

A facilitatory effect of rich stem allomorphy but not inflectional productivity on single-word recognition

ALEXANDRE NIKOLAEV
University of Helsinki

MINNA LEHTONEN
University of Helsinki, Abo Akademi University, and University of Oslo

EVE HIGBY
University of California, Riverside

JUNGMOON HYUN and SAMEER ASHAIE
Northwestern University

Received: March 12, 2017 Revised: March 19, 2018 Accepted: March 21, 2018

ADDRESS FOR CORRESPONDENCE

Alexandre Nikolaev, University of Helsinki, Helsinki Collegium for Advanced Studies, P.O. Box 4 (Fabianinkatu 24), FIN 00014 University of Helsinki, Finland. E-mail: alexandre.nikolaev@helsinki.fi

ABSTRACT

The aim of the present study was to investigate whether the recognition speed of Finnish nominal base forms varies as a function of their paradigmatic complexity (stem allomorphy) or productivity status. Nikolaev et al. (2014) showed that words with greater stem allomorphy from an unproductive inflectional class are recognized faster than words with lower stem allomorphy from a productive inflectional class. Productivity of an inflectional paradigm correlates with the number of stem allomorphs in languages like Finnish in that unproductive inflectional classes tend to have higher stem allomorphy. We wanted to distinguish which of these two characteristics provides the benefit to speed of recognition found by Nikolaev et al. (2014). The current study involved a lexical decision task comparing three categories of words: unproductive with three or more stem allomorphs, unproductive with two stem allomorphs, and productive with two stem allomorphs. We observed a facilitation effect for word recognition only for unproductive words with three or more stem allomorphs, but not for unproductive words with two allomorphs. This effect was observed particularly in words of low to moderate familiarity. The findings suggest that high stem allomorphy, rather than productivity of the inflectional class, is driving the facilitation effect in word recognition.

Keywords: Finnish language; inflectional productivity; lexical decision; morphology; stem allomorphy; word retrieval

Lexical characteristics such as frequency of use and phonological neighborhood density have been studied extensively in terms of how they affect the speed of word recognition and retrieval. Stem allomorphy (i.e., the number of different stems representing a lexeme) has not been considered among the variables affecting speed of lexical access because it does not exist in many languages. In languages such as English, Finnish, and Russian, words may present different stems in different morphological contexts, such as *foot-~~feet~~* to differentiate number and *teach-~~taught~~* to indicate different tenses. Greater stem allomorphy is associated with unproductive inflectional patterns (e.g., Eng. *foot-~~feet~~*, *teach-~~taught~~*, Fin. *vesi-~~vedet~~* “water,” *sulkea-~~suljen~~* “to close,” Rus. *дно-~~донья~~* “bottom,” *сесть-~~сяды~~* “to sit”) whereas lower stem allomorphy or no stem variation is associated with productive inflectional patterns (e.g., Eng. *table-~~tables~~*, *walk-~~walked~~*, Fin. *lasi-~~lasit~~* “glass,” *haluta-~~halusi~~* “to want,” Rus. *стол-~~столы~~* “table,” *читать-~~читал~~* “to read”).

It is not clear whether stem allomorphy is beneficial or costly for word processing. Evidence for a possible disadvantage of having stem allomorphs can be found in language acquisition studies. In the earliest stages of acquiring first language morphology, children usually prefer productive morphological patterns over unproductive ones (e.g., Kim, Marcus, Pinker, Hollander, & Coppola, 1994; Slobin, 1985). Further evidence comes from language ontogenesis: words with lower frequency of use tend to shift from an unproductive type to a competing productive inflectional type (e.g., Lieberman, Michel, Jackson, Tang, & Nowak, 2007). As a result, in unproductive inflectional classes, the number of words is usually small, and these few words tend to have relatively high frequency of use. The situation is typically the opposite for productive classes; they include a lot of words and most have relatively low frequency of use (e.g., Baayen, 2009; Nikolaev & Niemi, 2008). Because words with greater stem allomorphy tend to be part of unproductive inflectional classes, this may result in a disadvantage for processing words with greater stem allomorphy.

Psycholinguistic studies have shown that productive, regular words in English and Dutch are recognized more quickly than unproductive, irregular words (Baayen & Moscoso del Prado Martín, 2005; Jaeger et al., 1996; Tabak, Schreuder, & Baayen, 2005). According to Baayen and Moscoso del Prado Martín (2005), (ir)regularity of a stem form across an inflectional paradigm is confounded with differences in semantic density (i.e., the association of more meanings with a given word). They found that irregular verbs in English and Dutch tend to have more meanings, and the semantic neighborhoods of those words also contain more irregulars. High semantic density renders a word more ambiguous, which can explain the longer response latencies for irregular words (Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017).

However, in contrast to what has been found for English and Dutch, Nikolaev et al. (2014) showed that, in Finnish, words from unproductive (irregular) inflectional types with high stem allomorphy are recognized more quickly than words from productive (regular) types. Reaction times and an N400 component from an event-related potentials experiment were sensitive to stem allomorphy, showing a facilitative effect for words with higher stem allomorphy. For example, nouns with

Table 1. A partial number and case matrix of a subset of Finnish *i*-final noun paradigms

Singular nominative	Singular partitive	Singular genitive	Singular essive	Plural partitive	Plural illative	English gloss
vesi	vet-tä	vede-n	vete-nä	vesi-ä	vesi-in	“water”
savi	save-a	save-n	save-na	savi-a	savi-in	“clay”
lasi	lasi-a	lasi-n	lasi-na	lase-ja	lase-ihin	“glass”

rich stem allomorphy (e.g., *vesi* “water” ~ *vede-n* GEN, *vet-tä* PART, *vete-nä* ESS) exhibited shorter response latencies than nouns without stem variation in singular (e.g., *lasi* “glass” ~ *lasi-n* GEN, *lasi-a* PART, *lasi-na* ESS). To explain this pattern, the authors conjectured that activation of multiple stem allomorphs at the lemma level (see the Lemma model of lexical selection, Levelt 2001; see also Taft, 2003, 2004; Taft & Kongious, 2004) facilitates lexical access from form to meaning.

