

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Hedlund, Laura A.; Wikman, Patrik; Hut, Suzanne C.A.; Leminen, Alina  
**Neural responses to Finnish inflected forms during overt and covert production**

*Published in:*  
Journal of Neurolinguistics

*DOI:*  
[10.1016/j.jneuroling.2020.100953](https://doi.org/10.1016/j.jneuroling.2020.100953)

Published: 01/02/2021

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

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Hedlund, L. A., Wikman, P., Hut, S. C. A., & 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, Article 100953. <https://doi.org/10.1016/j.jneuroling.2020.100953>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Journal of Neurolinguistics

journal homepage: [www.elsevier.com/locate/jneuroling](https://www.elsevier.com/locate/jneuroling)

Research paper

# Neural responses to Finnish inflected forms during overt and covert production: The role of stem frequency and stem allomorphy

Laura A. Hedlund<sup>a,b,\*</sup>, Patrik Wikman<sup>a</sup>, Suzanne C.A. Hut<sup>a</sup>, Alina Leminen<sup>a,b,c</sup>

<sup>a</sup> Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

<sup>b</sup> Aalto NeuroImaging, AMI Centre, Aalto University, Espoo, Finland

<sup>c</sup> Cognitive Science, Department of Digital Humanities, Faculty of Arts, University of Helsinki, Helsinki, Finland

## ARTICLE INFO

## Keywords:

Inflection  
Morphology  
Allomorphy  
Pseudoword  
fMRI  
Production  
Finnish  
Gradation

## ABSTRACT

Despite extensive behavioral research on complex word recognition, the neural mechanisms involved in the production of inflections in agglutinative languages, such as Finnish, are still poorly understood. Finnish inflected nouns typically involve morphophonological alternations of the stem (i.e. consonant gradation; CG), which is less common in other languages. Behavioral research on recognition of inflected nouns containing consonant gradation has shown that the number of stem allomorphs results in faster recognition times. To our knowledge, no functional magnetic resonance imaging (fMRI) studies have explicitly investigated consonant gradation in word production. In this study, participants performed covert and overt production tasks during event-related fMRI. Our stimuli comprised real word stems of high and medium frequency as well as pseudoword stems. The stems either included consonant gradation or were non-gradating. Our findings showed that the production of inflected forms containing high frequency stems or CG stems (irrespective of frequency and lexicality) yield enhanced activation of the left inferior frontal and middle frontal gyri (LIFG and MFG, respectively). This suggests that CG stems, that is, stems with more than one allomorph, facilitate lexical lookup, and that the activation of multiple stem allomorphs is reflected in increased recruitment of frontal brain regions.

## 1. Introduction

In this study we investigated neural substrates during the online production of inflected words. We focused on the question of whether lexical or grammatical rule operations are utilized during the inflection of words that undergo stem alteration (consonant gradation; CG). More specifically, we investigated how these lexical or grammatical processes manifest on the neural level. In addition to the whole brain analysis, we performed region of interest (ROI) analysis, which allowed for a theory-driven investigation of inflectional production.

In the field of neurolinguistics, the perception and production of morphologically complex words has been a topic of vigorous research. An example of morphological complexity is plural case inflection in English (e.g., ‘shark + s’). Previous behavioral studies

\* Corresponding author. Cognitive Brain Research Unit, Dept. of Psychology and logopedics, Faculty of Medicine, University of Helsinki, Haartmaninkatu 3, PL 21, F 219, Finland.

E-mail address: [laura.hedlund@helsinki.fi](mailto:laura.hedlund@helsinki.fi) (L.A. Hedlund).

<https://doi.org/10.1016/j.jneuroling.2020.100953>

Received 17 October 2019; Received in revised form 17 August 2020; Accepted 3 September 2020

Available online 19 September 2020

0911-6044/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

have frequently shown that inflected words yield longer reaction times and higher error rates during recognition and production than their non-inflected counterparts (e.g. Laine, 1996; Laine & Koivisto, 1998; Laine, Vainio, & Hyönä, 1999; Niemi et al., 1994; Lehtonen et al., 2009; Hakala et al., 2018, see Cohen-Goldberg (2013) for a review on production), suggesting combinatorial processing of inflected forms. In some languages, parsing inflected words into their constituents seems to be the rule rather than the exception, since it promotes efficient storage and processing (Yang, 2016). A particularly interesting language, due to its rich and complex morphology is Finnish, an agglutinative language using as many as 17 inflectional cases (Karlsson, 2008). In Finnish, approximately 95% of polymorphemic words occur only once per million (Vannest et al., 2002). Thus, storing items of a low surface frequency as whole forms would be extremely taxing on memory systems. Indeed, a body of behavioral evidence suggests that the majority of inflected Finnish words are processed combinatorially, that is, inflected forms with high surface or lemma frequency are stored holistically (Lehtonen & Laine, 2003; Soveri et al., 2007; Vartiainen et al., 2009). Lehtonen et al. (2006), however, reported that in addition to high frequency inflected nouns, inflected nouns with medium lemma frequency seem to be stored as full forms as well, while low frequency items remain to be parsed and stored as decomposed units in the mental lexicon.

In addition to *lexical frequency*, the *lexicity* factor has also been frequently accounted for in several languages, by including inflected pseudowords as experimental stimuli. For instance, Meunier and Longtin (2007) reported that the parsing of French complex pseudowords into their constituents takes place early in the processing stage, while grammatical and semantic integration of morphemes is completed later. In Finnish, Laine (1996) and Hakala et al. (2018) reported longer reaction times and higher error rates for the processing of words where a real inessive suffix (-ssa/-ssä) was attached to a Finnish pseudoword stem (e.g., \*lora + ssa, 'in a \*lora') in comparison to monomorphemic pseudowords (e.g., \*peiksa). This indicates that the number of real word morphemes is correlated with morphological processing costs. The findings from these two studies have been interpreted as evidence for combinatorial processing of inflected pseudowords, as parsing should take more time to perform and are more prone to error. However, other studies (Laine, Vainio, & Hyönä, 1999; Leinonen et al., 2009), found no significant differences in behavioral responses between monomorphemic pseudowords and inflected pseudowords, suggesting that pseudoword stems do not activate a morpheme-based lexical access route. Thus, whether inflected pseudoword stems are processed holistically or combinatorially and which neural areas are functionally tied to these processes are still relevant questions.

Furthermore, a less addressed aspect of morphologically rich languages, such as Finnish and Hungarian, is *consonant gradation* (CG), despite its occurrence in over 30 percent of the Finnish vocabulary (Hakulinen et al., 2004). CG refers to morphophonological changes that mostly affect consonants *k*, *p*, and *t* that occur in the last syllable of a monomorphemic stem (Gehring & Heinzmann, 2011; Karlsson, 2008). This phenomenon is morphophonological in nature, that is, it is driven by grammar but manifests phonologically (Niemi et al., 1994). Nouns in Finnish may undergo three different types of consonant gradation that are driven by grammatical case endings, transitioning from a 'strong grade' to a 'weak grade' (Karlsson, 2008). For example, an inflected form such as 'in Helsinki', would involve adding an inflectional case marker for 'in' (-ssa/-ssä; -a/ä determined by vowel harmony rules), and the strong grade '-nk' would undergo assimilative CG to the weak grade '-ng' *Helsingissä* ('in Helsinki'). Another example of CG is quantitative gradation, in which a double consonant, such as the strong grade plosive '-pp' in 'kauppa' ('a store'), is shortened to one '-p' ('kauppa' + '-ssa' *kaupassa*/*\*kauppassa* ['in the store']). Lastly, a third type of CG is qualitative, in which, typically, plosives (*p*, *t*, *k*) change from being their strong voiceless grade to their weaker voiced grade. This is seen in the stem *luku* ('chapter') and its change to *luvussa* ('in the chapter'). Thus, in order to inflect a noun stem, particularly in the inessive case, changing of the stem into its bound stem allomorph and then adding the case ending (with its possible allomorph) is required (Karlsson, 2008). Despite some research on recognition of CG and stem allomorphy (e.g., Järvikivi & Niemi, 2002; Laine, 1996; Nemeth et al., 2015; Niemi et al., 1994), there are no studies examining the production of such items in Finnish.

