

Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing

Kaidi Lõo

University of Alberta, Canada

Juhani Järvikivi

University of Alberta, Canada

R. Harald Baayen

University of Tübingen, Germany and University of Alberta, Canada

Word count: 3298

Version: December 7, 2017

Corresponding author:

Kaidi Lõo

Department of Linguistics

4-32 Assiniboia Hall, University of Alberta

Edmonton, AB T6G 2E7, Canada

e-mail: kloo@ualberta.ca

Abstract

1 Estonian is a morphologically rich Finno-Ugric language with nominal paradigms that have at least 28 dif-
2 ferent inflected forms but sometimes more than 40. For languages with rich inflection, it has been argued
3 that whole-word frequency, as a diagnostic of whole-word representations, should not be predictive for lexical
4 processing. We report a lexical decision experiment, showing that response latencies decrease both with
5 frequency of the inflected form and its inflectional paradigm size. Inflectional paradigm size was also pre-
6 dictive of semantic categorization, indicating it is a semantic effect, similar to the morphological family size
7 effect. These findings fit well with the evidence for frequency effects of word n-grams in languages with little
8 inflectional morphology, such as English. Apparently, the amount of information on word use in the mental
9 lexicon is substantially larger than was previously thought.

10

11 *Keywords:* whole-word frequency; inflectional paradigm size; Estonian; inflectional morphology; semantic
12 categorization; Generalized Additive Mixed-Models

1 Introduction

Estonian is a Finno-Ugric language with remarkably productive morphology. A 15-million word corpus of Estonian¹ contains no less than 790,957 different words, a number similar to the total number of different words (794,771) in a 100-million word corpus of British English (Leech, Rayson & Wilson, 2014). This raises the question of how a speaker of Estonian can understand such a large number of different forms, as the probability of encountering an out-of-vocabulary word, i.e., a word the speaker has not yet seen or heard, is no less than 0.64.

However, the problem might not be as severe as it may seem, as out-of-vocabulary words are typically morphologically complex. In fact, roughly 95% of the forms in our corpus have morphological structure, including derived words (e.g., *töötaja* ‘worker’) and compounds (e.g., *käsitöö* ‘handwork’). Derived words and compounds, built on the same stem, cluster into morphological families. Inflected forms (e.g., *tööd* ‘works’, *töös* ‘at work’, *tööga* ‘with the work’), in turn, cluster into inflectional paradigms. Inflectional paradigms typically come with a few inflected variants, known as the principal parts, from which all other forms in the paradigm can be predicted (Blevins, 2006). Thus, the number of forms that one must know by heart is much smaller than the number of forms that one can understand or produce, given these basic forms and the rules of the language.

Several studies have argued that for morphologically rich languages, such as Finnish and Turkish, storing all word forms in a mental dictionary would exceed the storage capacity of the brain (Hankamer, 1989; Niemi, Laine & Tuominen, 1994; Yang, 2016). Crucially, mental dictionaries with only stored forms would not be able to deal with the large numbers of out-of-vocabulary words. Therefore, algorithms must be available for interpreting and producing novel complex words, both in natural language processing systems and in the human cognitive system (Hankamer, 1989; Kaalep, 1997; Karlsson & Koskenniemi, 1985; Sproat, 1992).

Although morphological decomposition has been argued to play a fundamental role in languages with simple morphologies, such as English (Fruchter & Marantz, 2015; Rastle, Davis & New, 2004; Taft, 1994; Taft & Forster, 1975), it seems especially attractive for languages with rich morphology, such as Estonian that minimization of storage and maximization of rule-driven computation (Pinker, 1999).

However, even though Estonian appears to be a prime candidate for a language dominated by rule-driven processing, recent findings place Estonian morphology in a different light. For languages as diverse as English and Mandarin, experimental evidence is accumulating that the frequency of occurrence of sequences of multiple words (e.g., *the president of the*) predicts a range of measures of lexical processing when other predictors, such as word frequency and length, are statistically controlled (Arnon & Snider, 2010; Janssen &

¹<http://www.cl.ut.ee/korpused/grammatikakorpus/> (15.01.2017)

44 Barber, 2012; Sun, 2016; Tremblay, Derwing, Libben & Westbury, 2011; Tremblay & Tucker, 2011). These
45 frequency effects have not only been found in studies with adults, but also in studies with children (Ambridge,
46 Kidd, Rowland & Theakston, 2015; Bannard & Matthews, 2008; Kidd, 2012), and second language learners
47 (Siyanova-Chanturia, Conklin & Van Heuven, 2011; Sonbul, 2015; Wolter & Gyllstad, 2013). Importantly,
48 sequences of words in English, such as *into to the house*, translate into Estonian with a single inflected form,
49 such as *majasse*. In the light of these frequency effects for English, we predict a similar frequency effect for
50 functional equivalence in Estonian.