Perhaps the reason that unproductive words were recognized more quickly than productive words in Nikolaev et al.’s (2014) study was because they had a high degree of stem allomorphy, which the English and Dutch words do not have. The aim of the present study is to test this hypothesis. Nikolaev et al. used only two Finnish noun types: nouns with higher stem allomorphy from an unproductive inflectional class and nouns with lower stem allomorphy from a productive inflectional class. Thus, it is unclear whether the facilitation was due to differences in the productivity/regularity of the word classes or differences in degree of stem allomorphy. In other words, productivity (regularity) was confounded with allomorphy. In the present study, we aim to disentangle these two variables by including a third noun type: an unproductive class of Finnish nouns (*savi* “clay”-type nouns) with a low degree of stem allomorphy (two allomorphs: sg. *save-*, pl. *savi-*). Table 1 provides an overview of the three noun paradigms.

ALLOMORPHY, REGULARITY, AND PRODUCTIVITY

Finnish inflected singular forms of *savi*-type and *vesi*-type nouns are unproductive and are completely unpredictable in their inflected targets (e.g., there is no way to predict that “*savi*” becomes “*saven*” in the genitive singular) and also in their form (“*savi*” → “*saven*,” but “*vesi*” → “*veden*”). Hence, they exhibit the same characteristics as English irregular forms. However, a crucial difference between English and Finnish irregulars is that in Finnish, there are segmental changes in the stem, whereas in English, stem changes are often formally fusional (there is no affix at all in “*sang*,” and stripping the affix off “*brought*” gives you “*brough*,” which is not a free stem). Finnish inflectional classes, by contrast, are fully regular (predictable) in the case of both number suffixes (singular or plural) and case suffixes (nominative, genitive, partitive, inessive, etc.). By contrast, stem changes across the inflectional paradigm are not regular, such as those seen in *savi*-type and *vesi*-type words.

The Finnish inflectional classes we chose are all *i*-final. In their nominative singular form, there are no phonological cues for Finnish speakers about the words' exact inflectional paradigms. Speakers have to know to which inflectional class each *i*-final word belongs. If they do not have this knowledge, as in the case of novel words or pseudowords, they choose the productive *lasi*-type class as a default (Nikolaev, 2002).

As can be seen from Table 1, *savi*-type nouns have the same number of allomorphs as the productive type *lasi*, namely, two allomorphs. However, having more than one allomorph already in the singular (see Table 1, sg. *savi*, sg. *save*-, cf. sg. *lasi*, sg. *lasi*-) is enough to make an inflectional type unproductive. *Savi*-type nouns are just as unproductive as *vesi*-type nouns, which have three or more allomorphs. Both *savi*- and *vesi*-type noun classes have approximately 100 words each (type frequency) compared to more than 4,000 words for the *lasi*-type noun class. Both *savi*- and *vesi*-type nouns come from an old lexical stratum, and the median of their token frequency in the corpus we used is 1,760 for *savi*-type and 1,867 for *vesi*-type (cf. to 34 for *lasi*-type). Thus, both noun types fulfill the basic criteria of an unproductive, frozen inflectional category.

There are reasons why words with multiple stem forms might be recognized more quickly. Different stem forms of the same lexical entry should be represented separately at the form level but unified at the lexical concept level since there is variation in the form, but no variation in the meaning. Thus, activation of multiple forms at the lemma level, and, at the same time, activation of the same meaning at the lexical concept level should result in a broader neuronal network for words with higher stem allomorphy compared to words with no stem variants. When a participant must initiate a word-recognition response (e.g., pressing a button) as quickly as possible following the detection of a stimulus, the possibility of parallel pathways to the motor system facilitates response latency by utilizing the principle of the fastest racer or the fastest group of racers (Miller & Ulrich, 2003; Raab 1962; Schröter, Frei, Ulrich, & Miller, 2009). Thus, when a stimulus activates multiple parallel codes, it leads to statistical facilitation.

The three inflectional classes we used (*lasi*, *savi*, and *vesi*) differ on two aspects: (a) the number of allomorphs (two for *lasi* and *savi*, three or more for *vesi*), and (b) productivity (*lasi* being productive and *savi* and *vesi* being unproductive). If word recognition speed for *savi*-type nouns patterns with the productive and regular noun class with two allomorphs (*lasi* "glass" -type nouns: sg. *lasi*-, pl. *lase*-), in other words, showing slower word recognition than *vesi*-type nouns, this would provide evidence that rich stem allomorphy is driving the facilitation of word recognition.

EFFECT OF INFLECTED FORMS ON MONOMORPHIC WORD RECOGNITION

Traditional decompositional models (Frauenfelder & Schreuder, 1992; Taft, 2004) assume that morphologically complex words are processed through their morphemes. Inflected words like *tables* are composed in production and

decomposed in recognition, hence, existing only at the moment of processing. However, if these inflected forms have a very high frequency of use, they may nevertheless be stored in our mental lexicon as a whole form (Lehtonen et al., 2007; Lehtonen & Laine, 2003; Lehtonen, Niska, Wande, Niemi, & Laine, 2006; Soveri, Lehtonen, & Laine, 2007). In the race model (Frauenfelder & Schreuder, 1992; Schreuder & Baayen, 1995), a polymorphemic word is processed both as a whole word (e.g., *childcare*) and as separate morphemes (e.g., *child* and *care*) in parallel. Having these two routes (the whole word and its morphemes) allows language users to recognize some polymorphemic words faster than monomorphemic words (e.g., Bertram, Laine, & Karvinen, 1999; Ji, Gagné, & Spalding, 2011). According to another decompositional model (Taft, 2004), there is no whole-word processing stage at all. Instead, any facilitatory or inhibitory effect of decomposition can be explained by the interaction of the two stages of processing: an early stage, where the base frequency effect arises (e.g., the cumulative frequency of all the associated word forms, e.g., *seem*, *seemed*, *seems*, and *seeming*) and a late stage, where the surface frequency effect arises (e.g., the frequency of the specific word form, e.g., *seeming*).