To account for CG or stem allomorphy in other languages, for instance, Nemeth and colleagues (Nemeth et al., 2015) have examined the production of Hungarian regular (*alma*-apple *almák*-apples) and irregular (*ló*-horse *lovak*-horses) real nouns that undergo stem changes (allomorphy), similar to that of Finnish. They found that irregular inflected forms were more susceptible to error, and took longer to produce than regular inflected forms. The authors suggest that this could partially be due to the extra step required when retrieving the changed stem form for the corresponding suffix. This would further imply less automaticity for the production of graded words with stem changes.

Apart from behavioral investigation, the processing of monomorphemic and inflected stimuli has also been extensively studied with neuroimaging techniques. Accounting for lexicity, several studies have reported that the processing of monomorphemic pseudowords activates the left inferior frontal gyrus (LIFG) more strongly than real words in lexical decision tasks (Heim et al., 2005; Xiao et al., 2005; Hagoort et al., 1999). Enhanced activation in the LIFG has also been observed during the recognition of Finnish inflected real words, compared to monomorphemic words (Laine, Rinne, et al., 1999; Lehtonen et al., 2006, 2009). In addition to activation of the left inferior frontal cortex and the left middle frontal gyrus (MFG), monomorphemic word retrieval has activated the cerebellum (lobules VI as well as Crus I and II and of lobule VII) (Gurd et al., 2002; see Price, 2012 for a review).

While several studies have focused on the perception of monomorphemic and inflected words, neuroimaging studies targeting the production (overt and covert) of inflected real words or pseudowords are scarcer. One of the few fMRI studies on overt speech production of inflected words in Russian have shown that inflected real and pseudo-verbs are processed compositionally, irrespective of their regularity (Slioussar et al., 2014). Other covert (Balaguer et al., 2006; Beretta et al., 2003; de Diego; Marangolo et al., 2006) and overt (Desai et al., 2006; Joanisse & Seidenberg, 2005) fMRI production studies in German, Spanish, Italian, and English have found that generation of inflected words mainly activates the LIFG. This corroborates findings on inflected word recognition, and suggests that BA 44 of the LIFG is involved in combinatorial processing of morphological units (Marslen-Wilson & Tyler, 2007; Ullman 2001, 2016; Vannest et al., 2005), both in recognition and production. Other brain regions consistently reported in the production of inflected words are the left cerebellum (de Diego-Balaguer et al., 2006; Oh et al., 2011), right cerebellum (Nevat et al., 2017; Oh et al.,

2011; Pliatsikas et al., 2014; Slioussar et al., 2014), left putamen (Desai et al., 2006), and left and bilateral thalamus (Desai et al., 2006; Tyler et al., 2004). While the caudate nucleus of the basal ganglia has been associated with lexical operations, the putamen has been tied to the application of grammatical rules (Teichmann et al., 2008).

Lexical and grammatical operations underlying morphological processing at the brain level have also been explained by at least two main theoretical accounts. For instance, the *Declarative-Procedural model* (DP model; Ullman, 2001) proposes that while the lexicon is dependent on declarative memory, grammar is mostly governed by procedural memory. However, when unproductive (irregular) transformations are required, declarative memory is involved in such processing. According to the model, BA 44, basal ganglia (via the thalamus), and cerebellum are putatively the neural correlates of the procedural memory system, while the declarative memory system is associated with activity in BA 45, the hippocampus, as well as connected temporal and parietal neocortical regions (Pliatsikas et al., 2014; Ullman, 2001, 2004, 2016). The *Core Decompositional Network model* (CDM; Marslen-Wilson & Tyler, 2007) does not include explicit predictions of whether subcortical structures, that is, the basal ganglia, thalamus, or cerebellum are involved in grammar processing. The CDM model postulates that all regularly inflected words undergo parsing, which engages the left fronto-temporal subsystem (anterior cingulate cortex, LIFG, STG, left inferior parietal lobule, and middle temporal gyrus). While BA45 of the LIFG, in particular, is associated with morphological parsing, the left STG and left anterior cingulate cortex sub-serve the application of grammatical rules, whereas bi-hemispheric IFG activation purportedly reflects holistic processing (Marslen-Wilson & Tyler, 2007). While both models are largely based on English data, fMRI and PET findings on the Finnish language (Laine, Rinne, et al., 1999; Lehtonen et al., 2006) seem to be in line with the Core Decompositional Model. Some evidence shows, however, that other inflectionally rich languages, such as Polish, may store regular inflections holistically in the mental lexicon (Reid & Marslen-Wilson, 2000; Bozic, Szlachta, & Marslen-Wilson, 2013).

To specifically account for Finnish inflection, the *Stem Allomorph/Inflectional Decomposition* (SAID) model was developed (Laine et al., 1994; Niemi et al., 1994). The model postulates that during recognition, inflected words are decomposed into a free stem + suffix, or a bound stem allomorph + suffix. During production, this model proposes that inflected forms are represented in such a manner that the nominative stem and possible decomposed bound stem allomorph(s) are activated in unison (e.g., 'kauppa' and 'kaup-'). These stem allomorphs are then used as indexes in the mental lexicon prior to the semantic-syntactic integration of the morphological constituents (e.g., inessive case: 'kaupa-ssa' ['in the store']). However, the SAID model does not have explicit predictions for the neural correlates of either production or recognition of inflected forms.

Taken together, there is a need for more neurocognitive evidence on how inflected words, particularly those with CG stems, are produced in morphologically rich languages. Hence, in response to this need, we investigated brain mechanisms underlying covert and overt production of inflected words in Finnish.

## 2. The present study

In this study, we specifically examined how (1) CG affects operational strategies (lexical or grammatical) that are involved in inflectional production and how this is reflected in the brain (2) *Word Status* (i.e. combined word frequency and lexicality); modulates inflection and associated brain mechanisms. To this end, we included different stimulus categories: high frequency real word stems (hf), middle frequency (mf) real word stems, and pseudoword stems (p). All stems either demanded consonant gradation while inflecting (CG) or not (NoCG). Thus, combining *Word Status* with CG resulted in six types of stems (CG-hf, CG-mf, CG-p, NoCG-hf, NoCG-mf, NoCG-p). The task of the participant was to inflect the visually presented monomorphemic stems either covertly (85% of conditions) or overtly (15% of the conditions).

Based on prior experimental evidence as well as the DP and CDM models, we made several hypotheses regarding expected performance (error rates). Because high frequency stems by definition have a high level of everyday exposure, they should have a more robust representation in the mental lexicon compared to middle frequency stems and pseudoword stems (having no exposure at all). Hence, inflected forms with high frequency stems should be the least laborious to inflect, resulting in fewer errors during covert and overt production. In contrast, because pseudoword stems by definition are not as such represented in the mental lexicon (Shtyrov et al., 2011), these stems should be the most laborious to inflect. Middle frequency stems might either cause error rates falling between these extremes, or show similar error rates to high frequency stems (Lehtonen et al., 2006).

With respect to CG, we hypothesized that inflecting the CG stems will be associated with more errors as compared to Non-CG stems, irrespective of *Word Status*. According to Nemeth et al. (2015), when inflecting words that undergo extra processing steps, that is, stem changes, one needs to retrieve the appropriate stem allomorph from the mental lexicon before an inflectional suffix can be added. This possible extra processing step should thus manifest as increased error rates in both covert and overt production.

At the brain level, we focused primarily upon regions such as the middle frontal gyrus (MFG), due to its importance in lexical word retrieval (for a review see Price, 2012). In addition, we expected that: (1) *Word Status* should influence activity in the BA45, MFG, the hippocampus, as well as connected temporal and parietal neocortical regions (Pliatsikas et al., 2014; Ullman, 2001, 2004, 2016). The DP model predicts that high frequency words should activate these brain regions more strongly than words with lower frequency and pseudowords. CG should activate BA45 (the retrieval of the correct allomorph), irrespective of *Word Status*. This is because CG words are associated with an extra processing step that involves retrieval of the correct allomorph stems (Nemeth et al., 2015); (2) Inflecting CG-pseudoword stems compared to NoCG-pseudoword stems might depend more heavily on procedural memory, because these words do not have lexical memory traces. Thus, according to the DP model, they should be associated with activation of the BA44 basal ganglia (via the thalamus) and cerebellum.

