisition regarding the representation of morphologically complex words in the bilingual lexicon. In the following sections, we will discuss the results in detail for both groups, focusing first on the results of VLD and then on the results of the eye–tracking task. After that, we will discuss how the results align with some of the existing models of morphological processing.

In the VLD for L1 speakers, we observed a difference in processing times between all three conditions. That is, monomorphemic nouns were processed fastest, followed by transparent inflections, and then semitransparent inflections. In terms of accuracy, the L1 group was expected at the ceiling level. Processing costs in morphologically complex nouns have often been interpreted to reflect decomposition to stems and affixes (e.g., Niemi et al., 1994; Lehtonen & Laine, 2003). If this hypothesis is true, the results for L1 speakers suggest that both transparent and semitransparent forms would be decomposed. In addition, it would imply that accessing semitransparent inflections via their stems is more demanding than accessing transparent forms.

In the VLD results for L2 speakers, we found a slightly different pattern: There was no statistically significant difference in processing times between monomorphemic nouns and transparent inflections, but instead, there was a difference between transparent and semitransparent forms. A similar processing cost was also detected for accuracy. Several earlier studies have indicated that L2 speakers exhibit limited or no sensitivity to morphological structure, as exposed by a lack of morphological priming effects (e.g., Babcock, Stowe, Maloof, Brovetto, & Ullman, 2012; Basnight-Brown et al., 2007; Bowden et al., 2010; Clahsen et al., 2013; Neubauer & Clahsen, 2009; Silva & Clahsen 2008). This has been suggested to be particularly true for L2 speakers at lower proficiency levels (Babcock et al., 2012; Basnight-Brown et al., 2007; Bowden et al., 2010). However, several studies have also demonstrated the opposite, i.e., suggesting that L2 speakers tend to be sensitive to morphologically structure and decompose morphologically complex nouns (Coughlin & Tremblay, 2013; Diependaele et al., 2011; Feldman et al., 2010; Foote, 2015; Gor & Cook, 2010; Gor & Jackson, 2013). It is interesting that the transparent inflections pattern with studies that show a lack of sensitivity to the morphological structure of the L2 speakers, whereas the semitransparent inflections pattern with studies that do show the sensitivity of L2 speakers to morphological structure. It is hard to imagine, though, that the different types of inflections would be processed in a fundamentally different way.

One explanation for the discrepancy would be that L2 speakers decompose both types of inflected words into their constituents, but that they, unlike L1 speakers, mainly focus on the stem to get the central meaning of the word and spend less time understanding the meaning of the suffix or integrating the meanings of stem and suffix. The procedure would be in line with the Word and Affix model, which presumes the activation of edge–aligned embedded word stems (Beyersmann & Grainger, 2023; for Finnish; Hyönä et al., 2021). This means that L2 speakers would employ a left–to–right word–scanning strategy, where they attempt to extract meaning from the first encountered embedded word. This could result in processing times for transparent inflections that are equivalent to processing times of monomorphemic words, as the stems are usually frequent and easily recognizable. At the same time, it results in a processing delay and more errors for semitransparent inflections in comparison to monomorphemic words and transparent inflections, stemming from a combination of reasons. First, for semitransparent forms it may be more difficult to determine which letters belong to the stem, as the stem does not correspond to the dictionary form of the word, i.e., the stem is less salient for the L2 speakers. Consequently, it will be more difficult to determine the location of the morpheme boundary between the stem and inflectional suffix. We thus propose that the processing delay tied to semitransparent inflections can be attributed to a more laborious decomposition process and a slower mapping of the orthographic input with the mental representation of the bound stem among L2 speakers. Similar reasons are put forth in other studies which found that lack of morphophonological transparency delays word processing in L2 (e.g., Basnight-Brown et al., 2007; DeKeyser, 2005; Goldschneider & DeKeyser, 2001; Hahne et al., 2006; Kempe & Brooks, 2008). This is also in line with Piccinin et al. (2018), who studied the processing of Italian–bound stems in L2. They posit that "the lack of a transparent, segmentable and autonomous status does not affect L1 processing mechanisms as predicted by paradigmatic approaches, while for L2 speakers the establishing of truly morphological relationships might be impaired by formal opacity." The authors conclude by stating that this interpretation aligns with models that recognize the role of morphology while acknowledging the potential interference of formal aspects (Piccinin et al., 2018).