51 Given the frequency effects for word sequences, it is unsurprising that whole-word frequency effects in the
52 processing of regular complex words are also attested (Dutch: Baayen, Dijkstra & Schreuder 1997; Baayen,
53 McQueen, Dijkstra & Schreuder 2003; Kuperman, Schreuder, Bertram & Baayen 2009; English: Baayen,
54 Kuperman & Bertram 2010; Baayen, Wurm & Aycock 2007; Vietnamese Pham & Baayen 2015; and Danish:
55 Balling & Baayen 2011). For Finnish, a Finno-Ugric language closely related language to Estonian, whole-
56 word frequency effects have been found for derived words, but not for most inflected forms (Bertram, Laine,
57 Baayen, Schreuder & Hyönä, 1999; Laine, Vainio & Hyönä, 1999; Niemi, Laine & Tuominen, 1994; Soveri,
58 Lehtonen & Laine, 2007; Vannest, Bertram, Järvikivi & Niemi, 2002). One reason may be that inflection
59 typically serves syntactic functions, such as grammatical role, number marking and agreement, whereas
60 derivation and compounding tend to result in the formation of new words (see e.g., Booij 2006, and for
61 detailed discussion that the distinction between inflection and word formation is not an absolute one, Booij
62 1993). However, a problem with previous studies on inflected forms in Finnish is the small number of subjects
63 and items as well as the concomitant lack of power (Westfall, Kenny & Judd, 2014). Hence, the first research
64 goal of the present study was to re-examine whole-word frequency effects for inflected words in Estonian
65 using a large regression design with thousands of items.

66 The consequences of the size of a word's morphological family (i.e., the count of derived words and
67 compounds sharing a constituent, e.g., *worker*, *workforce*, *handwork*, while excluding inflectional variants
68 from the counts) for lexical processing have been studied extensively (Bertram, Baayen & Schreuder, 2000;
69 De Jong, Schreuder & Baayen, 2000; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder & Baayen,
70 2004; Schreuder & Baayen, 1997). Words with larger families are processed faster, and this phenomenon has
71 been explained in two ways. Within the framework of interactive activation (De Jong, Schreuder & Baayen,
72 2003), words from larger families receive more activation from their family members, resulting in a critical
73 threshold activation level being reached earlier in time. According to learning models (e.g., Baayen, Milin,
74 Filipović Đurđević, Hendrix & Marelli 2011), as long as complex forms share some element of meaning, that
75 element will be strengthened for all the family members each time it is encountered, allowing for a faster
76 reaction time for words with larger families. The morphological family size effect is generally understood

77 as a semantic effect, as it appears to be driven primarily by semantically transparent family members or
 78 semantically relevant subsets of family members (Moscoso del Prado Martín et al., 2004; Mulder, Dijkstra,
 79 Schreuder & Baayen, 2014). As semantic transparency is greater for inflection as compared to derivation and
 80 compounding, an effect of inflectional paradigm size should be detectable for languages with large inflectional
 81 paradigms.

82 Only a few studies have looked at the role of inflectional paradigms in lexical processing. Moscoso del
 83 Prado Martín, Kostić & Baayen (2004) studied the processing consequences of inflectional paradigms in
 84 English and Dutch, using summary measures characterizing the probability distribution of inflectional vari-
 85 ants. Specifically inflectional entropy and the Kulback-Leibler divergence have been found to predict the
 86 consequences of inflectional paradigmatic relations in the lexical decision task (Milin, Filipović Durđević &
 87 Moscoso del Prado Martín 2009, see also Baayen et al. 2011 for prepositional entropy effects for English).

88 The size of Estonian nominal paradigms offers further opportunities for investigating the consequences of
 89 paradigm complexity. Estonian nominal paradigms are characterized by 14 cases in both singular and plural,
 90 but may have several additional parallel forms. However, in practice, most words are actually not used in
 91 all their cases and numbers. For example, for *jalg* ‘foot, leg’ out 46 grammatically possible forms only 36
 92 inflected forms are present in the Balanced Corpus of Estonian (Table 1).