In contrast, the CDM (Marslen-Wilson & Tyler, 2007) does not include explicit predictions of whether subcortical structures, such as the basal ganglia, thalamus, or cerebellum are involved in morphological processing. This model also makes opposite predictions

about the effect of Word Status in the BA45. According to the CDM, the BA45 supports stem decomposition and stem-lookup prior to suffix concatenation for lower frequency words and pseudowords. Thus, there should be an increased activation of the BA45 for pseudoword stems compared to high and medium frequency real word stems, because high and medium frequency items should mostly depend on full-form retrieval (Lehtonen et al., 2006; Lehtonen & Laine, 2003; Soveri et al., 2007; Vartiainen et al., 2009).

### 3. Materials and methods

#### Ethical approval

This study was carried out according to the regulations and guidelines set forth by the Ethical principles of research in the Humanities and Social and Behavioral Sciences; conveyed by the University of Helsinki Ethical review board in Humanities and Social and Behavioral Sciences. All participants gave written informed consent in accordance with the Declaration of Helsinki.

#### 3.1. Participants

Fifteen right-handed native speakers of Finnish, seven females, mean age 31.3, range 18–44 participated in the experiment. Handedness was assessed by a shortened Finnish version of the Edinburgh handedness inventory (Oldfield, 1971). There were no reports of neurological, psychiatric disorders, or alcohol/drug dependencies. All participants had normal or corrected-to-normal vision, and normal hearing. Each participant was given a consent form, MRI safety checklist in her/his native language.

#### 3.2. Stimuli

Stimuli consisted of six stem types, of which half of them would require CG in the inessive case inflection: CG-p (pseudo), CG-mf (medium frequency), CG-hf (high frequency); stems that would not need CG: NoCG-p NoCG-mf, and NoCG-hf. To obtain the final list of CG- and NoCG-high and medium frequency stems (45 in each group), we used the Match program with an initial list of 270 words (see Van Casteren & Davis, 2007) to balance for: 1) length of words (in letters) for lemma and surface forms, 2) logarithmic corpus lemma frequencies within frequency categories, 3) gradation type within gradation categories 4) rated word frequency. We obtained lemma frequencies from the Language Bank of Finland's, Korp [Corpora: Finnish Tree Bank, 1990–2000 Luvun Suomalaisia Aikakausi-ja Sanomalehtiä, Suomi 24 2016 H2, KLK Suomenkieliset Lehdet (1980–2000)] using 132 million tokens (Borin et al., 2012). In addition, we wanted to verify whether the corpus lemma frequencies would reflect everyday word usage (see Shtyrov et al., 2011; Tryk, 1968). To this end, we additionally estimated a rated word frequency for each lemma (based on ratings from 5 Finnish speaking subjects, not included in the experiment). The rated word frequencies had a high correlation with the corpus-based lemma frequencies ( $r = 0.533$ ,  $n = 180$ ,  $p = <0.001$ ). Initially, we had aimed at having both high and low per million frequencies stems, but other factors such as lemma length, surface length (in letters), as well as matching across gradation, had to be carefully considered. Thus, in order to get two statistically different sets of words we had to make low frequency words slightly higher than typically used (see e.g., Lehtonen et al., 2006). Hence, they are more closely comparable with “medium” surface frequency (per million) than with low frequency. The inclusion of pseudoword stems into the stimuli design was strictly to investigate the online application of consonant gradation rules without possible confounding variables of high and medium frequency real word stems.

Examples of stimuli and their lemma frequencies are shown in Table 1. The factor Consonant Gradation (CG, NoCG) was subjected to a one-way repeated measures ANOVA to ensure that high frequency words ( $[F(1,44) = 0.985$ ,  $p = 0.326$ ,  $\eta^2 = 0.022]$ ) and medium frequency words ( $[F(1,44) = 0.612$ ,  $p = 0.438$ ,  $\eta^2 = 0.014]$ ) did not differ in their lemma log frequencies.

**Table 1**

Stimuli Properties.

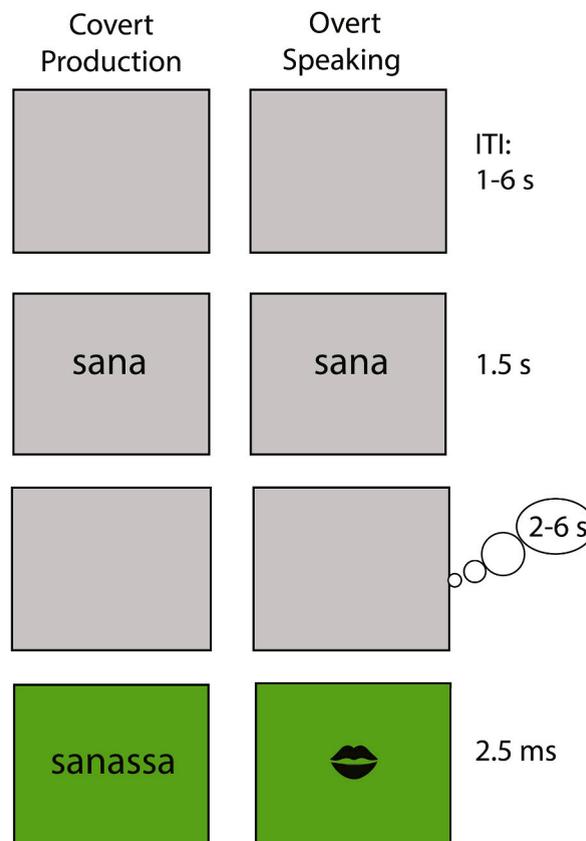
Examples of stimuli and their average lemma log frequency, as well as their corresponding standard deviations (SD) and p-values. Four, one-way repeated measures ANOVAs (factor of Word Status: high, medium, pseudo) were used to test the mean difference of the average surface as well as the average lemma lengths for gradated and non-gradated stems.

Word Type	Example	Average Lemma Log Frequency (SD)	t-test (p)	Average Lemma Length (SD)	ANOVA	Average Surface Length (SD)	ANOVA
<b>Gradated high frequency (CG-hf)</b>	Akku → Akussa ( <i>Battery → In the battery</i> )	1.85 (0.4)	$p = < 0.00$	5.5 (0.9)	$p = .938$	8.0 (0.6)	$p = .777$
<b>Gradated medium frequency (CG-mf)</b>	Tupa → Tuvassa ( <i>Cottage → In the cottage</i> )	1.10 (0.4)		5.4 (1.1)		7.9 (0.8)	
<b>Gradated pseudo (CG-p)</b>	Mappa → Mapassa	N/A	N/A	5.5 (0.9)		7.9 (0.7)	
<b>Non-Gradated high frequency (NoCG-hf)</b>	Asema → Asemassa ( <i>Station → In the station</i> )	1.94 (0.3)	$p = < 0.00$	5.4 (0.9)	$p = .968$	8.4 (0.9)	$p = .975$
<b>Non-gradated medium frequency (NoCG-mf)</b>	Aula → Aulassa ( <i>Lobby → In the lobby</i> )	1.18 (0.4)		5.4 (1.0)		8.4 (1.0)	
<b>Non-gradated pseudo (NoCG-p)</b>	Puovi → Puovissa	N/A	N/A	5.4 (0.8)		8.4 (0.8)	

### 3.3. Procedure

In the experimental task (See Fig. 1), the participants were presented with a randomized series of the monomorphemic stems from each of the six word categories. Participants were asked to always silently generate the inflected form of the monomorphemic stems in the inessive case. In 85% of the cases, the silent generation was followed by an evaluation of the inflected form, where either a correctly or incorrectly inflected form would appear. The participant would press their right pointer finger on a response pad if the inflected word was the same as they silently produced, and their middle finger if the inflected form was different from the one they had generated silently. In 15% of the trials, the participants were prompted to overtly produce the inflected form. A mouth icon appearing after the monomorphemic stem word/pseudoword stem presentation would signal the overt production. Prior to the experiment, each participant was instructed to speak into a microphone (Optoacoustics two-way full duplex FOMRI, Or Yehuda, Israel) in a ventriloquist-like manner, i.e., as loudly and as clearly as possible, while limiting the use of facial muscles. The task consisting of 15% overt generation allowed us to ensure that the participant would always silently generate the inflected word during the given time period, as we randomized the appearance of the overt production prompt and its appearance was unbeknownst to the participants (for a similar method, see (Hut & Leminen, 2017)). This kind of a paradigm allowed us to verify the accuracy of the inflected form generation, opposed to a fully silent generation task. Furthermore, due to only 15% of trials involving minimal facial movement, motion artifacts were not a significant factor, thus allowing us to include the overt speech trials into the fMRI analysis.