One important question, however, is related to task demands, as for the L1, the results were different in lexical decision tasks and natural reading. In Experiment 2, the morphological characteristics of the manipulated words did not impact the processing times in any of the three measures in L1. This is in line with the results of Hyönä et al. (2002), who found that for L1 speakers, the inflectional processing cost observed in isolation did not extend to sentence context. More precisely, they found that fixation durations were highly similar for transparent inflected and monomorphemic words. According to them, this suggests that the morphological effect observed for isolated words mainly derives from the lack of syntactic and/or semantic context, i.e., the context facilitates the recognition of transparent and semitransparent inflections more strongly relative to monomorphemic words. However, this was not the case for L2 speakers. Their pattern of differences in total fixation duration remained similar to that observed in the VLD: there was no statistically significant difference in processing times between monomorphemic nouns and transparent inflections, but instead, there was a difference between transparent and semitransparent forms. However, it is worth noting that this effect became only statistically significant in total fixation duration, which incorporates later regressions and rereadings of the target word as well. In first–pass measures such as gaze duration and selective regression path duration it did not reach significance. Our interpretation of this relatively late effect in sentence context is that

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

the processing of the semitransparent words is often not finalized or not completely successful during first–pass reading, so rereadings and regressions are frequently needed. This is in line with the notion that difficult words—whether they are unfamiliar, ambiguous, long, or unexpected—often are not fully processed upon first encounter and require more regressive saccades and more frequent rereadings (Rayner, 1998). With respect to the task demands, it is important to point out that the VLD results should be taken with caution, as inflected forms are rarely encountered in isolation. We contend that the eye–tracking study has higher ecological validity, as it investigates inflectional word processing in the context of natural reading with inflected word forms presented in their natural environment, i.e., part of the text.

In sum, this study has implications for theories in second language acquisition regarding the representation of morphologically complex words in the bilingual lexicon. Some of the previous studies have shown that late L2 learners may not be as sensitive to morphologically complex words as native speakers (Babcock et al., 2012; Basnight-Brown et al., 2007; Bowden et al., 2010). Our findings suggest that this may not be the whole truth: even beginner–level L2 speakers seem to process morphologically complex words longer when there is a stem allomorph involved. This finding is in line with the Word and Affix model, according to which stems appear to represent prominent units in the reading system, and thus, a less salient stem makes the embedded word activation more challenging. For transparent inflections, these types of challenges do not exist as the embedded transparent stem is easily recognizable.

From a more practical point of view, the results indicate that stem changes present challenges for morphological processing among L2 speakers in both single–word recognition and during sentence reading. This is in line with the findings of Vainio et al. (2014) and suggests that the difficulty in complex word recognition in L2 is more related to fusional than agglutinative features of the language. While it is acknowledged that many morphological characteristics, such as input frequency and morphophonological regularity are known to correlate with how quickly morphology is learned (Goldschneider & DeKeyser, 2001), direct comparisons between the processing of fusional language characteristics and agglutinative features in late L2 speakers has not been conducted before. The current findings suggest that fusionality may require more attention in educational settings than agglutinativity. It can be expected that similar principles could apply to other languages demonstrating agglutinative and fusional characteristics, such as Turkish or Hungarian.

Limitations and future directions

In this study, we have matched the items based on basic lexical characteristics, i.e., lemma and surface frequency and word length. We decided to match on these factors because they are the most crucial ones affecting word processing. Moreover, matching for other characteristics was challenging after matching the conditions on these critical aspects and given the restrictions in selecting inflected words with specific stem alternations for the CG conditions.

