Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning.

Yu-Ying Chuang[1], Kaidi Lõo[2], James P. Blevins[3], and R. Harald Baayen[1],

University of Tübingen[1], University of Alberta[2], and University of Cambridge (UK)[3]

abstract

According to Word and Paradigm Morphology (Matthews, 1974; Blevins, 2016), the word is the basic cognitive unit over which paradigmatic analogy operates to predict form and meaning of novel forms. Baayen et al. (2019b, 2018) introduced a computational formalization of word and paradigm morphology which makes it possible to model the production and comprehension of complex words without requiring exponents, morphemes, inflectional classes, and separate treatment of regular and irregular morphology. This computational model, Linear Discriminative Learning (LDL), makes use of simple matrix algebra to move from words' forms to meanings (comprehension) and from words' meanings to their forms (production). In Baayen et al. (2018), we showed that LDL makes accurate predictions for Latin verb conjugations. The present study reports results for noun declension in Estonian. Consistent with previous findings, the model's predictions for comprehension and production are highly accurate. Importantly, the model achieves this high accuracy without being informed about stems, exponents, and inflectional classes. The speech errors produced by the model look like errors that native speakers might make. When the model is trained on incomplete paradigms, comprehension accuracy for unseen forms is hardly affected, but production accuracy decreases, reflecting the well-known asymmetry between comprehension and production. Unseen principal parts (i.e., nominative, genitive, and partitive singulars) are particularly difficult to produce, possibly due to their more distinctive forms. Removing principal parts from training, however, does not affect accuracy for other case forms. Model performance does not degrade either when the training data includes the alternative forms that are ubiquitous in Estonian. These results are consistent with the claim of Blevins (2008) that Estonian number and case inflection is organized in a way that facilitates the deduction of full paradigms from only a small number of forms.

1 Introduction

The morpheme, understood as the smallest meaning-bearing unit (Plag, 2003; Lieber, 2010), has played an important role in theories of morphology. For a long time, the focus of morphological analysis has been centering on identifying morphemes and creating rules that put them together to form words. Combined with the axiom that language is at heart a compositional system, morphologically complex words are then analysed as hierarchically organized graphs with morphemes as nodes. These graphs underlie both words' forms and their semantic interpretation. This particular view of the morpheme has been extremely influential in psychology and cognitive science (see, e.g., Zwitserlood, 2018; Butz and Kutter, 2016, who take the cognitive reality of morphemes as minimal signs to be an established fact). Yet, there are at least seven good reasons why the morpheme, in the sense of a minimal sign, is an undesirable linguistic construct.

1. **Allomorphy**. The exact form of morphemes may vary. Thus, the English plural morpheme has three variants (allomorphs), [s] as in *cats*, [z] as in *hands*, and [ɪz] as in *houses*. The plural morpheme of Turkish has two allomorphs, *ler* as in *evler* (houses), and *lar* as in *adamlar* (men). Stem morphemes can likewise alternate, compare English *sing* and *sang*, and Classical Hebrew *katab-nu* (we write) and *ni-ktob* (we wrote). Some, but not all allomorphy appears to be driven by pressures to simplify articulation. And at least one language, Pirahã, spoken

in Brazil, is known for abundant unconditioned allomorphy. This language, which has only 11 phonemes, and thus has one of the smallest phoneme inventories in the world, allows the stops [p, t, k, ?] to freely interchange. Thus, the morphs *xapapaí, kapapaí, papapaıı, xaxaxaí*, and *kakakaí* are interchangeble and all mean 'head' (O'Neill, 2014), although the segmental variation appears to differ across idiolects, and the segmental variability is counterbalanced by a rich suprasegmental system. Languages allow words to have forms that are 'good enough', rather than having the unique and fixed forms one would expect for yoked calculi of form and meaning.