Stems were displayed via a Panasonic PT-DZ110XEJ projector (Osaka, Japan), on a back projection screen for 1.5 s (black text on grey background, Arial font, 24 pt) using a mirror, which was attached to the head coil. Thereafter a blank grey screen was displayed for an interval randomly chosen between 2 and 6 s, serving as the silent generation period. After this interval, in 85% of the trials the subjects were presented with either the correct or incorrect inflected form (2.5 s) to respond to (covert task). In 15% of the trials a mouth icon (overt task) was displayed for 2.5 s serving as a prompt that the subjects were to overtly produce the inflected word aloud. There was a randomly varying interval between 1 and 6 s of rest between the trials (inter trial interval; ITI), see Fig. 1. During each run, there were three 40-s breaks denoted by a hollow circle, allowing the participants' BOLD signal to abate. In between the two runs, 5 minutes of anatomical scans were completed, while the participant watched a silent, subtitled movie or closed their eyes.



**Fig. 1.** Experimental Task.

The task started with a word stem being displayed, in this case a non-gradated high frequency 'sana' (i.e. word), after that the subject covertly inflected the word in inessive case (2-6 s). After this either the correct or wrongly inflected form was displayed (85% of the cases) and the participant answered "correct/incorrect", or a mouth icon was displayed (15% of the trials, overt task) where after the participant overtly produced the inflected word.

The presentation of the stimuli was written and commanded by a script, in Presentation 18.1 (Neurobehavioral Systems, Albany, CA, United States). Each participant was in the MRI scanner for a combined time of approximately 75 min, which was split into two experimental task runs (~25 min each), as well as structural and diffusion tensor imaging (DTI, results reported elsewhere) scans.

### 3.4. Behavioral data acquisition and analysis

The accuracy of the covert button press responses was first analyzed by extracting any misses (no button press), and extremely fast reaction times (<100 ms). Any statistical operations were only carried out on the remaining trials. As the task was not designed to test how quickly participants responded, reaction times were not necessarily accurate, and thus we report only error rate data.

The accuracy of the participants' overt speech was evaluated offline for correctness by two independent Finnish-speaking raters. Their inter-rater reliability agreement (McHugh, 2012) was 98.8%, with an intraclass correlation coefficient (Pearson) of 0.83. The final error rate percentage of the overt spoken responses was determined after unintelligible and missed trials were excluded. One participant's overt data had to be excluded due to an MRI noise-filter malfunction, where the voice responses were indiscernible from scanner noise, however the participant successfully performed the task involving button presses, which comprised 85% of the experiment trials. The statistical analysis was carried out on the remaining trials.

### 3.5. fMRI data acquisition and analysis

Whole-brain functional and anatomical images were acquired using a 3T Siemens MAGNETOM Skyra MRI scanner (Siemens Healthcare, Erlangen, Germany) with Syngo MR D13C software and a 32-channel head-neck coil. Functional images were acquired using a T2 gradient-echo echo planar imaging (EPI) pulse sequence with  $42 \times 3.5$  mm axial slices, with axial slices gap: 0.5 mm, TE: 32 ms, TR: 2330 ms, flip angle:  $75^\circ$ , FOV:  $224 \times 224$  mm, in-plane matrix resolution:  $64 \times 64$ . High-resolution T1-weighted MPRAGE gradient-echo anatomical images were collected with  $176 \times 1$  mm slices (TE: 3.30 ms, TR: 2530 ms, TI: 1100 ms, FOV:  $256 \times 256$  mm, image matrix resolution:  $256 \times 256$ ).

Data analysis was carried out in FSL version 5.0 (Jenkinson et al., 2012). Motion correction was completed using MCFLIRT (Jenkinson et al., 2002). The data quality, and possible artifacts of the motion from the overtly spoken trials were carefully checked, as they could have caused noise in the data. Image distortion was corrected using EPI unwarping (unwrap direction: y, EPI TE: 35 ms, EPI dwell time: 0.7 ms, signal loss threshold: 10%). Each image was spatially smoothed with a Gaussian kernel, with Full Width at Half Maximum (FWHM) of 9 mm (Pajula & Tohka, 2014); grand-mean intensity normalization of the entire 4D dataset was done by a single multiplicative factor; high-pass temporal filtering was achieved using Gaussian-weighted least-squares straight line fitting, with sigma equal to 30.0s.

The first level analysis was carried out using FSL's fMRI Expert Analysis Tool (FEAT) version 6.0. The six different word types (CG-hf, CG-mf, CG-p; NoCG-hf, NoCG-mf, NoCG-p) were then modeled in a general linear model (GLM) as different explanatory variables (EVs) (see Supplementary Table 1). All participants' data was registered to MNI space using FLIRT (Greve et al., 2009; Jenkinson et al., 2001; Jenkinson et al., 2002). B0 unwarping and simultaneous registration was carried out with boundary-based registration (BBR) with 3 degrees of freedom to the initial structure.

Based on the first level statistics, we performed six,  $2 \times 2$  repeated measures ANOVAs across subjects using FSL. The first-level statistics were modeled this way because FSL does not allow correction for sphericity that is necessary in repeated measures ANOVAs with more than two levels (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/GLM>). The significance was determined by Z-transformed F-maps in FEAT (version 6.0; FMRIB's Software Library, [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)), the cluster defining threshold was determined by  $Z > 2.3$  and a cluster significance threshold of  $p$  being  $<0.05$  (Worsley, 2001). The rendered images were made using Freesurfer (version 6.0, [freesurfer.net](http://freesurfer.net)), and the coordinates of the significant cortical regions reported were retrieved using Freesurfer's Freeview, and automatic cortical parcellation (Destrieux et al., 2010). The coordinates and Z-scores of the subcortical areas and areas of the cerebellum were retrieved with FSL's (version 5.0) FSLeves.

The contrasts included in these ANOVAs were: 1) NoCG- and CG-hf vs. NoCG- and CG-p; 2) NoCG- and CG-hf vs. NoCG- and CG-mf; 3) NoCG-hf and NoCG-mf vs. CG-hf and CG-mf; 4) NoCG-hf and NoCG-p vs. CG-hf and CG-p; 5) NoCG-mf and NoCG-p vs. CG-mf and CG-p; as well as 6) NoCG- and CG-mf vs. NoCG- and CG-p.

### 3.6. Region of interest (ROI) analysis

Based on our a priori theory-driven assumptions and the significant main effects for the whole brain ANOVAs, we selected the BA44, BA45, and MFG as regions of interest (ROI). For each ROI, the mean percentage signal change against silent rest was extracted for the six conditions: CG-hf, CG-mf, CG-p; NoCG-hf, NoCG-mf, NoCG-p. The ROIs were further divided according to the Juelich histological atlas for the regions BA44, BA45 (Amunts et al., 1999); MNI Cerebellar Atlas in MNI-152 Space After Normalization with FLIRT was used for cerebellar regions Crus I, Crus II, VI, VIIb, VIIIa (Diedrichsen et al., 2009); and the Harvard-Oxford Subcortical Structural Atlas for the thalamus, caudate, putamen, and pallidum.