How do the inflected nouns *feet* or *tables* influence recognition of their base form counterparts *foot* and *table*, respectively? Because inflected forms such as *feet* (“irregular” verbs in English) are fusional, they are problematic for these models, which assume decomposition of polymorphemic words. The race model does not specify the possible influence of inflected morphologically complex words on the processing of their monomorphemic counterparts. The obligatory decomposition model, in contrast, includes irregular forms like *feet* to the base frequency of the word *foot* because these two word forms share a lemma (Taft, 2004). Thus, if we attempt to explain reaction times for monomorphemic words like *foot*, *teach*, *table*, and *walk* (or, as in the current study, Finnish nouns like *lasi*, *vesi*, and *savi*) by incorporating into the model their base frequencies in addition to their surface frequencies among other explanatory variables, then, according to Taft (2004), we would not find any effect of stem allomorphy because the variance will be explained by the interaction of an early stage of processing (the base frequency effect) and a late stage (the surface frequency effect).

Nevertheless, two recent studies have demonstrated that inflected words influence the processing of their monomorphemic forms. Caselli, Caselli, and Cohen-Goldberg (2016, Experiment 2) investigated the production of monomorphemic words, and Nikolaev et al. (2014) assessed the speed of recognition of monomorphemic nouns.

Caselli et al.’s (2016) findings challenge the prediction of traditional decompositional models that productive and regularly inflected words are less likely to have whole-word representations than derived polymorphemic words. The authors found that phonological neighborhood density of monomorphemic words predicted their duration during speech production. The phonological neighborhood density of their inflected counterparts also independently predicted the duration of monomorphemic words. The authors concluded that all forms of a word, whether morphologically simple or complex, influence spoken-word production.

Nikolaev et al. (2014) came to a similar conclusion for word recognition. The authors argued that stem allomorphs (i.e., bound morphemes) are represented at the lemma level, which mediates two major subsystems (form and meaning). Thus, the recognition of monomorphemic words includes the intermediate level of the lemma model (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999), where, according to Nikolaev et al. (2014), stem allomorphs are explicitly represented. The activation of multiple stem allomorphs allows speakers to recognize a target word faster, even if the word itself is monomorphemic. The benefits of such processing are explainable by probability summation (Raab, 1962), which posits that the reaction time in a given trial reflects the processing time of the route that happens to be the fastest on that trial. In other words, the route from a form to a meaning can go either via a base stem allomorph (e.g., *vesi*) or, possibly, via any other accessible allomorph (e.g., *vede*, *vet*, *vete*, *ves*).

PRESENT STUDY

By comparing the speed of responses to *savi*-type nouns, which have low stem allomorphy but are unproductive, to the two previously studied types (*lasi*- and *vesi*-type nouns), we can better understand the source of the facilitation in word recognition speed for *vesi*-type nouns, namely, as an effect of stem allomorphy or of unproductivity. The comparison of *savi*-type nouns with the other two noun classes allows us to identify a crucial variable driving word recognition speed in Finnish.

In the current experiment, the only word forms the participants saw were nominative singular forms of monomorphemic nouns. In other words, the participants did not see any inflected forms of the experimental items. We collected or calculated the values of 19 different lexical variables (see Materials section) that have been reported in the psycholinguistic literature to potentially influence word recognition speed. These measures were included as additional explanatory variables in the mixed-effects model based on a maximum likelihood method (e.g., Bates, Mächler, Bolker, & Walker, 2015a), in which logarithmically transformed reaction times were the response variable.

METHOD

Participants

We tested 31 native Finnish-speaking young adults (mean age = 25.4 years, $SD = 5.3$; 24 females) with normal or corrected-to-normal vision and with no history of neurological disorders or diagnosed language difficulties. All participants were university students, and they were compensated for their time with a movie ticket.

Materials

Ninety-nine Finnish *i*-final monomorphemic nouns were selected and divided into three sets of 33 nouns from three inflectional types (*lasi*, *savi*, and *vesi*). The

three lists were not matched for common psycholinguistic variables as the design of the experiment was not factorial.

The base and surface frequencies were extracted from the Language Bank (of Finland) corpus (<http://www.csc.fi>), which includes 131.4 million word tokens from written texts. From the same corpus, we calculated morphological family size and family frequencies (see, e.g., Schreuder & Baayen, 1995). We analyzed morphological family separately for compounds and derived words as the former category is extremely productive in Finnish, allowing speakers to form new compounds at a high rate (Niemi, 2009) due to their semantic transparency. Derived words in Finnish are more opaque, and thus, usually are more lexicalized.

Neighborhood density, as well as Hamming distance of 1 (Coltheart, Davelaar, Jonasson, & Besner, 1977), were calculated from the Basic Dictionary of Finnish (1990/1994). Phonological neighborhood density was calculated by counting the number of words with the same length but differing in the initial letter, and Hamming distance was calculated as the number of words with the same length but differing in any one letter. Since Finnish orthography–phonology mapping is isomorphic, in the present study phonological neighbors are equivalent to orthographic neighbors. Bigram frequency, initial trigram frequency, and final trigram frequency (i.e., the average number of times that all combinations of two or three subsequent letters occur in the corpus) were obtained from the Turun Sanomat Corpus (22.7 million word tokens) using a computerized search program (Laine & Virtanen, 1999). For the subjective frequency (familiarity) ratings, level of concreteness, and pictureability of test items, 16 additional participants indicated on a 6-point scale (from 0 to 5) their estimates of the target words. Participant characteristics may also influence reaction times (Baayen & Milin, 2010), so we added participants' gender and age as explanatory variables.