Table 1: Inflectional paradigm of *jalg* ‘foot, leg’ with 46 paradigm members. The 36 paradigm members present in the corpus are marked in bold.

Case	Singular	Plural	English translation
Nominative	jalg	jalad	foot (subject)
Genitive	jala	jalgade, jalge	of a foot; foot (as a total object)
Partitive	jalga	jalgasid, jalgu	foot (as a partial object)
Illative-1	jalga	-	into a foot
Illative-2	jalasse	jalgadesse, jalusse, jalgesse	into a foot
Inessive	jalas	jalgades, jalus, jalges	in a foot
Elative	jalast	jalgadest, jalust, jalgest	from a foot
Allative	jalale	jalgadele, jalule, jalgele	onto a foot
Adessive	jalal	jalgadel, jalul, jalgel	on a foot
Ablative	jalalt	jalgadelt, jalult, jalgelt	from a foot
Translative	jalaks	jalgadeks, jaluks, jalgeks	[to turn] into a foot
Terminative	jalani	jalgadeni, jalgeni	up to a foot
Essive	jalana	jaladena	as a foot
Abessive	jalata	jalgadeta	without a foot
Comitative	jalaga	jalgadega	with a foot

93

94 Karlsson (1986) made a similar observation in a Finnish corpus study, and pointed out that even though
 95 in theory a word can appear in any of the inflected forms defined by grammar, only a subset of the possible
 96 forms actually occur. Figure 1 illustrates this point for Estonian. Due to a large number of words that occur
 97 only once, most of which are the result of productive compounding, many Estonian nouns are attested with

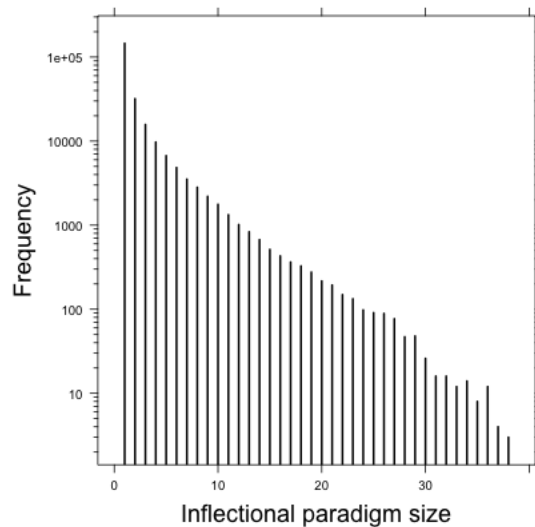


Figure 1: Histogram of 231,891 noun paradigms in the Balanced Corpus of Estonian, 75% of which are paradigms of compound. The x-axis represents inflectional paradigm size, the y-axis shows the frequency of a particular paradigm size. Small inflectional paradigms are more common than large inflectional paradigms.

98 only one inflected case form. However, there are also nouns that have well-filled paradigms, that, including
 99 variant forms, can comprise up to 38 different forms.

100 [Karlsson \(1986\)](#) argues that which forms are attested, depends on the meaning of the word. For example,
 101 *kesä* ‘summer’ primarily has a temporal meaning and therefore, the most frequent inflected form is adessive
 102 *kesällä* ‘in the summer’, whereas other forms are used less or not used at all. Likewise, although in theory
 103 the number of paradigm members for Estonian nouns can be quite high, in practice, not all possible inflected
 104 forms are semantically felicitous. This observation brings us to our second research goal, namely whether the
 105 actual size of Estonian nominal paradigms has consequences for lexical processing. Specifically, does a large
 106 inflectional paradigm size facilitate processing, just as large morphological families do? Experiment 1 makes
 107 use of a visual lexical decision task to address these issues.

108 2 Experiment 1: Lexical decision task

109 2.1 Participants

110 Twenty-four native speakers of Estonian (14 females; 24-67 years, mean 43.69) with normal or corrected-to-
 111 normal vision completed the experiment. They received 10 euros for their participation.

2.2 Materials

1,000 nouns were retrieved from the Balanced Corpus of Estonian. For each noun, three different case-inflected forms were selected (equal number of nominative, genitive, partitive; inessive, allative, ablative; translative, essive, comitative singular, and nominative and two variants of plural illative case). The whole-word frequency of the stimuli ranged between 1 and 3,402 occurrences per million (median 4). Length in letters varied between 3 and 21 characters (mean 7 characters), and inflectional paradigm size ranged between 2 and 36 (median 19).