## 4. Results

### 4.1. Behavioral data

Statistical analyses were performed in IBM SPSS Statistics, Version 25.0 (Armonk, NY: IBM Corp). Analysis of the error rates for the covert production task was performed with two-way repeated-measures ANOVA with the factors Gradation (NoCG, CG) and Word Status (pseudowords, real words with a high frequency, real words with a medium frequency). In case of violations of sphericity, Greenhouse-Geisser adjustments were performed (corrected degrees of freedom are reported). The ANOVA showed that there was a main effect of Gradation [ $F(1,14) = 57.24, p < .0001, \eta^2 = 0.804$ ] because CG stems elicited higher error response, as well as Word Status [ $F(1.458, 20.416) = 56.81, p < 0.0001, \eta^2 = 0.802$ ] because pseudoword stems yielded higher error rates compared to real word stems with a high and medium frequency. There was also an interaction between Consonant Gradation  $\times$  Word Status [ $F(1.357, 18.99) = 28.08, p < 0.0001, \eta^2 = 0.66$ ], because the gradation effect was strongest for inflected forms with pseudo word stems. See Fig. 2 for post hoc pairwise comparisons (Bonferroni corrected).

For overt responses, there were only errors in the CG-mf condition (1.5%, SEM: 1.03 and the CG-p condition (12.08%, SEM: 2.84). The performance was error free in all other conditions.

### 4.2. fMRI results

For the whole brain analysis, the role of word status (frequency and lexicality) as well as gradation was investigated using six separate two-way ANOVAs (initial cluster defining threshold,  $Z = 2.3$ , and cluster probability being  $p < 0.05$ ; Worsley, 2001). We observed significant main effects in the following ANOVAs for **Consonant Gradation**: NoCG-hf and NoCG-mf vs. CG-hf and CG-mf (Fig. 3a); NoCG-hf and NoCG-p vs. CG-hf and CG-p (Fig. 3c). These main effects were significant because CG-stems, regardless of their frequency or lexicality, yielded a greater response than NoCG-stems (see rendered images depicted in Fig. 3, and a complete list of the significant brain areas in Supplementary Table 1). We also observed significant main effects in the following ANOVAs for **Word Status**: NoCG- and CG-hf vs. NoCG- and CG-p (Fig. 3b); NoCG- and CG-mf vs. NoCG- and CG-p (Fig. 3d).

### 4.3. Regions of interest results

Three separate four-way ANOVAs were performed for the different ROIs. The first group consisted of frontal regions (BA44, BA45, MFG; Fig. 4); the second of subcortical areas (Caudate, Pallidum, Putamen, Thalamus; Supplementary Fig. 1A), and the third consisted of different subcomponents of the cerebellum (Crus I, Crus II, VI, VIIB, VIIB; Supplementary Fig. 1B). The other factors in each ANOVA were Hemisphere (left, right), Consonant Gradation (CG, NoCG), and Word Status (high, medium, pseudo). All ANOVA results were Greenhouse-Geisser corrected when necessary.

The ANOVA for the frontal regions revealed significant main effects of CG [ $F(1,14) = 12.2, p = 0.004, \eta^2 = 0.466$ ], and Word Status

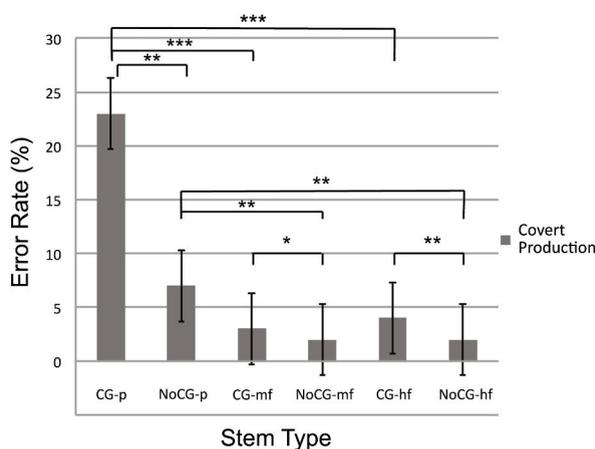
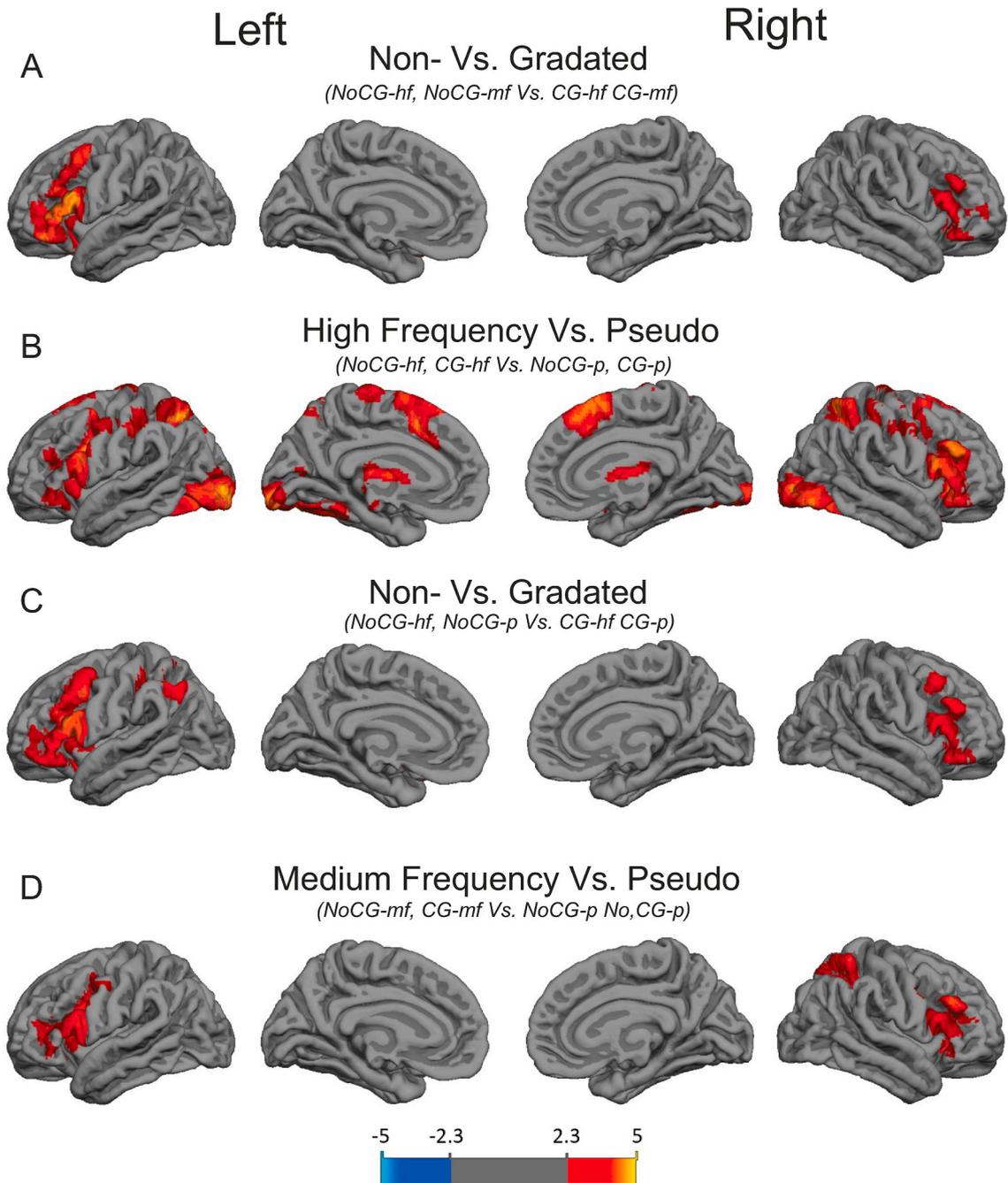


Fig. 2. Covert Production Button Presses.