In addition to the 99 *i*-final monomorphemic nouns, 99 *i*-final pseudowords were created, the phonotactics of which did not violate Finnish phonology. In order to prevent participants from guessing the right answer based on pseudowords' lower bigram frequency (Grainger, Dufau, Montant, Ziegler, & Fagot, 2012), the pseudowords were matched with the target word groups for graphemic length and bigram frequency. Seventy-eight *a*-final nouns from two different inflectional types and 78 *a*-final pseudowords acted as fillers. A practice session included 30 trials with 15 words and 15 pseudowords not included in the actual experiment.

Procedure

The participants were given written instructions to decide as quickly and accurately as possible whether the letter string on the screen was a real Finnish word or not by pressing the corresponding button (*yes* for words and *no* for pseudowords) with their dominant hand. The experiment was divided into two blocks of 177 items each. The order of the items was randomized across the blocks for each participant. There were short pauses after the practice session and between the two blocks. The experiment lasted approximately 25 min. Each stimulus was

visible for 2500 ms or until a button press was made, whichever came first. Each stimulus was preceded by an asterisk in the middle of the screen for 500 ms, after which the screen was blank for 500 ms before the stimulus appeared in the same position.

Data analysis

We analyzed the data using a mixed-effects model (Bates, Mächler, Bolker, & Walker, 2015b). Our model included participants, items, and trial numbers as random intercepts and a variety of other explanatory variables as fixed effect factors. We also added possible interactions between word type and the other explanatory variables into the model. Log-transformed reaction times were the response variable.

Before analysis, we removed trials in which the participants' response was incorrect (*no* to real words or *yes* to pseudowords) or exceeded 2500 ms (5% of responses). The fastest reaction time was 440 ms, so we did not remove any short reaction times because 440 ms is an acceptable time in which to make this type of decision. Following Baayen and Milin (2010), to improve the model and remove the influence of possible outliers, we excluded data points with absolute standardized residuals exceeding 2.5 *SD* (2.5% of the data), which improved the R^2 of the model from .435 to .438. After removing outliers, all variables remained significant.

An estimate of the variance explained by the random effect of trial number was close to zero (variance = 0.0006, $SD = 0.026$), so this was removed from the model. In order to choose the best fitting model, we used a stepwise regression (step(model, direction = "both")) in the package "lmerTest" (Kuznetsova, Brockhoff, & Christensen, 2013), in which the models' p values were also calculated. Table 2 and Figure 1 were made using the "sjPlot" package (Lüdtke, 2017). For word-type effects, *vesi*-type serves as the reference for *savi*- and *lasi*-types.

RESULTS

Facilitation for *vesi*-type words was found (see Table 2), replicating Nikolaev et al. (2014). Both of the low allomorphy types, *savi* and *lasi*, showed significantly slower reaction times than *vesi*-type nouns ($t = 2.68$, $p = .009$, and $t = 3.08$, $p = .003$, respectively). Reaction times for low allomorphy types, *savi*-type nouns, which are unproductive, and *lasi*-type, which are productive, did not differ from each other.¹ Thus, the reaction times for the two low-allomorphy types differed from those for the high-allomorphy type, showing an effect of the number of allomorphs on word recognition but no significant effect for inflectional class productivity.

With the exception of five variables and one interaction term, most of the explanatory variables and their interactions with word type did not reach the level of significance in predicting word recognition response times. The effect of four variables was facilitatory: phonological/orthographic neighborhood density (marginally significant), pictureability, morphological family size (from derived words), and

Table 2. *Estimated coefficients, confidence intervals, and p values for the mixed-model fitted to the latencies elicited for i-final words in the lexical decision experiment*

	Reaction times		
	<i>B</i>	<i>CI</i>	<i>p</i>
Fixed parts			
(Intercept)	6.91	[6.81, 7.02]	< .001
Type			
<i>savi</i>	0.15	[0.04, 0.27]	.009
<i>lasi</i>	0.19	[0.07, 0.32]	.003
Familiarity rating	-0.05	[-0.07, -0.02]	< .001
Pictureability	-0.02	[-0.03, -0.00]	.012
Morph. family size derived words (log)	-0.03	[-0.06, -0.01]	.016
Morph. family freq. derived words (log)	0.02	[0.01, 0.02]	< .001
Morph. family freq. compounds (log)	-0.01	[-0.03, -0.00]	.007
Neighborhood density	-0.01	[-0.02, 0.00]	.055
Type <i>savi</i> : Familiarity rating	-0.04	[-0.07, -0.01]	.019
Type <i>lasi</i> : Familiarity rating	-0.05	[-0.08, -0.01]	.009
Random parts			
N_{Item}		99	
N_{Subject}		31	
Observations		2842	
R^2 / Ω_0^2		.438 / .438	
Akaike information criterion		-1082.774	

subjective familiarity of words. The effect of one variable was inhibitory: morphological family frequency (from derived words). The interaction between familiarity and word type (see Figure 1) reveals that a facilitation effect for word recognition for unproductive *lasi*-type words with three or more stem allomorphs was observed particularly in words of low to moderate familiarity.