From the list of 3,000 words, twelve experimental lists were created, each contained 250 words in randomized order. Each list was enlarged with 250 inflected nonwords, created by changing one or two letters of existing Estonian inflected forms while respecting Estonian orthophonotactics. Each subject was presented with 500 items. Across the experiment, each inflected form received two responses, and each lemma six responses.

2.3 Procedure

We used a standard visual lexical decision task. Stimuli were presented on a 15-inch Dell laptop in 18 point Courier New Bold font on a white background, using E-Prime 2.0 experimental software (Psychology Software Tools). Each trial started with the visual presentation of a blank screen for 1,000 ms, after which a fixation point appeared in the centre of the screen for 750 ms, followed by the target stimulus in the same position for a maximum of 2,500 ms. After responding or after time-out, the stimulus disappeared and a new trial was triggered. The experiment was divided into five blocks with 100 trials each. The first experimental block was preceded by 20 practice trials. A break followed each block and lasted until the participant was ready to continue. The entire experimental session lasted approximately 45 minutes.

2.4 Analysis and Results

Nonword trials and trials with response times less than 50 ms were removed prior to analysis, trials with incorrect responses were removed from response time analysis (approximately 13 % of the trials, i.e., 813 data points.). Mean reaction time was 863 ms and standard deviation 332 ms in the response time analysis, mean error rate was 7% in the accuracy analysis. We analysed the data with the generalized additive mixed model (GAMM, Wood 2006; Baayen, Vasishth, Bates & Kliegl 2017) using the *R*-package *mgcv* (version 1.8-12), inversed-transformed reaction time (-1000/RT) served as the response variable and inflectional paradigm size (inflectional variants of a paradigm attested in the corpus), whole-word frequency, lemma frequency (the summed frequency of a noun's inflectional variants), word length and participant's age in years as predictors.

142 To avoid outlier effects, frequency was log-transformed. For visualization, we made use of the *itsadug* package
143 (version 2.2, van Rij, Baayen, Wieling & van Rijn, 2016).

144 Reaction times increased linearly with age ($\hat{\beta} = 0.0077, t(5130) = 3.2921, p = 0.001$). Reaction times
145 decreased nonlinearly for increasing inflectional paradigm size ($F = 6.558, edf = 1.6554, p = 0.0013$). Whole-
146 word frequency and word length entered into a non-linear interaction that was modelled with a tensor product
147 smooth ($F = 46.6386, edf = 4.7394, p < 0.0001$, see Baayen et al. 2017 for an introduction to the generalized
148 additive model). As expected, reaction times decreased with frequency and increased with length. The
149 interaction indicated that the frequency effect was slightly larger for the shortest words.

150 Whole-word frequency was correlated with lemma frequency ($r = 0.43$). However, replacing whole-word
151 frequency by lemma frequency, resulted in a model with a substantially worse fit (increase in AIC 147.27),
152 and adding lemma frequency to the model with whole-word frequency revealed lemma frequency not to be
153 significant when whole-word frequency was present as a predictor.

154 The model included by-subject factor smooths for trial ($F = 778.5655, edf = 136.1335, p < 0.001$), as well
155 as by-subject random slopes for length ($F = 11.8300, edf = 20.2773, p < 0.0001$), frequency ($F = 2.6188, edf =$
156 $13.7509, p < 0.001$), and paradigm size ($F = 1.0644, edf = 10.0954, p < 0.0063$). The model also included
157 by-item random intercepts ($F = 0.2316, edf = 473.2887, p < 0.0001$). Further random effects and interactions
158 did not reach significance.

159 A logistic GAMM fitted to the accuracy data supported the conclusions drawn from the analysis of reaction
160 times. The probability of an error decreased with inflectional paradigm size ($\chi^2 = 17.3905, edf = 2.7258, p =$
161 0.0010) and with age ($\hat{\beta} = 0.0206, z(5945) = 2.8442, p = 0.0045$). Whole-word frequency and length showed
162 a strongly non-linear interaction (see Figure 2, $\chi^2 = 46.6386, edf = 4.7394, p < 0.0001$): accuracy increased
163 with whole-word frequency and word length, but was especially high for long words of intermediate frequency.
164 The random effect structure was the same as for the reaction times (all $p < 0.05$). The data and the code
165 for statistical analyses of Experiment 1 are available on Open Science Framework ([dataset] Lõo, 2017).