Caption: Percentage of error rates for covert (button presses) conditions for each stem type: CG-p (gradated pseudo), NoCG-p (non-gradated pseudo), CG-mf (gradated medium frequency), NoCG-mf (non-gradated medium frequency), CG-hf (gradated high frequency), NoCG-hf (non-gradated high frequency). Error bars denote SEM, and asterisks indicate significant differences between the stem categories\*

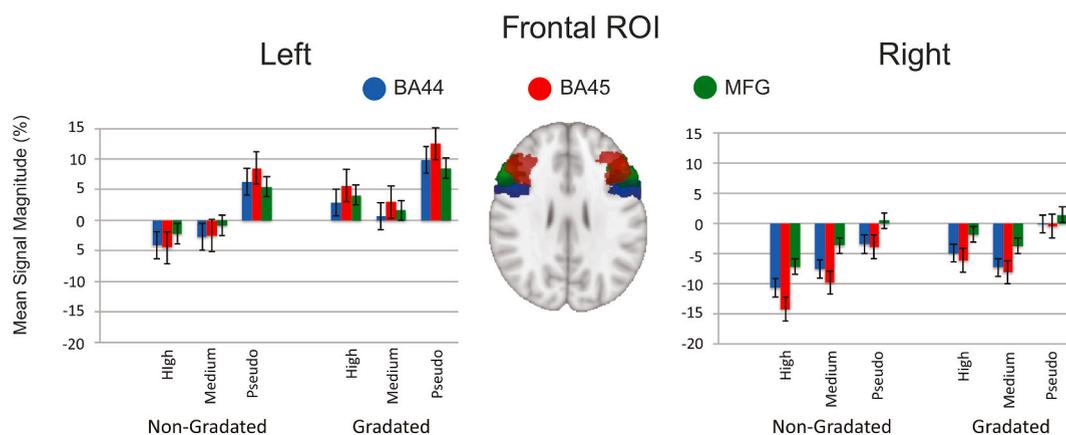
$p < 0.05$ , \*\*  
 $p < 0.01$ , \*\*\*  
 $p < 0.001$ .



**Fig. 3.** Rendered Images.

Significant Main effects of Task in all six ANOVAs projected onto an average cortical surface (fsaverage). The brightness of the color indicates the p-value. Cluster-wise  $p < 0.05$  (FDR-corrected). A: Main effect of Gradation in the ANOVA for high, low frequency stems. B: Main effects of Word Status in the ANOVA for high frequency versus pseudo stems. C: Main effects of Gradation in the ANOVA with high frequency and pseudo stems. D: Main effects of Word Status in the ANOVA for medium frequency versus pseudo stems. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

[ $F(1.449, 20.289) = 13.167, p = 0.001, \eta^2 = 0.485$ ]. The Main effect of Consonant Gradation was due to a higher activation for CG-stems versus NoCG-stems. The main effect of Word Status in the frontal ROI was due to a lower activation for pseudoword stems compared to stems with a medium frequency (mean differences: 0.071;  $p = 0.009$ ), and stems with a high frequency (mean difference: 0.072;  $p = 0.002$ ). An interaction between Region and Consonant Gradation was also significant [ $F(2, 28) = 8.4, p = 0.001, \eta^2 = 0.375$ ]. A post hoc test revealed significantly higher BOLD activations for CG-stems versus NoCG-stems in all three frontal regions:



**Fig. 4. Region of Interest.**

Bar charts represent mean % signal change ( $\pm$ SEM) as compared to rest. The color-coded brain plots are used only for illustrative purposes to accompany the bar graphs. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

BA44 (CG-stems versus NoCG-stems mean difference: 0.038;  $p = 0.011$ ), BA45 (CG-stem versus NoCG-stem mean difference: 0.057;  $p = 0.0003$ ) and MFG (CG-stem versus NoCG-stem mean difference: 0.030;  $p = 0.030$ ).

In the ANOVA for the cerebellar regions and the subcortical regions there were only significant main effects of Word Status [ $F(2, 28) = 9.9$ ,  $p = 0.001$ ,  $\eta^2 = 0.415$ ;  $F(1.304, 18.257) = 6.7$ ,  $p = 0.013$ ,  $\eta^2 = 0.325$ ]. In both cases, this was due to a lower BOLD activation during covert and overt production of inflected pseudowords compared to real stems with a high frequency (subcortical mean difference: 0.043;  $p = 0.003$ ; cerebellum mean differences: 0.060;  $p = 0.007$ ), see [Supplementary Fig. 1](#).

## 5. Discussion

We investigated brain mechanisms underlying covert and overt production of inflected words and pseudowords in Finnish. We included real word stems with a medium or high frequency, as well as pseudoword stems. When inflected in the inessive case, stems either underwent CG stem changes or did not. Our behavioral results on covert production demonstrated a significant increase in error rates for the production of inflected forms containing CG stems as compared to No-CG stems. Irrespective of CG, there were more errors when inflecting pseudoword stems compared to real word stems with a medium and high frequency. The only errors in overt production were made in the CG-mf and CG-p stem conditions. The lower error rate in the overt production task, compared to covert production, suggests that there might be an inner inspection of phonological speech errors, prior to the inflected form being uttered aloud. One explanation for increased errors in covert production is offered by the covert repair hypothesis ([Postma & Kolk, 1993](#)). According to this hypothesis, the covert production task could trigger purely lexical access, since there is no need for an internal phonological check for correctness, the way there is for overt production ([Hartsuiker et al., 2005](#)). This also explains an increase in error rates in covert production for CG stems, as this phenomenon is purely morphophonological in nature ([Gehring & Heinzmann, 2011](#); [Karlsson, 2008](#)). Our ROI fMRI results illustrated that there was a significant increase in BOLD activation in BA44, BA45, and MFG when inflecting medium and high frequency stems, compared to pseudo stems. In the same three ROIs, there was a significant increase in BOLD activation for inflected word CG-stems, compared to NoCG-stems. In addition, subcortical and cerebellar regions showed higher activation for forms with high and medium frequency than for those with pseudo stems. Next, we will discuss each result in more detail.

According to the Stem Allomorph/Inflectional Decomposition (SAID) model ([Laine et al., 1994](#); [Niemi et al., 1994](#)), native speakers of Finnish utilize combinatorial processing for inflected words during both recognition and production, except for those of a very high frequency. We, however, did not find significant differences in error rates between high and medium frequency stems during covert production. Our results on the production of inflected forms suggest both high and medium frequent items are stored in their whole-forms in the output lexicon. This corroborates the findings by [Lehtonen et al. \(2006\)](#), who reported that the recognition of high and medium frequency stems was evidently performed in a holistic manner.

Moreover, observing decreased rates of errors for NoCG-stems suggests that inflection of NoCG-mf and NoCG-hf stems might activate the full-form. Increased error rates of CG-stems could, in turn, signify a mistake during stem allomorph lookup. Additionally, we observed an increase in error rates for CG-p stems, as compared to NoCG-p stems, suggesting application of combinatorial processes when producing forms with a pseudoword stem. Our results are in line with the work of Laine and colleagues (1999), where they observed increased error rates for inflected forms consisting of pseudoword stems and real suffixes, in comparison with monomorphemic pseudowords. We also suggest that the increase in error rates for inflections with CG-stems, compared to NoCG-stems, could be attributed to the need for online stem gradation, or the increased effort of activating the correct allomorph prior to suffix concatenation ([Laine et al., 1994](#); [Niemi et al., 1994](#)).