The effect of allomorphy was operationalized by comparing inflections that either included or did not include CG. There are two types of CG: qualitative, where phonemes change (e.g., *pata: padassa;* ‘pot: in a pot’), and quantitative, where phonemes vary in length (*hattu: hatussa;* ‘hat: in a hat’). In this study, we used both types. However, it is possible that there are differences in processing between quantitative and qualitative CG, as the former is slightly more productive and pertains to a change in phoneme length, whereas qualitative CG pertains to a change of the phoneme. In our study the number of items for each type of alternation was equal within the semitransparent condition, but the number of items per type was too small to explore this factor separately. It would nevertheless be interesting to explore the potential differences in processing of qualitative vs. quantitative CG. We leave this to future research.

Due to the limited access to participants, gender balance was not fully optimal in the L1 group (44 women, 8 men). In the analyses, we followed the principle of using the most parsimonious, theoretically motivated model and avoided the risk of overfitting given the relatively small sample size (Babyak, 2004). For this reason, our models included only condition, language group, trial number, and lemma frequency, but future studies with a larger sample size would benefit of exploring additional factors.

The linguistic background of our participants was diverse: among 39 participants in Experiment 1, there were 26 different first languages, and in Experiment 2, there were 18 distinct first languages in a group of 35 participants. This means that we could not study any cross–linguistic effects, as the number of individuals per language was not large enough for statistical analysis. However, it is noteworthy that even if there is expectedly a lot of L1–related noise in the data, we still observed a robust effect for stem allomorphy. It would be extremely interesting to study cross–linguistic effects more closely in the future, not only for their practical implications (e.g., language instruction for people coming from different language backgrounds), but also to shed light on the theoretical question of how L1 and L2 interact in the bilingual mental lexicon.

In L2 word processing, proficiency often plays a crucial role. The proficiency scores in this study were based on sum scores from self-assessments and exposure for the L2 group. The rationale for using these measures lies in their frequent application for assessing the proficiency levels of L2 speakers, as they have been shown to correlate with more objective measures (Marian, Blumenfeld, & Kaushanskaya 2007). In future studies, it would nevertheless be valuable to also assess Finnish proficiency against a standardized tool like the Lexize vocabulary test (Salmela, Lehtonen, Garusi, & Bertram, 2021). Note that Lexize was not utilized for this study, as it was still under development during the course of data collection.

There are two further avenues for research that could be explored in the future. In our study, there was a relatively limited range in proficiency—most participants were of low relatively proficiency (recruited from A2–B1 level language courses). However, it would be worth to investigate the interaction between L2 proficiency and morphological processing and knowledge with participants representing a wider proficiency range. By involving L2 participants covering a large range of proficiency levels, we will get a better understanding about the link between proficiency and morphological knowledge and processing skills, which may be especially important in a language as rich in morphology as Finnish.

Finally, we would like to point out that the current study is just the tip of the iceberg as the fusional characteristics in Finnish language are abundant. Other distinct fusional features in Finnish include concurrent simultaneous consonant and vowel gradation (e. g, *silta + plural marker + adessive => silloilla;* 'on the bridges'), reverse consonant gradation in nominal forms (e.g., *ranne: ranteet*; 'wrist: wrists') and verb conjugations (*hylätä: hylkäsi;* 'to reject: s/he rejected'). Moreover, Finnish may also exhibit an accumulation of fusional process when stacking multiple affixes, for instance combining derivative and inflectional affixes (e.g., *vesi + tön + ssä => vede+ttömä+ssä*; 'in (a place) without water.' It is likely that these intricate fusional processes pose even more significant challenges for L2 learners than the ones examined in this study and it would therefore be valuable to explore these further.