166 Given that inflectional paradigm size reflects semantic richness, the number of contexts a word can be used
167 in, we expect it would also persist in a task such as semantic categorization (Experiment 2). For example, a
168 word such as *jalg* ‘foot, leg’ is semantically rich as it is used to express several different affordances, such as
169 *jalal* ‘on a foot’, *jalaga* ‘with a foot’ or *jalata* ‘without a foot’, whereas *kuvöös* ‘incubator’ is semantically poor
170 and can be used only in the nominative base form or to express location, as in *kuvöösis* ‘in an incubator’.
171 Therefore, words with larger inflectional paradigm size offer participants more information to base their
172 decision on in a semantic categorization task, for instance, when deciding whether a word is animate or
173 inanimate. Whether a noun refers to an animate or inanimate object is fairly easy to decide, which is why
174 we used this particular binary opposition (see Andrews & Heathcote 2001; Balota & Chumbley 1984; Forster

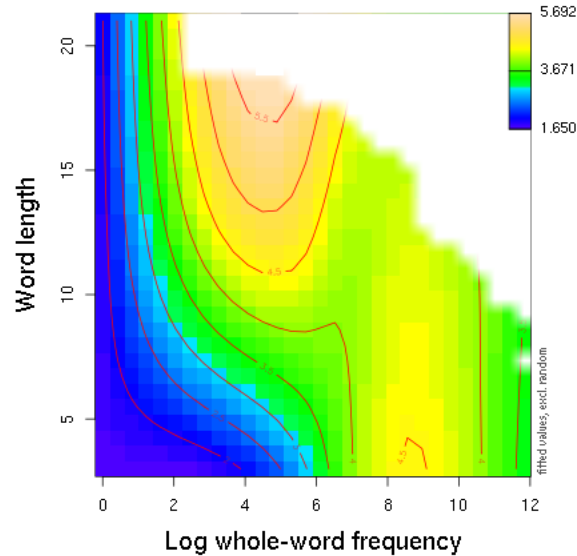


Figure 2: Tensor product smooth for the interaction of word length by word frequency. Colour coding is used to represent model predictions, with darker shades of yellow indicating higher log-odds for correct responses, and darker shades of blue representing lower log-odds ratios.

175 & Shen 1996 for other similar studies). Although we expect inflectional paradigm size effect to arise, a
 176 whole-word frequency effect may well be absent in this task as it gauges the probability of a word, but not
 177 the probability of one of the many features of a given word, namely, that it refers to an animate or inanimate
 178 entity.

179 3 Experiment 2: Semantic categorization task

180 3.1 Participants

181 26 native speakers of Estonian (18 female; age 21-67 years, mean 38.66) with normal or corrected-to-normal
 182 vision were recruited. They received 10 euros for their participation.

183 3.2 Materials

184 200 case-inflected nouns were selected from the Balanced Corpus of Estonian, 100 animate and 100 inanimate
 185 nouns. The stimuli were selected in such a way that the correlation between the inflectional paradigm size
 186 and whole-word frequency was low ($r = 0.3$). However, the correlation between inflectional paradigm size
 187 and lemma frequency remained high ($r = 0.8$), whole-word frequency ranged between 1 and 213 per million
 188 (median 5), length in letters varied between 4 and 15 characters (mean 8 characters), inflectional paradigm
 189 size ranged between 2 and 36 (median 22). The same 200 items were presented to all participants in a

190 randomized order.

191 3.3 Procedure

192 Participants classified stimuli as animate or inanimate by pressing the relevant computer key. Stimuli were
193 presented on a 17-inch Dell computer screen in 26 point Courier New Bold font on a light grey background,
194 using ExperimentBuilder software (SR Research Ltd). Each trial started with the visual presentation of a
195 blank screen for 1,000 ms, followed by a fixation cross for 500 ms, after which the word appeared at the
196 center of the screen for 2,500 ms or until a decision was made. The experiment started with six practice
197 trials, followed by 200 experimental trials. The experiment took approximately 20 minutes.

198 3.4 Analysis and Results

199 Trials with response times less than 50 ms were removed prior to analysis, trials with incorrect responses were
200 removed from response time analysis (approximately 12 % of the trials, i.e., 589 data points). Mean reaction
201 time was 811 ms and standard deviation 292 ms in the response time analysis, mean error rate was 10 % is the
202 accuracy analysis. A GAMM with log-transformed reaction time as response variable revealed that inanimate
203 nouns were responded to more slowly than animate nouns ($\hat{\beta} = 0.0528, t(4527) = 2.4057, p = 0.0162$).
204 Reaction times increased linearly with age ($\hat{\beta} = 0.0074, t(4527) = 3.3253, p = 0.0009$), and nonlinearly
205 with word length ($F = 5,796, edf = 3.1301, p = 0.0007$), and decreased with inflectional paradigm size
206 ($F = 10.9108, edf = 2.6432, p < 0.0001$). As shown in Figure 3, the effect of paradigm size is restricted to the
207 smaller paradigm sizes. Effects of word frequency or lemma frequency were not significant.