In line with our initial expectations, (c)overt production of inflections with high frequency stems elicited the highest BOLD

activation in the LIFG. This also supports previous findings on recognition of pseudowords. For instance, in a visual lexical decision paradigm, Binder et al. (2003) reported that while pseudoword processing was behaviorally more taxing, the BOLD signal showed that the real words elicited stronger activations as compared to non-words. This pattern of activity has also been observed in neuromagnetic studies, where monomorphemic real words yielded stronger neural activation than pseudowords (MacGregor et al., 2012; Shtyrov et al., 2007). Thus, this lexicality effect observed in all three of our ROI areas (frontal, subcortical, and cerebellar), supports the assumption that these areas are involved in lexical access (Ullman, 2001, 2006), also in the output lexicon. Moreover, we observed that the subcortical and cerebellar structures were only sensitive to word frequency, but not to CG. Our word frequency finding is in line with electrophysiological studies (e.g., Alexandrov et al., 2011; Shtyrov et al., 2011), where stronger neural activation was interpreted as reflecting stronger memory traces for high frequency words, when compared to lower frequency words. In our study, pseudowords stems had the least amount of BOLD activation in the VIIB for both the CG and NoCG conditions. The VIIB has been commonly associated with working memory processes, and the creation of motor memory traces ((Durisko & Fiez, 2010); (Stoodley & Schmahmann, 2009)). Our finding suggests that VIIB might be associated with, for instance, articulatory rehearsal needed for speech production.

We could not fully align our neuroimaging results with any particular model. Recall that according to Ullman's (2016) DP model, BA45 is related to the lexical retrieval, whereas BA44 is related to rule-governed grammatical operations. The CDM, in turn, proposes that BA45 is associated with morphological decomposition (Marslen-Wilson & Tyler, 2007; see also (Schremm et al., 2018)). We, however, did not observe significant differences in the activation of BA44 and BA45, when using stimuli to specifically target rule application (i.e., CG stems) and lexical retrieval (Word Status, i.e., word frequency and lexicality). Hence, we could not find concrete evidence for a functional specificity of these areas and this still requires closer scrutiny in the future.

Our prediction that there would be greater BOLD signal in the three ROI clusters for CG stems compared to No-CG stems was confirmed. Recent studies on the processing of Finnish allomorphy have examined factors of stem frequency and productivity (inflectional regularity and irregularity; Nikolaev et al., 2014; Nikolaev et al., 2018). In addition to stem frequency, the number of allomorphs a particular stem may have affected the processing. Namely, the recognition of inflected words is quicker for unproductive stems that have at least three allomorphs (e.g., water 'vesi': ves-, vet-, vete-, ved-), as compared to productive stems (e.g., plastic 'muovi': muovi-, muove-). Nikolaev and colleagues therefore postulated that a higher number of allomorphs facilitate recognition. This is because during lexical access, the allomorphs are unified, since no meaning is assigned until a suffix is attached. Thus, there might be a stronger or broader neuronal network for words with more stem allomorphs compared to words with less or no stem variants. While we did not specifically contrast productivity and unproductivity of allomorphs, we nevertheless witnessed stronger BOLD activity to stems that have more allomorphs (CG) than those that have fewer (NoCG). A significant effect of CG was observed in the frontal ROI, suggesting that MFG, BA44, and BA45 are involved in the allomorph retrieval.

Finally, according to the SAID model, during recognition, morphological decomposition is applied in all but highly frequent inflections, taking place before semantic integration of morphological constituents. We observed the increased BOLD signal for real word CG stems with high frequency and lower BOLD activation for CG pseudoword stems. Hence, our findings suggest that production of inflected forms is compositional even for highly frequent stems when inflection requires consonant gradation.

## 6. Summary and conclusions

Our findings suggest that the increased BOLD signal in the frontal, subcortical, and cerebellar ROIs may be related to lexical memory traces of stem or allomorph frequency. Our findings also suggest that there is a facilitation effect when producing inflected forms with CG stems. More specifically, there is an initial activation of all stem allomorphs and a higher number of allomorph variants for a stem would overall strengthen the memory trace of the stem. The allomorph selection is followed by choosing the proper stem allomorph and the concatenating it with the correct suffix. Furthermore, we propose that the inflections of CG stems are produced by activating all stem allomorphs, and this process is tied to a network of frontal areas (BA44, BA45, and MFG). To further investigate the effect of stem allomorphs in production of inflected forms, future research should also control for a number of allomorphs a stem possesses, as well as stem productivity.

### CRedit authorship contribution statement

**Laura A. Hedlund:** Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Investigation. **Patrik Wikman:** Supervision, Validation, Writing - review & editing. **Suzanne C.A. Hut:** Conceptualization, Methodology, Data curation, Writing - review & editing. **Alina Leminen:** Supervision, Validation, Writing - review & editing, Funding acquisition.

### Acknowledgements

This study was supported by Kone Foundation (PI AL), University of Helsinki (own research funds, PI AL) and by the Russian Science Foundation (grant code RSF 19-18-00534), Doctoral Programme in Psychology, Learning and Communication (University of Helsinki, Finland), and Helsinki University Research Foundation. We thank Tatu Huovilainen for his help with stimulus preparation and data collection.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jneuroling.2020.100953>.