Conclusion

The main finding of this study is that semitransparent inflections are more demanding to process than transparent inflections for L2 speakers. In other words, the difficulty of processing bimorphemic inflections in L2 Finnish seems to primarily lie in the fusional characteristics of the language, not in the agglutinative ones. The results suggest that fusional characteristics may require extra attention in language teaching and assessment, most likely not only in Finnish but also in other languages with agglutinative-fusional typologies. This should be implemented in language instruction by raising awareness of the challenges associated with fusional morphology and by practicing stem allomorphy change patterns efficiently. However, the complexity of the topic calls for more studies focusing on the effect of L2 proficiency and comparing learners with different L1 typologies. Understanding the possible difficulties more profoundly will help us to better support learners of Finnish and other agglutinative–fusional languages.

Acknowledgments. This study was financially supported by the Åbo Akademi Minority Research Profile and the EDUCA Flagship project by Research Council of Finland. We thank Prof. Matti Laine and Prof. Jukka Hyönä for their consultation, research assistant Satu Savo for the help with data collection, and Ali Moazami Goodarzi for the statistical consultation. We also thank all the Finnish L2 teachers at various institutions in Turku who kindly allowed us to visit their classes for participant recruitment.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Babcock, L., Stowe, J., Maloof, C., Brovetto, C., & Ullman, M. (2012). The storage and composition of inflected forms in adult-learned second language: A study of the influence of length of residence, age of arrival, sex, and other factors. *Bilingualism: Language and Cognition*, *15*(4), 820–840. https://doi.org/10.1017/S1366728912000053

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411–421.

Barton, J. J. S., Hanif, H. M., Björnström, L. E., & Hills, C. (2014). The word-length effect in reading: a review. *Cognitive Neuropsychology*, *31*(5–6), 378–412. PMID: 24665973.

Basnight-Brown, D., Chen, L., Hua, S., Kostić, A., & Feldman, L. (2007). Monolingual and bilingual recognition of regular and irregular English verbs: Sensitivity to form similarity varies with first language experience. *Journal of Memory and Language 57*, (1), 65–80. https://doi.org/10.1016/j.jml.2007.03.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *i*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models.* arXiv preprint arXiv: 1506.04967.

Bertram, R. (2011). Eye movements and morphological processing in reading. *Mental Lexicon*, *6*, 83–109.

Bertram, R., Laine, M., Karvinen, K. (1999). The interplay of word formation type, Affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, *28*(3), 213–226. https://doi.org/10.1023/A:1023200313787

Bertram, R., & Hyona, J. (2023). The Importance of Eye Movement Research for Gaining Insight into Morphological Processing. In *Linguistic Morphology in the Mind and Brain* (pp. 121–136). Routledge.

Beyersmann, E., & Grainger, J. (2023). The role of embedded words and morphemes in reading. In D. Crepaldi (Ed.), *Linguistic morphology in the mind and brain* (pp. 26–49). Routledge.

Booij, G. (2010). Construction morphology. *Language and linguistics compass*, *4*(7), 543–555.

Bowden, H. W., Gelfand, M. P., Sanz, C., & Ullman, M. T. (2010). Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition. *Language Learning*, *60*(1), 44–87.

Bresnan, J., Asudeh, A., Toivonen, I., & Wechsler, S. (2015). *Lexical-functional syntax.* John Wiley & Sons.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548.

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. *Frontiers in psychology*, *6*, 645.

Clahsen, H., Balkhair, L., Schutter, J. S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research*, *29*(1), 7–31.

Coughlin, C., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34*(3), 615–646. https://doi.org/10.1017/S0142716411000890

DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, *55*(Suppl1), 1–25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

Diependaele, K., Duñabeitia, J., Morris, J., & Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *6*(4), 344–358. https://doi.org/10.1016/j.jml.2011.01.003

Feldman, L., Kostić, A., Basnight-Brown, D., Đurđević, D., & Pastizzo, M. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: Little evidence for two distinct mechanisms. *Bilingualism: Language and Cognition*, *13*(2), 119–135. https://doi.org/10.1017/S1366728909990459

Foote, R. (2015). The production of gender agreement in native and L2 Spanish: The role of morphophonological form. *Second Language Research*, *I*(3), 343–373. https://doi.org/10.1177/0267658314565691

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review*, *8*, 127–131.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*(1), 1–50.