208 The model included by-subject factor smooths for trial ($F = 215.8889, edf = 120.8497, p < 0.0001$), as
209 well as by-subject random slopes for length ($F = 3.2682, edf = 17.6420, p < 0.0001$) and condition ($F =$
210 $2.9903, edf = 21.5936, p < 0.0001$). The model also included by-item random intercepts ($F = 3.3404, edf =$
211 $146.6001, p < 0.0001$). Further random effects and interactions did not reach significance.

212 A logistic GAMM fitted to the accuracy data did not support the effects of order, age, and length.
213 However, a larger paradigm size predicted fewer errors ($\hat{\beta} = 0.0313, z(5058) = 2.1712, p = 0.0299$), and
214 inanimate nouns elicited less errors ($\hat{\beta} = 0.9850, z(5058) = 3.0472, p = 0.0023$). The random effect structure
215 was the same as for the reaction times (all $p < 0.001$). The data and the code for statistical analyses of
216 Experiment 2 are available on Open Science Platform ([dataset] Lõo, 2017).

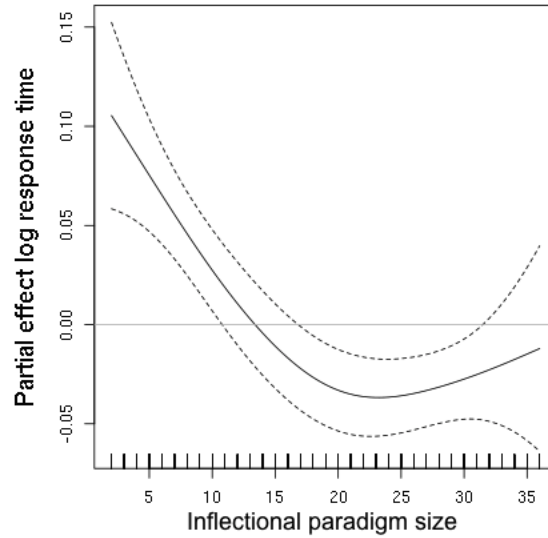


Figure 3: The partial effect for inflectional paradigm size in semantic categorization task. Increasing inflectional paradigm size decreases response times. The effect disappears for larger paradigm sizes.

4 General Discussion

Even though Estonian has highly productive and complex inflectional morphology, making it a prime candidate for decompositional processing, Experiment 1 showed that whole-word frequency was a better predictor than lemma frequency. This dovetails with our second finding in both experiments that inflectional paradigm size also predicted reaction times.

Are these effects due to irregularities in Estonian paradigms? Estonian has not only agglutinative inflected forms, but also fusional forms. For example, in the partitive plural form *kalu* ‘fish’ a suffix *-u* stands for both partitive and plural, whereas in the inessive plural *kalades* ‘in the fishes’, one can distinguish a plural suffix *-de* and an inessive suffix *-s*. Further, case-inflected forms are often used as lexicalized adverbs, e.g., *käes* ‘there’, literally ‘in the hand’. Furthermore, many paradigm members have alternative parallel forms, which may express subtly different meanings. For example, *kalasid* and *kalu* are both plural partitive forms of *kala* ‘fish’, with the same meaning, whereas *jalgadel*, *jalul* and *jalgel* are all adessive plurals of *jalg* ‘foot’ and vary slightly in meaning. *Jalgadel* has an external locational meaning (something is *on the feet*), whereas *jalul* and *jalgel* translate as ‘back on the feet’. Nevertheless, regular inflected forms are in the majority.

Since irregularity does not provide a full explanation of the present frequency and paradigm effects, the question remains of how to account for these effects in current models of the mental lexicon. One possibility is offered by dual-route models, which hold that forms are stored but rules are also available (Baayen et al., 1997). If so, human memory is capable of storing much more information than previously

235 assumed (Hankamer, 1989; Niemi et al., 1994; Yang, 2016). Alternatively, we may ask whether it is fruitful to
236 discuss these issues in terms of rules and representations. In learning-driven computational models of lexical
237 processing (Baayen et al., 2011; Seidenberg & Gonnerman, 2000), frequency effects, as well as paradigmatic
238 effects, can arise without representations for whole words. Thus, frequency effects for Estonian inflected
239 forms, just as frequency effects for English word n-grams, require rethinking the traditional Bloomfieldian
240 division of labour between storage and computation.