Perhaps the effect of stem allomorphy is linear, so that it makes a difference to have three allomorphs instead of two, or four allomorphs instead of three. Some of the *vesi*-type words we used in this study have three allomorphs (20/33) and some have four allomorphs (13/33). Thus, we investigated this possibility by replacing our word-type variable in the model with number of allomorphs. Number of allomorphs (two, three, or four) was significant in the model. However, when we added word type (with three values: *lasi*, *savi*, and *vesi*) back into the model, the number of allomorphs variable was no longer significant. When we tried to explain the variance in reaction times for the 33 *vesi*-type words only, number of allomorphs (three or four) was not significant. Thus, although the number of allomorphs matters, it is the difference between having two or more allomorphs, but not between having three or four allomorphs, that matters most.

For some of the words the stem allomorph used in the nominative singular form is more frequent/probable than for other forms. We took this into account by

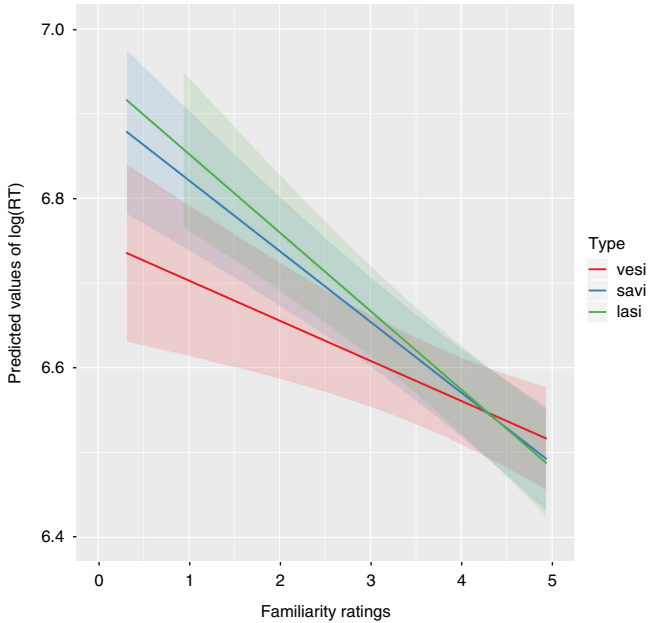


Figure 1. Interaction effect of familiarity ratings and type (*lasi*, *savi*, and *vesi*) with confidence intervals.

adding stem allomorph probability into the model (probability of the surface form of, e.g., *vesi*, divided by its lemma frequency). This variable did not reach significance in the model.

We also considered whether subtypes of each word-type category could be driving the effect. According to the Basic Dictionary of Finnish, all *lasi*-type words we used in the experiment belong to the same inflectional type (#5). This holds for *savi*-type words as well (type #7). However, the dictionary divides the *vesi*-type words we used into seven different inflectional categories. Based on Baayen's conception of statistical productivity (1994, 2001, 2003), Nikolaev and Niemi (2008) calculated two estimated indices of productivity, category-conditioned and hapax-conditioned degree of productivity, for each of these seven inflectional types (as well as for the types *lasi* and *savi*). For example, hapax-conditioned p for the inflectional type *lasi* is .243, for *savi* it is .000042, and for *vesi*-categories it ranges from .00089 to $9.78E-14$. We added these two indices of productivity into the model in order to see if a more precise measure of productivity can explain the variance better. Neither of these variables was significant in the model. However, since the indices for *vesi*-type inflectional categories were extremely small (e.g., $9.78E-14$), the variable is right-skewed. Thus, we applied a common transformation for right-skewed data, cube root ($\text{abs}(x)^{1/3}$), to these two variables. This did not change the result; transformed variables were also not significant in the model.

DISCUSSION

Effect of word type

The purpose of the current study was to investigate whether a high degree of stem allomorphy or a word's productivity status contributes to facilitation of word recognition of noninflected monomorphemic nouns in Finnish. To disentangle these variables, which are typically confounded, we compared word recognition speed for words with three or more stem allomorphs from an unproductive nominal class (*vesi*-type words) to words with fewer stem allomorphs from an unproductive class (*savi*-type words) and a productive class (*lasi*-type words). As expected, we observed faster responses for *vesi*-type words than for *lasi*-type words, replicating Nikolaev et al. (2014). The *vesi*-type words were also faster than *savi*-type words, which differ in their degree of stem allomorphy but are similarly unproductive, revealing that the speed advantage of *vesi*-type words is due to its rich stem allomorphy as opposed to the (un)productivity status of the class.

Our results contradict those of Baayen and Moscoso del Prado Martín (2005), Jaeger et al. (1996), and Tabak et al. (2005), who found facilitation for productive, regular words compared to unproductive irregular words, which were found to have higher semantic density and are thus more ambiguous. The reason for the contradiction might be due to properties of the languages studied. Irregular forms in English and Dutch tend to have few stem allomorphs (usually two) while the advantage we found in Finnish was for words that have at least three allomorphs. Thus, the irregular words in those studies might not have been able to benefit from the stem allomorphy. However, our results are consistent with those of Baayen, Wurm, and Aycok (2007), who found that semantic density was not predictive of reaction times for low-frequency words. In the current study, a significant interaction between word type and word familiarity (see Table 2 and Figure 1) demonstrated that the facilitatory effect of stem allomorphy was seen only for words with low to moderate levels of frequency/familiarity.

The function of language is to convey meaning. A greater number of stem allomorphs does not serve any facilitatory function of communicative efficiency because the meaning (lexical concept) is the same for all of the allomorphs. Greater stem allomorphy is merely a rudimentary feature of language ontogenesis (e.g., Dressler, 1985). However, the results of the present study show that the mind takes advantage of the idiosyncrasies of stem variation by retrieving the meanings of words with higher stem allomorphy faster than words with lower stem allomorphy, which may in part aid the preservation of stem allomorphy.