Gor, K. & Jackson, S. (2013) Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes*, *28*(7), 1065–1091. https://doi.org/10.1080/01690965.2013.776696

Gor, K. and Cook, S. (2010), Nonnative Processing of Verbal Morphology: In Search of Regularity. *Language Learning*, *I*, 88–126. https://doi.org/10.1111/j.1467-9922.2009.00552.x

Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, *18*(1), 121–134.

Halle, M., & Marantz, A. (1994). Some key features of Distributed Morphology. *MIT Working Papers in linguistics*, *21*(275), 88.

Harley, H., & Noyer, R. (1999). Distributed morphology. *Glot International*, *4*(4), 3–9.

Hedlund, L. A., Wikman, P., Hut, S. C., & Leminen, A. (2021). Neural responses to Finnish inflected forms during overt and covert production: The role of stem frequency and stem allomorphy. *Journal of Neurolinguistics*, *57*, 100953.

Hippisley A. and Stump, G. (2016). *The Cambridge handbook of morphology*. Cambridge: Cambridge University Press.

Hyönä, J., Vainio, S., & Laine, M. (2002). A morphological effect obtains for isolated words but not for words in sentence context. *European Journal of Cognitive Psychology*, *14*(4), 417–433.

Itkonen, S., Häikiö, T., Vainio, S., & Lehtonen, M. (2024). LASTU: A psycholinguistic search tool for Finnish lexical stimuli. *Behavior Research Methods*, *14*(4), 417–433. https://doi.org/10.1080/09541440143000131

Jacob, G., Fleischhauer, E., & Clahsen, H. (2013). Allomorphy and affixation in morphological processing: A cross-modal priming study with late bilinguals. *Bilingualism: Language and Cognition*, *16*(4), 924–933.

Järvikivi, J., & Niemi, J. (2002a). Stem allomorphs as units in the mental lexicon. In R. Rapp (Ed.), *Linguistics on the way into the third millennium, Part 2* (pp. 47–58). Frankfurt am Main: Peter Lang.

Järvikivi, J., & Niemi, J. (2002b). Form-based representation in the mental lexicon: Priming (with) bound stem allomorphs in Finnish. *Brain and Language*, *I*, 412–423. https://doi.org/10.1006/brln.2001.2534

Karlsson, F. (1982). *Suomen kielen äänne ja muotorakenne. [The phonological and morphological structure of Finnish.]* Juva: Werner Söderström.

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*(4), 703–746.

Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing*, *21*, 783–804.

Kimppa, L., Shtyrov, Y., Hut, S. C., Hedlund, L., Leminen, M., & Leminen, A. (2019). Acquisition of L2 morphology by adult language learners. *Cortex*, *116*, 74–90.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876.

Laine, M., & Virtanen, P. (1999). *WordMill lexical search program*. University of Turku, Center for Cognitive Neuroscience.

Laine, M., Vainio, S., & Hyönä, J. (1999). Lexical access routes to nouns in a morphologically rich language. *Journal of Memory and Language*, *40*, 109–135.

Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research*, *35*(2), 121–46. https://doi.org/10.1007/s10936-005-9008-1

Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, *6*(3), 213–225. https://doi.org/10.1017/S1366728903001147

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. *Eye guidance in reading and scene perception*, pp. 55–75, Elsevier Science Ltd. https://doi.org/10.1016/B978-008043361-5/50004-3.

Luotolahti, J., Kanerva, J., Laippala, V., Pyysalo, S., & Ginter, F. (2015). Towards universal web parsebanks. In *Proceedings of the Third International Conference on Dependency Linguistics* (Depling 2015) (pp. 211–220). Uppsala University.