241 5 Acknowledgements

242 We would like to thank Abigail Toth, Fabian Tomaschek and two anonymous reviewers for their comments
243 on an earlier version of this paper. This research was supported by the national scholarship program Krist-
244 jan Jaak, funded and managed by Archimedes Foundation in collaboration with the Estonian Ministry of
245 Education and Research and by an Alexander von Humboldt research award (grant number 1141527) to the
246 first and last author, respectively.

247 References

- 248 Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in
249 first language acquisition. *Journal of Child Language*, *42*(02), 239–273.
- 250 Andrews, S. & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identifica-
251 tion: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*,
252 *27*(2), 514.
- 253 Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of*
254 *Memory and Language*, *62*(1), 67–82.
- 255 Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel
256 dual route model. *Journal of Memory and Language*, *36*, 94–117.
- 257 Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. In S. Scalise
258 & I. Vogel (Eds.), *Compounding*. Amsterdam/Philadelphia: Benjamins.
- 259 Baayen, R. H., McQueen, J., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional
260 morphology: Revisiting Dutch plurals. In R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in*
261 *language processing* (pp. 355–390). Berlin: Mouton de Gruyter.

- 262 Baayen, R. H., Milin, P., Filipović Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model
263 for morphological processing in visual comprehension based on naive discriminative learning. *Psychological*
264 *Review*, *118*(3), 438–481.
- 265 Baayen, R. H., Vasishth, S., Bates, D., & Kliegl, R. (2017). The cave of shadows. addressing the human
266 factor with generalized additive mixed models. *Journal of Memory and Language*, *56*, 206–234.
- 267 Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words. a
268 regression study across tasks and modalities. *The Mental Lexicon*, *2*, 419–463.
- 269 Balling, L. & Baayen, R. H. (2011). Probability and surprisal in auditory comprehension of morphologically
270 complex words. Manuscript submitted for publication.
- 271 Balota, D. A. & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? the role of
272 word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception*
273 *and performance*, *10*(3), 340.
- 274 Bannard, C. & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity
275 on children’s repetition of four-word combinations. *Psychological Science*, *19*, 241–248.
- 276 Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of*
277 *Memory and Language*, *42*, 390–405.
- 278 Bertram, R., Laine, M., Baayen, R. H., Schreuder, R., & Hyönä, J. (1999). Affixal homonymy triggers full-
279 form storage even with inflected words, even in a morphologically rich language. *Cognition*, *74*, B13–B25.
- 280 Blevins, J. P. (2006). English inflection and derivation. In B. Aarts & A. M. McMahon (Eds.), *Handbook of*
281 *English Linguistics* (pp. 507–536). London: Blackwell.
- 282 Booij, G. E. (1993). Against split morphology. In G. E. Booij & J. v. Marle (Eds.), *Yearbook of Morphology*
283 *1993* (pp. 27–50). Dordrecht: Kluwer Academic Publishers.
- 284 Booij, G. E. (2006). Inflection and Derivation. In *Encyclopedia of Language & Linguistics, Second Edition*,
285 volume 5 (pp. 654–661).
- 286 [dataset] Lõo, K. (2017). Estonian lexical processing data. Available from Open Science Framework:
287 osf.io/grbdx.
- 288 De Jong, N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology.
289 *Language and Cognitive Processes*, *15*, 329–365.