2. **Suppletion**. English *go* and *went* exemplify the phenomenon of suppletion, the use of stems that are so different that they cannot be plausibly categorized as allomorphs. One might think that morphs such as *went* are the incidental by-product of historical accidents — English *went* is originally the past tense form of *to wend*, which has replaced the original past tense of *go*. And as a consequence, one might reason that suppletion, which flies in the face of compositionality, should always be at the fringes of morphological systems. This, however, turns out not to be the case. The Yeli Dnye language, spoken on Rossel Island, of the eastern tip of Papua New Guinea, has hardly a verb that does not have many suppletive alternative forms (Levinson and Majid, 2013; Majid et al., 2004; Levinson, 2006).

3. **Portmanteau morphs**. Allomorphs such as English *sing* and *sang*, and Hebrew *katab-nu* and *ni-ktob*, challenge decompositional systems in that inflectional functions (in the present example, tense) and lexical meanings are packed together in basic forms. This packing together of different meanings in simple forms is not restricted to suppletion. It is found for basic word forms as well. For instance, some English verbs jointly express motion and direction of motion, compare *come* and *go*, but most English verbs of motion pack together the fact of motion with the manner of motion, compare *walk, run, saunter, slither, ...* (Talmy, 1985).

   In many languages, inflectional exponents also express multiple inflectional functions simultaneously, compare Latin *hortus* (garden), where the *-us* exponent realizes both nominative case and singular number. Yeli Dnye combines extensive stem allomorphy with a large collection of over 1000 portmanteau morphs, often monosyllables, that express combinations of negation, tense, aspect, person and number of subject, deixis, evidentiality, associated motion, and counterfactuality (Levinson and Majid, 2013). The existence of portmanteau morphs and their presence in many languages shows that languages can function perfectly well without isomorphisms between building blocks, and combinations thereof, of forms and meanings.

4. **Homophony**. The mirror image of allomorphy is homophony, with one form linked to several meanings. Homophony occurs both for content words (compare English *time* and *thyme*), and for inflectional exponents, such as English *-s* (compare *she walks*, third person singular on verbs; *the walks*, plural on nouns; *the girl's house*, genitive singular on nouns; *the girls' house*, genitive plural on nouns; and *he's here*, pronominal clitic). Recent research shows things are not as bad as it might seem, as there appear to be at least some systematic differences in acoustic duration between homophones (Gahl, 2008; Plag et al., 2015, 2017). However, the differences observed are differences in the mean, and the overlap in distributions of durations is much larger than one would expect under a calculus driven by basic form-meaning pairings.

5. **Multiple exponents**. In some languages, one finds that a single set of morphosyntactic features is realized in multiple places in a word's form. This phenomenon is known as multiple exponence. Matthews (1974) and Harris (2009) point out that multiple exponence is highly

problematic for morpheme-based theories, for an overview of the phenomenon and the theoretical discussions it has given rise to, see Harris (2017). An example of multiple exponence, taken from Harris (2009), is the Archi inflected form *das:árejrut:ur* (my own, refering to a female possessor), which is analyzed as d-as:á-r-ej-r-u-t:u-r, II-OF.MYSELF-II-SUFFIX-II-SUFFIX-SUFFIX-II, where the exponent glossed with the roman numeral II is a class marker for the possessed noun. This class marker is expressed at no less than four different places in this form. A form of more modest multiple exponence is evident in the above example of the Hebrew verb forms *katab-nu* (we write) and *ni-ktob* (we wrote). Tense is expressed twice, once in the stem allomorph and once in the affix (*-nu* for the present, *ni-* for the past).

6. **Distribution classes**. At first blush, one might think that case in a language such as Latin expresses a unitary semantic function. However, the Latin ablative can realize very diverse semantic functions that in English would be expressed with a range of different prepositions, such as *from, with, by, in*, and *at*. The only reason that the tabulation of Latin case forms by case and number brings together disparate uses under the label of 'ablative case' is that words with an ablative exponent occur with the abovementioned range of meanings. Accordingly, 'homophonous' case endings, such as the *ae* exponent of Latin feminine nouns (which is classified by the standard paradigm tabulations as genitive and dative singular and nominative plural), tell us no more than that words with this exponent are used to express a wide range of very different semantic functions. The non-unitary nature of case inflections is entirely unexpected from the perspective of morpheme-based theories.