Marantz, A. (2013). No escape from morphemes in morphological processing. *Language and Cognitive Processes*, *28*(7), 905–916.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4): 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martin, M. 1995. The map and the rope: Finnish nominal inflection as a learning target. Studia Philologica Jyväskylaensia 38. University of Jyväskylä, Jyväskylä.

Meunier, F. & Longtin, C. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, *56 (4)*, 457–471. https://doi.org/10.1016/j.jml.2006.11.005

Mousikou, P., & Schroeder, S. (2019). Morphological processing in single-word and sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(5), 881.

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language: An Experimental Study of German Participles. *Studies in Second Language Acquisition*, *31*(3), 403–435. https://doi.org/10.1017/S0272263109090354

Niemi, J., Laine, M., & Tuominen, J. (1994). Cognitive morphology in Finnish: Foundations of a new model. *Language and Cognitive Processes*, *9*(3), 423–446. https://doi.org/10.1080/01690969408402126

Nikolaev, A., & Bermel, N. (2023). Studying negative evidence in Finnish language corpora. *Word Structure*, *16*(2–3), 206–232.

Nikolaev, A., Lehtonen, M., Higby, E., Hyun, J., & Ashaie, S. (2018). A facilitatory effect of rich stem allomorphy but not inflectional productivity on single-word recognition. *Applied Psycholinguistics*, *39*(6), 1221–1238. https://doi.org/10.1017/S0142716418000292

Nikolaev, A., Pääkkönen, A., Niemi, J., Nissi, M., Niskanen, E., Könönen, M., Mervaala E. & Soininen, H. (2014). Behavioural and ERP effects of paradigm complexity on visual word recognition, *Language, Cognition and Neuroscience*, *29*(10), 1295–1310, https://doi.org/10.1080/23273798.2014.912341

Orsolini, M. & Marslen-Wilson, W. (1997). Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes*, *12*(1), 1–47. https://doi.org/10.1080/016909697386899

Piccinin, S., Dal Maso, S., & Giraudo, H. (2018). Bound stem processing in L1 and L2 Italian. *Lingue e Linguaggio, Rivista Semestrale*, *2*, 289–306, https://doi.org/10.1418/91870

Pinker, S. (1999). *Words and rules*. New York, NY: Basic Books.

Portin, M., Lehtonen, M., Harrer, G., Wande, E., Niemi, J., & Laine, M. (2008). L1 effects on the processing of inflected nouns in L2. *Acta Psychologica 128* (3), 452–465. https://doi.org/10.1016/j.actpsy.2007.07.003.

R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language*, *26*(4), 392–405.

Rastle, K. & Davis, M. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*, 942–971. https://doi.org/10.1080/01690960802069730

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506.

Salmela, R., Lehtonen, M., Garusi, S., & Bertram, R. (2021). Lexize: A test to quickly assess vocabulary knowledge in Finnish. *Scandinavian Journal of Psychology*, *62*, 806– 819.

Schmidtke, D., Matsuki, K., & Kuperman, V. (2017). Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1793–1820. https://doi.org/10.1037/xlm0000411

Shmueli, G. (2010). To explain or to predict? Statistical Science, 25, 289–310.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(2), 245–260. https://doi.org/10.1017/S1366728908003404

Soveri, A., Lehtonen, M., & Laine, M. (2007). Word frequency and morphological processing in Finnish revisited. *The Mental Lexicon*, *2*(3), 359–385. https://doi.org/10.1075/ml.2.3.04sov

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of verbal learning and verbal behavior*, *14*(6), 638–647.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of verbal learning and verbal behavior*, *15*(6), 607–620.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*, 263–272.

Vainio, S., Pajunen, A., & Hyönä, J. (2014). L1 and L2 word recognition in Finnish: Examining L1 effects on L2 processing of morphological complexity and morphophonological transparency. *Studies in Second Language Acquisition*, *36*, 133–162. https://doi.org/10.1017/S0272263113000478

Wang, M., Cheng, C., & Chen, S. W. (2006). Contribution of morphological awareness to Chinese-English biliteracy acquisition. *Journal of Educational Psychology*, *98*(3), 542.