- 290 De Jong, N. H., Schreuder, R., & Baayen, R. H. (2003). Morphological resonance in the mental lexicon. In
291 R. H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 65–88). Berlin:
292 Mouton de Gruyter.
- 293 Forster, K. & Shen, D. (1996). No enemies in the neighborhood: absence of inhibitory neighborhood effects
294 in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory,*
295 *and Cognition, 22*, 696–713.
- 296 Fruchter, J. & Marantz, A. (2015). Decomposition, lookup, and recombination: Meg evidence for the full
297 decomposition model of complex visual word recognition. *Brain and Language, 143*, 81–96.
- 298 Hankamer, J. (1989). Morphological parsing and the lexicon. In Marslen-Wilson, W. (Ed.), *Lexical Repre-*
299 *sentation and Process*, (pp. 392–408)., Cambridge, MA, USA. MIT Press.
- 300 Janssen, N. & Barber, H. A. (2012). Phrase frequency effects in language production. *PloS one, 7*(3), e33202.
- 301 Kaalep, H.-J. (1997). An Estonian morphological analyser and the impact of a corpus on its development.
302 *Computers and the Humanities, 31*(2), 115–133.
- 303 Karlsson, F. (1986). Frequency considerations in morphology. *STUF-Language Typology and Universals,*
304 *39*(1-4), 19–28.
- 305 Karlsson, F. & Koskeniemi, K. (1985). A process model of morphology and lexicon. *Folia Linguistica, 29,*
306 207–231.
- 307 Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Develop-*
308 *mental Psychology, 48*(1), 171.
- 309 Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading of multimorphemic Dutch
310 compounds: Towards a multiple route model of lexical processing. *Journal of Experimental Psychology:*
311 *HPP, 35*, 876–895.
- 312 Laine, M., Vainio, S., & Hyönä, J. (1999). Lexical access routes to nouns in a morphologically rich language.
313 *Journal of Memory and Language, 40*(1), 109 – 135.
- 314 Leech, G., Rayson, P., & Wilson, A. (2014). *Word frequencies in written and spoken English: Based on the*
315 *British National Corpus*. Routledge.
- 316 Milin, P., Filipović Durdević, D., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of
317 inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and*
318 *Language, 50–64*.

- 319 Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological
320 family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal*
321 *of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1271–1278.
- 322 Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information
323 theoretical perspective on morphological processing. *Cognition*, *94*, 1–18.
- 324 Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morpho-
325 logical family size in monolingual and bilingual word processing. *Journal of Memory and Language*, *72*,
326 59–84.
- 327 Niemi, J., Laine, M., & Tuominen, J. (1994). Cognitive morphology in Finnish: foundations of a new model.
328 *Language and Cognitive Processes*, *9*, 423–446.
- 329 Pham, H. & Baayen, H. (2015). Vietnamese compounds show an anti-frequency effect in visual lexical
330 decision. *Language, Cognition and Neuroscience*, *30*(9), 1077–1095.
- 331 Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. London: Weidenfeld and Nicolson.
- 332 Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic
333 segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*, 1090–1098.
- 334 Schreuder, R. & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*,
335 *37*, 118–139.
- 336 Seidenberg, M. S. & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of
337 codes. *Trends in Cognitive Sciences*, *4*(9), 353–361.
- 338 Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. (2011). Seeing a phrase time and again matters:
339 The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology:*
340 *Learning, Memory, and Cognition*, *37*(3), 776.
- 341 Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? frequency effects on off-line/on-line
342 collocational processing. *Bilingualism: Language and Cognition*, *18*(03), 419–437.
- 343 Soveri, A., Lehtonen, M., & Laine, M. (2007). Word frequency and morphological processing in Finnish
344 revisited. *The Mental Lexicon*, *2*(3), 359–385.
- 345 Sproat, R. (1992). *Morphology and Computation*. Cambridge, Mass.: The MIT Press.

- 346 Sun, C.-C. (2016). *Lexical Processing in Simplified Chinese: An Investigation Using a New Large-Scale*
347 *Lexical Database*. PhD thesis, Eberhard Karls Universität Tübingen.
- 348 Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language*
349 *and Cognitive Processes*, 9(3), 271–294.
- 350 Taft, M. & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning*
351 *and Verbal Behavior*, 14, 638–647.
- 352 Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles:
353 evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613.
- 354 Tremblay, A. & Tucker, B. V. (2011). The effects of n-gram probabilistic measures on the recognition and
355 production of four-word sequences. *The Mental Lexicon*, 6(2), 302–324.
- 356 van Rij, J., Baayen, R. H., Wieling, M., & van Rijn, H. (2016). itsadug: Interpreting time series, autocorre-
357 lated data using GAMMs. R package version 2.2.
- 358 Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More
359 morphological computation in english than in finnish. *Journal of Psycholinguistic Research*, 31(2), 83–106.
- 360 Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in
361 which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*,
362 143(5), 2020.
- 363 Wolter, B. & Gyllstad, H. (2013). Frequency of input and l2 collocational processing. *Studies in Second*
364 *Language Acquisition*, 35(03), 451–482.
- 365 Wood, S. N. (2006). *Generalized Additive Models*. New York: Chapman & Hall/CRC.
- 366 Yang, C. (2016). *The Price of Linguistic Productivity*. The MIT Press.