Our findings fit with the lemma model of lexical access during speech production (Levelt, 1989; Levelt et al., 1999). This model postulates three major subsystems of language production: between the conceptual level and the form level is a level in which lemmas are represented. A word's lemma contains the representations of the word's morphological units, in addition to the word's syntactic properties. This model was developed to account for lexical selection during production, but the theoretical architecture of the lemma model is suitable

not only for word production but also for word recognition. Word recognition does not recruit all of the stages of the lemma model, such as articulation, but the underlying architecture provides a basis for understanding word recognition effects. For example, priming effects have been reported for word roots in Arabic (Boudelaa & Marslen-Wilson, 2004), and arguments have been made about the existence of roots and word-pattern units at the level mapping lemmas to their phonological word-forms in Hebrew (Deutsch & Malinovitich, 2016). Furthermore, Whiting, Shtyrov, and Marslen-Wilson (2014) have shown that, in visual word recognition, native English speakers segment the visual input into sub-lexical units (existing free morphemes and potentially meaningful bound morphemes). These studies do not provide direct evidence for the mental representation of stem allomorphs in languages such as Finnish, however, because of the differences in their morphological architecture. Nonetheless, a growing body of research on Finnish allomorphs (Järvikivi & Niemi, 2002; Järvikivi & Pyykkönen, 2011; Nikolaev et al., 2014) suggests that stem allomorphs are separately represented in our mental lexicon.

In order to illustrate how words with three or more stem allomorphs might be recognized faster than words with two stem allomorphs, we can use the parallel grains model (Miller & Ulrich, 2003; Schröter et al., 2009) as a conceptual framework. Each stimulus (e.g., word *vesi*) activates multiple codes (grains) in parallel. These grains can be conceptualized as different types of stimulus features. In our case, the grains would be different stem allomorphs (*vesi*, *vede*, *vete*, *vet*) at the intermediate level between form decoding and the conceptual level. Each grain is assumed to present the concerted activity of a large number of neurons (cf. units in a neural network model, Rumelhart & McClelland, 1982). When the word *vesi* is read by the participant, each stem allomorph (grain) becomes activated with a certain probability. Each grain's activation is then transmitted after a random delay to a central decision center. The detection process is viewed as a race between these grains, where the average time of the winner in a race is usually shorter than the average detection time of each single process.

Words with four allomorphs did not afford additional facilitation above that seen for words with three allomorphs. This might be due to saturation in the system (i.e., a floor effect on speed of word recognition) or due to the fact that there are no low to moderate frequency nouns with four stem allomorphs. Figure 1 shows an interaction between word type and familiarity. The reaction time benefit for words with three or more stem allomorphs (type *vesi*) applies to words in the low to moderate familiarity range. More research with additional languages could be used to test whether there is a linear effect of number of stem allomorphs on word recognition speed.

Järvikivi, Bertram, and Niemi (2006) studied an effect of allomorphy versus productivity of five Finnish derivational suffixes. The authors found that structural invariance of an affix enhances decomposition of derived words in word recognition tasks. In other words, allomorphy decreases the probability of a derivational affix serving as a processing unit. The authors explain their findings in terms of the morphological transparency of a complex word, which increases

with invariance and decreases with allomorphy. In less transparent complex words, the allomorphic variants generate some kind of competition, which would require computational resources, complicating processing of complex words during the decomposition route (Järvikivi et al., 2006, p. 422). One apparent difference between Järvikivi et al.'s (2006) study and the one presented here is that their study focused on the recognition of derived words, whereas the present study presented monomorphemic words in nominative singular form. Another difference is that Järvikivi et al. (2006) focused on suffix allomorphy, whereas we focused on stem allomorphy.

Effect of psycholinguistic variables

Since there are not many studies describing the effects of lexical variables on word recognition in the Finnish language, in what follows, we discuss the additional explanatory variables that significantly predicted word recognition speed.

In addition to familiarity ratings, we included two additional variables of subjective judgment: concreteness, which is the directness with which words refer to concrete entities; and pictureability, which is the ease and speed with which words elicit mental images (Kemmerer, 2015). Concreteness and pictureability are highly correlated in our data ($r = .98$), though pictureability was a more significant predictor of reaction times in the model. In order to avoid the influence of collinearity between these two variables, we included only pictureability in the model, which turned out to be a significant predictor of word recognition.

Familiarity was another significant predictor of word recognition. The more familiar a word was rated, the faster the reaction time for that word was. However, the significant interaction between familiarity and word type shows that the effect of familiarity differed across word types. Figure 1 shows that words with low or average familiarity are recognized faster when they have a higher number of stem allomorphs (*vesi*-type), whereas words with high familiarity are recognized quickly across all word types but slightly more slowly for *vesi*-type words. This finding provides additional evidence for our hypothesis: when words are of low familiarity, words with a higher number of stem allomorphs (*vesi*-type) are generally recognized faster than words with a lower number of stem allomorphs (*savi*- and *lasi*-types). However, allomorphy does not resonate enough in the system to exert an effect when words are very familiar and hence recognized quickly.

The variables that we obtained ratings for (familiarity ratings and pictureability) were better predictors than the corpus variables (base frequency and surface frequency). This may be due to the reliance of corpora on text materials as opposed to a combination of written and spoken modalities. According to Baayen and Moscoso del Prado Martín (2005), native speakers tend to rate words that are used more often in spoken language higher than those used in written language. Another reason may be that raters tend to integrate multiple words from the same morphological family when making their judgments, which can increase familiarity ratings.

Both of the variables representing morphological family (size and frequency) that were measured in derived words were significant predictors of word recognition speed. By contrast, for compounds, only morphological family frequency, not size, was significant in the model. This suggests that a semantic variable such as morphological family should not only be measured according to its size and frequency but also be measured separately for compounds and derived words, at least in languages with rich morphology.