Whaley, C. P. (1978). Word—nonword classification time. *Journal of Verbal learning and Verbal behavior*, *17* (2), 143–154.

Wu, Z., & Juffs, A. (2022). Effects of L1 morphological type on L2 morphological awareness. *Second Language Research*, *38*(4), 787–812.

Zhao, J., Joshi, R. M., Dixon, L. Q., & Chen, S. (2017). Contribution of phonological, morphological and orthographic awareness to English word spelling: A comparison of EL1 and EFL models. *Contemporary Educational Psychology*, *49*, 185–194.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

APPENDIX A

Table A1. Separate model for L1 reaction times in the VLD Experiment

	log (RT)			
Predictors	Estimates	*std. error*	*CI*	*t*
(Intercept)	6.56	.05	6.46, 6.66	130.55***
Condition [Monomorphemic]	–.03	.01	–.06, –.01	–2.37*
Condition [Semitransparent]	.03	.01	.00, .06	2.15*
Lemma Frequency (log)	–.04	.01	–.06, –.01	–2.75**
Trial Number	–.00	.00	–.00, –.00	–4.36***
Random Effects				
σ^2	.04			
$\tau_{00\,\mathrm{Itemnr}}$	.00			
ICC	.28			
N_{Subj}	52			
N_{Itemnr}	141			
Observations	7212			
Marginal R^2/Conditional R^2	.018/.296			

*p<0.05; **p<0.01; ***p<0.001.

Table A2. Separate model for L2 reaction times in the VLD Experiment

	log (RT)			
Predictors	Estimates	*std. error*	*CI*	*T*
(Intercept)	7.12	.10	6.93, 7.31	73.45***
Condition [Monomorphemic]	–.04	.03	–.09, .01	–1.46**
Condition [Semitransparent]	.08	.03	.03, .14	3.11
Proficiency	–.16	.04	–.24, –.08	–4.10***
Lemma Frequency (log)	–.09	.03	–.14, –.04	–3.51***
Trial Number	–.00	.00	–.00, –.00	–2.16*
Random Effects				
σ^2	.09			
$\tau_{00\,\mathrm{Itemnr}}$	.01			
$\tau_{00\,\mathrm{subj}}$	.04			
ICC	.39			
N_{Subj}	37			
N_{Itemnr}	141			
Observations	4438			
Marginal R^2/Conditional R^2	.142/.478			

*p<0.05; **p<0.01; ***p<0.001.

Table A3. Separate model for L2 accuracy in the VLD Experiment

	Correct			
Predictors	*Odds Ratios*	*Std.Error*	*CI*	*Z*
(Intercept)	.04	.05	.01, .38	–2.85**
Condition [Monomorphemic]	.82	.26	.44, 1.54	–.61
Condition [Semitransparent]	.38	.12	.21, .70	–3.08**
Proficiency	2.43	.51	1.61, 3.67	4.21***
Lemma Frequency (log)	6.31	1.93	3.46, 11.48	6.02***

*p<0.05; **p<0.01; ***p<0.001.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table A4. Global model for gaze duration in the Eye–tracking Experiment

	log(Gaze)			
Predictors	Estimates	*Std. Error*	*CI*	*t*
(Intercept)	5.48	.10	5.28, 5.68	54.14***
Condition [Monomorphemic]	–.02	.03	–.07, .04	–.52
Condition [Semitransparent]	.02	.03	–.04, .07	.56
Language Group [L2]	.61	.06	.50, .71	10.86***
Trial number	–.00	.00	–.00, .00	–1.72
Lemma Frequency (log)	–.03	.03	–.08, .03	–.96
Condition [Monomorphemic]* Language Group [L2]	.01	.02	–.04, .05	.24
Condition [Semitransparent]* Language Group [L2]	.03	.03	–.02, .08	1.36
Random Effects				
σ^2	.16			
$\tau_{00\ \text{item\_number}}$	.01			
$\tau_{00\ \text{Subject}}$	.05			
ICC	.27			
N_{Subject}	72			
$N_{\text{item\_number}}$	104			
Observations	6440			
Marginal R^2/Condition R^2	.297/.489			