## References

- Alexandrov, A. A., Boricheva, D. O., Pulvermüller, F., & Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS One*, 6(8).
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412(2), 319–341.
- Beretta, A., Campbell, C., Carr, T. H., Huang, J., Schmitt, L. M., Christianson, K., et al. (2003). An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain and Language*, 85(1), 67–92.
- Binder, J. R., McKiernan, K. A., Parsons, M. E., Westbury, C. F., Possing, E. T., Kaufman, J. N., et al. (2003). Neural correlates of lexical access during visual word recognition. *Journal of Cognitive Neuroscience*, 15(3), 372–393.
- Borin, L., Forsberg, M., & Roxendal, J. (2012). *Korp – the corpus infrastructure of språkbanken proceedings of LREC 2012* (pp. 474–478). Istanbul: ELRA.
- Bozic, M., Szlachta, Z., & Marslen-Wilson, W. D. (2013). Cross-linguistic parallels in processing derivational morphology: Evidence from Polish. *Brain and Language*, 127(3), 533–538.
- Cohen-Goldberg, A. M., Cholin, J., Miozzo, M., & Rapp, B. (2013). The interface between morphology and phonology: Exploring a morpho-phonological deficit in spoken production. *Cognition*, 127(2), 270–286.
- Desai, R., Conant, L., Waldron, E., & Binder, J. R. (2006). fMRI of past tense processing: the effects of phonological complexity and task difficulty. *Journal of Cognitive Neuroscience*, 18, 278–297.
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1), 1–15.
- Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage*, 46(1), 39–46.
- de Diego-Balaguer, R., Rodríguez-Fornells, A., Rotte, M., Bahlmann, J. J., Heinze, H.-J. J., & Münte, T. F. (2006). Neural circuits subserving the retrieval of stems and grammatical features in regular and irregular verbs. *Human Brain Mapping*, 27, 874–888.
- Durisko, C., & Fiez, J. A. (2010). Functional activation in the cerebellum during working memory and simple speech tasks. *Cortex*, 46(7), 896–906.
- Gehring, S., & Heinzmann, S. (2011). *Suomen mestari 1: Suomen kielen oppikirja aikuisille* (pp. 28–29). Helsinki: Finn Lectura.
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72.
- Gurd, J. M., Amunts, K., Weiss, P. H., Zafiris, O., Zilles, K., Marshall, J. C., & Fink, G. R. (2002). Posterior parietal cortex is implicated in continuous switching between verbal fluency tasks: an fMRI study with clinical implications. *Brain*, 125(5), 1024–1038.
- Hagoort, P., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. *The neurocognition of language* (pp. 273–316). Oxford: Oxford University Press.
- Hakala, T., Hultén, A., Lehtonen, M., Lagus, K., & Salmelin, R. (2018). Information properties of morphologically complex words modulate brain activity during word reading. *Human Brain Mapping*, 39(6), 2583–2595.
- Hakulinen, A., Vilkkuna, M., Korhonen, R., Koivisto, V., Heinonen, T. R., & Alho, I. (2004). *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hartsuiker, R. J., Kolk, H. H., & Martensen, H. (2005). In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *The division of labor between internal and external speech monitoring: Phonological encoding and monitoring in normal and pathological speech (Chapter 11)*. New York: Psychology Press.
- Heim, S., Alter, K., Ischebeck, A. K., Amunts, K., Eickhoff, S. B., Mohlberg, H., et al. (2005). The role of the left Brodmann's areas 44 and 45 in reading words and pseudowords. *Cognitive Brain Research*, 25(3), 982–993.
- Hut, S. C., & Leminen, A. (2017). Shaving bridges and tuning kitaraa: the effect of language switching on semantic processing. *Frontiers in Psychology*, 8, 1438.
- Järviö, J., & Niemi, J. (2002). Form-based representation in the mental lexicon: Priming (with) bound stem allomorphs in Finnish. *Brain and Language*, 81(1–3), 412–423.
- Jenkinson, M., Bannister, P., Brady, J. M., & Smith, S. M. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). *Fsl NeuroImage*, 62(2), 782–790.
- Jenkinson, M., & Smith, S. M. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156.
- Joanisse, M. F., & Seidenberg, M. S. (2005). Imaging the past: Neural activation in frontal and temporal regions during regular and irregular past-tense processing. *Cognitive, Affective, & Behavioral Neuroscience*, 5(3), 282–296.
- Karlsso, F. (2008). *Finnish: An essential grammar*. New York: Routledge.
- Laine, M. (1996). Lexical status of inflectional and derivational suffixes: Evidence from Finnish. *Scandinavian Journal of Psychology*, 37(3), 238–248.
- Laine, M., & Koivisto, M. (1998). Lexical access to inflected words as measured by lateralized visual lexical decision. *Psychological Research*, 61(3), 220–229.
- Laine, M., Niemi, J., Koivuselkä-Sallinen, P., Ahlsén, E., & Hyönä, J. (1994). A neurolinguistic analysis of morphological deficits in a Finnish-Swedish bilingual aphasic. *Clinical Linguistics and Phonetics*, 8(3), 177–200.
- Laine, M., Rinne, J. O., Krause, B. J., Teräs, M., & Sipilä, H. (1999). Left hemisphere activation during processing of morphologically complex word forms in adults. *Neuroscience Letters*, 271(2), 85–88.
- Laine, M., Vainio, S., & Hyönä, J. (1999). Lexical access routes to nouns in a morphologically rich language. *Journal of Memory and Language*, 40(1), 109–135.
- Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6(3), 213–225.
- Lehtonen, M., Vorobyev, V. A., Hugdahl, K., Tuokkola, T., & Laine, M. (2006). Neural correlates of morphological decomposition in a morphologically rich language: An fMRI study. *Brain and Language*, 98(2), 182–193.
- Lehtonen, M., Vorobyev, V., Soveri, A., Hugdahl, K., Tuokkola, T., & Laine, M. (2009). Language-specific activations in the brain: Evidence from inflectional processing in bilinguals. *Journal of Neurolinguistics*, 22(5), 495–513.
- Leinonen, A., Grönholm-Nyman, P., Järvenpää, M., Söderholm, C., Lappi, O., Laine, M., et al. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: Evidence from a morphologically rich language. *Brain Research*, 1275, 54–66.
- MacGregor, L. J., Pulvermüller, F., Van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, 3, 711.
- Marangolo, P., Piras, F., Galati, G., & Burani, C. (2006). Functional anatomy of derivational morphology. *Cortex*, 42(8), 1093–1106.
- Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 362(1481), 823–836.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Meunier, F., & Longtin, C. M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56(4), 457–471.
- Nemeth, D., Janacek, K., Turi, Z., Lukacs, A., Peckham, D., Szanka, S., Gazso, D., Lovassy, N., & Ullman, M. T. (2015). The production of nominal and verbal inflection in an agglutinative language: Evidence from Hungarian. *PLoS one*, 10(3), Article e0119003.
- Nevat, M., Ullman, M. T., Eviatar, Z., & Bitan, T. (2017). The neural bases of the learning and generalization of morphological inflection. *Neuropsychologia*, 98, 139–155.

- Niemi, J., Laine, M., & Tuominen, J. (1994). Cognitive morphology in Finnish: Foundations of a new model. *Language & Cognitive Processes*, 9(3), 423–446.
- 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*, 1–18.
- Nikolaev, A., Pääkkönen, A., Niemi, J., Nissi, M. J., Niskanen, E., Könönen, M., et al. (2014). Behavioural and ERP effects of paradigm complexity on visual word recognition. *Language, Cognition and Neuroscience*, 29(10), 1295–1310.
- Oh, T. M., Tan, K. L., Ng, P., Berne, Y. I., & Graham, S. (2011). The past tense debate: Is phonological complexity the key to the puzzle? *NeuroImage*, 57, 271–280.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Pajula, J., & Tohka, J. (2014). Effects of spatial smoothing on inter-subject correlation based analysis of fMRI. *Magnetic Resonance Imaging*, 32(9), 1114–1124.
- Pliatsikas, C., Johnstone, T., & Marinis, T. (2014). fMRI evidence for the involvement of the procedural memory system in morphological processing of a second language. *PLoS one*, 9(5), Article 97298.
- Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech, Language, and Hearing Research*, 36(3), 472–487.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816–847.
- Reid, A., & Marslen-Wilson, W. (2000). Organising principles in lexical representation: Evidence from Polish. *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 387–392).
- Schremm, A., Novén, M., Söderström, P., van Westen, D., & Roll, M. (2018). Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain and Language*, 176, 42–47.
- Shtyrov, Y., Kimppa, L., Pulvermüller, F., & Kujala, T. (2011). Event-related potentials reflecting the frequency of unattended spoken words: A neuronal index of connection strength in lexical memory circuits? *NeuroImage*, 55(2), 658–668.
- Shtyrov, Y., Osswald, K., & Pulvermüller, F. (2007). Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. *Cerebral Cortex*, 18(1), 29–37.
- Slioussar, N., Kireev, M. V., Chernigovskaya, T. V., Kataeva, G. V., Korotkov, A. D., & Medvedev, S. V. (2014). An ER-fMRI study of Russian inflectional morphology. *Brain and Language*, 130, 33–41.
- Soveri, A., Lehtonen, M., & Laine, M. J. (2007). Word frequency and morphological processing in Finnish revisited. *The Mental Lexicon*, 2(3), 359–385.
- Stoodley, C. J., & Schmahmann, J. D. (2009). The cerebellum and language: evidence from patients with cerebellar degeneration. *Brain and Language*, 110(3), 149–153.
- Teichmann, M., Gaura, V., Demonet, J. F., Supiot, F., Delliaux, M., Verny, C., et al. (2008). Language processing within the striatum: Evidence from a PET correlation study in Huntington's disease. *Brain*, 131(4), 1046–1056.
- Tryk, H. E. (1968). Subjective scaling of word frequency. *American Journal of Psychology*, 81(2), 170–177.
- Tyler, L. K., Bright, P., Fletcher, P. C., & Stamatakis, E. A. (2004). Neural processing of nouns and verbs: the role of inflectional morphology. *Neuropsychologia*, 42(4), 512–523.
- Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30(1), 37–69.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1–2), 231–270.
- Ullman, M. T. (2006). Is Broca's area part of a basal ganglia thalamocortical circuit? *Cortex*, 42(4), 480–485.
- Ullman, M. T. (2016). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In *Neurobiology of language* (pp. 953–968). Academic Press.
- Van Casteren, & Davis. (2007). Match, A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, 39(4), 973–978.
- Vannest, J., Bertram, R., Järviö, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *Journal of Psycholinguistic Research*, 31(2), 83–106.
- Vannest, J., Polk, T. A., & Lewis, R. L. (2005). Dual-route processing of complex words: New fMRI evidence from derivational suffixation. *Cognitive, Affective, & Behavioral Neuroscience*, 5(1), 67–76.
- Vartiainen, J., Aggularo, S., Lehtonen, M., Hultén, A., Laine, M., & Salmelin, R. (2009). Neural dynamics of reading morphologically complex words. *NeuroImage*, 47(4), 2064–2072.
- Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional MRI: An introduction to methods*. Oxford: Oxford Press (Chapter 14).
- Xiao, Z., Zhang, J. X., Wang, X., Wu, R., Hu, X., Weng, X., et al. (2005). Differential activity in left inferior frontal gyrus for pseudowords and real words: An event-related fMRI study on auditory lexical decision. *Human Brain Mapping*, 25(2), 212–221.
- Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press (Chapter 3).