Two related and highly correlated variables exhibited opposite patterns of influence on word recognition: morphological family size (for derived words) being facilitatory and morphological family frequency (for derived words) being inhibitory. This pattern of findings is similar to those found by Baayen, Tweedie, and Schreuder (2002) and Pyllkänen, Feintuch, Hopkins, and Marantz (2004). One possible explanation for the inhibitory effect of high family frequency could be competition at the lemma level between a target word and its derived words with higher frequency of use.

According to more recent studies on morphological family (e.g., Mulder, Dijkstra, Schreuder, & Baayen, 2014), a secondary morphological family size measure should be included in the model as well. According to Mulder et al. (2014), the compound *horsefly* belongs to the primary family size of the noun *horse*, whereas the compound *flypaper* belongs to the secondary family size of the noun *horse*. However, in Finnish, words like *pele* “game” can occur in as many as 3,569 different compounds (the Language Bank of Finland corpus; <http://www.csc.fi>). This means that the secondary family size for the word *pele* would be in the hundreds of thousands. For the 99 *i*-final words we used in the experiment, 11 have a primary family size of more than 1,000 compounds each. In speech, we transmit information at a rate that exceeds that of any other acoustic signal (Lieberman, 2015). The system would not be very efficient if we activated hundreds of thousands of morphological family members for each word. Thus, we argue that the obtained effect of morphological family does not follow from activation of all members of the family during word recognition. Rather, the morphological family variables that we used in the model reflect activation of at least some core members of the family. Bertram, Baayen, and Schreuder (2000) have shown that the removal of loosely related members of the morphological family improves the statistical power of the morphological family variable, lending support to the idea that morphological family size represents the activation of a subset of morphological family members rather than all possible members.

We found a facilitatory effect of neighborhood density on word recognition, which was marginally significant ($t = -1.95$, $p = .055$). By contrast, Hamming distance of 1 was not statistically significant in our model. The finding that the number of neighbors differing in their first phoneme is a better predictor of lexical retrieval than the number of neighbors differing in other positions has been reported in a number of other studies (e.g., Bien, Baayen, & Levelt, 2011; Vitevitch, Armbrüster, & Chu, 2004; see also Caselli et al. 2016). Previous reports of the effect of neighborhood density on word recognition are inconsistent. Andrews (1989) found a facilitatory effect of neighborhood density for lexical recognition, whereas Luce and Pisoni (1998) claim the effect of neighborhood density is

inhibitory due to lexical competition among phonologically similar forms. Furthermore, the effect of neighborhood density depends on task demands (language comprehension vs. production) and on language (e.g., Grainger & Jacobs, 1996).

Our claim about the advantage of words with greater stem allomorphy is based on the present study and on Nikolaev et al. (2014), which both examined the mental lexicon in written-word perception, not spoken-word perception or production. Although we controlled for a number of variables known to affect word recognition speed, it is in principle possible that there are still other lexical characteristics that the results could be ascribed to. It is important, then, that additional studies attempt to investigate this issue further using a new set of stimulus words and in other languages.

Conclusion

Previous research suggested that despite a disadvantage for inflectionally unproductive words in language acquisition and language ontogenesis, they provide a benefit for word recognition, reflected in faster word recognition for words from unproductive inflectional types (Nikolaev et al., 2014). However, stem allomorphy is often confounded with productivity of the inflectional paradigm, making it difficult to determine which of these factors (productivity or stem allomorphy) leads to faster word recognition.

In order to resolve the question of whether it is (un)productivity of the inflectional class or rich stem allomorphy that facilitates word recognition, we contrasted three inflectional types: unproductive with high stem allomorphy (*vesi*-type), unproductive with low stem allomorphy (*savi*-type), and productive with low stem allomorphy (*lasi*-type). We found that the degree of stem allomorphy predicted response times, but productivity type did not. Unproductive, irregular words with three or four stem allomorphs were recognized more quickly than irregular words with two stem allomorphs. We propose that multiple stem allomorphs are activated in parallel at the lemma level, a level between form decoding and lexical selection. This implies that words with rich stem allomorphy involve a larger neural network. The word recognition process is viewed as a race between these allomorphs, where the average time of the winner in the race is usually shorter than the average word recognition time of each single process.

ACKNOWLEDGMENTS

We thank Raymond Bertram and an anonymous reviewer for their constructive comments and suggestions. Alexandre Nikolaev was supported by the Department of Foreign Languages and Translational Studies at the University of Eastern Finland. Minna Lehtonen was supported by the Academy of Finland (Grant 288880).

NOTE

1. In an alternative model, we changed the reference from *vesi*-type to *lasi*-type; in this model, $t = -2.86$ ($p = .005$) for *vesi*, and $t = -0.87$ ($p = .38$) for *savi*.

REFERENCES

- Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 802–814.
- Baayen, R. H. (1994). Productivity in language production. *Language and Cognitive Processes*, *9*, 447–469.
- Baayen, R. H. (2001). *Word frequency distributions*. Dordrecht: Kluwer.
- Baayen, R. H. (2003). Probabilistic approaches to morphology. In R. Bod, J. B. Hay & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 229–287). Cambridge, MA: MIT Press.
- Baayen, R. H. (2009). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 899–919). Berlin: Mouton de Gruyter.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28.
- Baayen, R. H., & Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, *81*, 666–698.
- Baayen, R. H., Tweedie, F. J., & Schreuder, R. (2002). The subject as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain and Language*, *81*, 55–65.
- Baayen, R. H., Wurm, L., & Aycocock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *Mental Lexicon*, *2*, 419–463.
- Basic Dictionary of Finnish *Suomen kielen perussanakirja* (1990–1994). Helsinki: Edita Oyj.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015a). lme4: Linear mixed-effects models using eigen and S4 (R package version 1.1-10). Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*, 390–405.
- 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*, 213–226.
- Bien, H., Baayen, R. H., & Levelt, W. J. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, *26*, 683–715.
- Boudelaa, S., & Marslen-Wilson, W. D. (2004). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, *90*, 106–116.
- Caselli, N. K., Caselli, M. K., & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology*, *69*, 432–454.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornick (Ed.), *Attention and performance* (Vol. 6, pp. 535–556). Hillsdale, NJ: Erlbaum.
- Deutsch, A., & Malinovitch, T. (2016). The role of the morpho-phonological word-pattern unit in single-word production in Hebrew. *Journal of Memory and Language*, *87*, 1–15.
- Dressler, W. (1985). *Morphonology*. Ann Arbor, MI: Karoma.
- Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1991* (pp. 165–183). Dordrecht: Kluwer Academic.
- Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (*Papio papio*). *Science*, *336*, 245–248.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.