**p*<0.05; ***p*<0.01; ****p*<0.001.

Table A5. Global model for selective regression path duration in the Eye–tracking Experiment.

	log(SRPD)			
Predictors	Estimates	*std. Error*	*CI*	*t*
(Intercept)	5.58	.11	5.36, 5.80	50.59***
Condition [Monomorphemic]	–.01	.03	–.07, .05	–.21
Condition [Semitransparent]	.04	.03	–.02, .10	1.26
Language Group [L2]	.73	.06	.61, .85	11.64***
Trial number	–.00	.00	–.00, –.00	– 4.09***
Lemma Frequency (log)	–.04	.03	–.10, .02	–1.37
Condition [Monomorphemic]* Language Group [L2]	–.02	.02	–.07, .02	–.089
Condition [Semitransparent]* Language Group [L2]	.01	.02	–.04, .06	.43
Random Effects				
σ^2	.14			
$\tau_{00\ \text{item\_number}}$	.01			
$\tau_{00\ \text{Subject}}$	.07			
ICC	.36			
N_{Subject}	72			
$N_{\text{item\_number}}$	104			
Observations	6429			
Marginal R^2/Conditional R^2	.372/.597			

**p*<0.05; ***p*<0.01; ****p*<0.001.

https://doi.org/10.1017/S0272263125000026 Published online by Cambridge University Press

Table A6. Separate model for L1 total fixation duration in the Eye–tracking Experiment

	log(ToFD)			
Predictors	Estimates	*std. Error*	*CI*	*t*
(Intercept)	5.61	.15	5.33, 5.90	38.67***
Condition [Monomorphemic]	–.02	.04	–.10, .06	–.51
Condition [Semitransparent]	.04	.04	–.04, .12	1.00
Trial number	–.00	.00	–.00, –.00	–7.43***
Lemma Frequency (log)	.04	.04	–.03, .12	1.09
Random Effects				
σ^2	.20			
$\tau_{00\ \mathrm{item\_number}}$	.02			
$\tau_{00\ \mathrm{Subject}}$	.09			
ICC	.35			
N_{Subject}	37			
$N_{\mathrm{item\_number}}$	104			
Observations	3640			
Mariginal R^2/Conditional R^2	.014/.360			

**p*<0.05*; ***p*<0.01*; ****p*<0.001*.

Table A7. Separate model for L2 total fixation duration in the Eye–tracking Experiment

	log(ToFD)			
Predictors	Estimates	*std. Error*	*CI*	*t*
(Intercept)	6.95	.19	6.58, 7.33	36.46***
Proficiency	–.32	.07	–.45, –.18	–4.67***
Condition [Monomorphemic]	–.02	.05	–.13, .08	–.45
Condition [Semitransparent]	.13	.05	.02, .23	2.40*
Trial number	–.00	.00	–.00, –.00	–3.99***
Lemma Frequency (log)	–.08	.05	–.18, .02	–1.55
Random Effects				
σ^2	.22			
$\tau_{00\ \mathrm{item\_number}}$	.04			
$\tau_{00\ \mathrm{Subject}}$	.12			
ICC	.41			
N_{Subject}	34			
$N_{\mathrm{item\_number}}$	104			
Observations	2696			
Marginal R^2/Conditional R^2	.194/.528			

**p*<0.05*; ***p*<0.01*; ****p*<0.001*.

Cite this article: Salmela, R., Lehtonen, M., Vainio, S., & Bertram, R. (2025). Challenges in inflected word processing for L2 speakers: The role of stem allomorphy. *Studies in Second Language Acquisition,* 1–28. https://doi.org/10.1017/S0272263125000026