- Jaeger, J., Lockwood, A., Kemmerer, D., Van Valin, R., Murphy, B., & Khalak, H. (1996). A positron emission tomography study of regular and irregular verb morphology in English. *Language*, 72, 451–497.
- Järvikivi, J., Bertram, R., & Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21, 394–431.
- Järvikivi, J., & Niemi, J. (2002). Form-based representation in the mental lexicon: Priming (with bound stem allomorphs in Finnish. *Brain and Language*, 81, 412–423.
- Järvikivi, J., & Pyykkönen, P. (2011). Sub- and supralephical information in early phases of lexical access. *Frontiers in Psychology*, 2, 1–11.
- Ji, H., Gagné, C. L., & Spalding, T. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, 65, 406–430.
- Kemmerer, D. (2015). *Cognitive neuroscience of language*. New York: Psychology Press.
- Kim, J., Marcus, G., Pinker, S., Hollander, M., & Coppola, M. (1994). Sensitivity of children's inflection to grammatical structure. *Journal of Child Language*, 21, 173–209.
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). (R Package Version 2.0–33). Retrieved from <https://CRAN.R-project.org/package=lmerTest>
- Laine, M., & Virtanen, P. (1999). WordMill Lexical Search Program. Centre for Cognitive Neuroscience, University of Turku.
- Lehtonen, M., Cunillera, T., Rodríguez-Fornells, A., Hultén, A., Tuomainen, J., & Laine, M. (2007). Recognition of morphologically complex words in Finnish: Evidence from event-related potentials. *Brain Research*, 1148, 123–137.
- Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in mono- and bilinguals. *Bilingualism: Language and Cognition*, 6, 213–225.
- 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, 121–146.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences*, 98, 13464–13471.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). Multiple perspectives on word production. *Behavioral and Brain Sciences*, 22, 61–69.
- Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., & Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 713–716.
- Lieberman, P. (2015). Evolution of language. In M. P. Muehlenbein (Ed.), *Basics in human evolution* (pp. 493–503). New York: Academic Press.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Lüdtke, D. (2017). sjPlot: Data visualization for statistics in social science (R package version 2.3.1). Retrieved from <https://CRAN.R-project.org/package=sjPlot>
- Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PLOS ONE*, 12, e0171935. doi: 10.1371/journal.pone.0171935.
- Miller, J., & Ulrich, R. (2003). Simple reaction time and statistical facilitation: A parallel grains model. *Cognitive Psychology*, 46, 101–151.
- Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, R. H. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, 72, 59–84.
- Niemi, J. (2009). Compounds in Finnish. *Lingue e Linguaggio*, 2, 233–252.

- Nikolaev, A. (2002). Eräiden suomen kielen taivutustyyppien produktiivisuudesta [Productivity of Finnish inflectional types]. *Puhe ja kieli*, 22, 113–124.
- Nikolaev, A., & Niemi, J. (2008). Suomen nominien taivutusjärjestelmän produktiivisuuden indekseistä [Indices of productivity in Finnish inflection]. *Virittäjä*, 112, 518–544.
- Nikolaev, A., Pääkkönen, A., Niemi, J., Nissi, M. J., Niskanen, E., Könönen, M., ... Soininen, H. (2014). Behavioural and ERP effects of paradigm complexity on visual word recognition. *Language, Cognition and Neuroscience*, 10, 1295–1310.
- Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition*, 91, B35–B45.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24, 574–590.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.
- Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–157). Hillsdale, NJ: Erlbaum.
- Schröter, H., Frei, L. S., Ulrich, R., & Miller, J. (2009). The auditory redundant signals effect: An influence of number of stimuli or number of percepts? *Attention, Perception, & Psychophysics*, 71, 1375–1384.
- Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues* (pp. 1157–1256). Hillsdale, NJ: Erlbaum.
- Soveri, A., Lehtonen, M., & Laine, M. (2007). Word frequency and morphological processing revisited. *Mental Lexicon*, 2, 359–385.
- Tabak, W. M., Schreuder, R., & Baayen, R. H. (2005). Lexical statistics and lexical processing: Semantic density, information complexity, sex, and irregularity in Dutch. In S. Kepser & M. Reis (Eds.), *Linguistic evidence* (pp. 529–555). Berlin: Mouton de Gruyter.
- Taft, M. (2003). Morphological representation as a correlation between form and meaning. In E. Assink & D. Sandra (Eds.), *Reading complex words* (pp. 113–137). Amsterdam: Kluwer.
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, 57A, 745–765.
- Taft, M., & Kougious, P. (2004). The processing of morpheme-like units in monomorphemic words. *Brain and Language*, 90, 9–16.
- Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 514–529.
- Whiting, C., Shtyrov, Y., & Marslen-Wilson, W. (2014). Real-time functional architecture of visual word recognition. *Journal of Cognitive Neuroscience*, 27, 246–265.