7. **Semantic opacity**. Yoked calculi for form and meaning are possible only when changes in form are paralleled by straightforward and well-definable changes in meaning. In derivational word formation, such clear semantic changes are sometimes analytically relatively straightforward, especially when a morphological operation has a syntactic parallel, as for *unkind* and *not kind*. But the majority of derived words is semantically opaque. For instance, English *worker* does not simply mean 'someone who works' (the meaning advocated in the typical linguistics 101 textbook). Its normal uses are to denote "1a: one that works especially at manual or industrial labor or with a particular material a factory worker — often used in combination, b: a member of the working class"; and "2: any of the sexually underdeveloped and usually sterile members of a colony of social ants, bees, wasps, or termites that perform most of the labor and protective duties of the colony" (https://www.merriam-webster.com/dictionary/worker). The degree to which languages can be non-compositional is especially well exemplified by the Navajo language. As Young and Morgan (1980) put it,

> ... the verb stem is a root morpheme that rarely exists as a lexical form independently of prefix elements, and a verb base is merely the "skeleton" of a meaningful lexical construction, devoid of meaning until the blanks representing paradigmatic prefixes have been filled in. (p. v)

For derived words (and compounds), the standard strategy in formal linguistics is to relegate semantically opaque forms to the lexicon, the repository of the unpredictable. This strategy, however, is not available for 'systematic idiosyncracy' in inflectional systems.

For instance, in Estonian, the regular locative case forms are based on a stem that is identical to the corresponding genitive form, singular or plural.[1] The rules here are simple and uniform:

---

[1] The term 'locative' case is used in an extended sense here to include translative, essive, abessive and comitative cases. This usage avoids the potentially misleading connotations of 'semantic' cases, given that these case forms do not preserve the meaning of the grammatical case forms that serve as their bases.

to form the singular elative, take the genitive singular and add -*st*; to form the plural elative, take the genitive plural and again add the same -*st* ending. The same strategy applies to all 11 locative cases; each consists of a number-appropriate genitive stem and a number-invariant ending. In addition, the nominative plural of every Estonian noun is formed from the genitive singular and the ending -*d*. The formation of the other grammatical case forms exhibits some class-specific variation, which is summarized in section 2 below. These simple rules illustrate the severe challenges that Estonian presents for morpheme-based theories: locative case forms such as the plural abessive, elative, or translative, are semantically not a function of the corresponding genitive, and the genitive singular does not even provide a number-appropriate stem for the nominative plural.

Several solutions have been put forward for addressing the above fundamental problems with the theoretical construct of the morpheme as minimal sign. One solution is to use a variety of formal mechanisms that whip a morphological system into the mold of an item-and-arrangement system such as Turkish. An example is the analysis of Hebrew stem allomorphy proposed by McCarthy (1981), according to which the allomorphs *katab* (present) and *ktob* (past) are comprised of two morphemes, a root morpheme consisting of the consonants *ktb* and a vowel morpheme, *a-a* for present and *o* for past tense. For a critique, see Ussishkin (2005).

Another solution is to give up the idea that morphemes are linguistic signs. Morphemes can then be reconceptualized as units of form. Realizational theories of morphology, such as developed by Stump (2001), avoid the term 'morpheme' and use the term 'exponent' to refer to units of form. Given a lexical meaning and a set of inflectional features (spelling out number, person, case, tense, etc.), rules and representations are set up that formalize how bundles of inflectional features are realized at the form level.

Morphemes can also be reconceptualized as minimal units of grammatical analysis (Bauer, 1983). The English form *was* is then analyzed as combining the morphemes BE, PRETERITE, and SINGULAR, all of which are expressed in a portmanteau morph, *was*. The theory of distributed morphology (Halle and Marantz, 1993) takes inflectional features to be syntactic in nature, and 'distributes' these features to different parts of syntactic tree graphs. Unlike realizational theories, distributed morphology retains the term 'morpheme', but uses it to denote alternatively units of form and units of meaning, but without positing one-to-one links between the two (Marantz, 2013).

Word and paradigm morphology (Matthews, 1974; Blevins, 2016) eschews splitting words' forms into sublexical form units. Whole words are the central and only lexical units. Proportional analogies between words are taken to drive generalization and productivity. A computational formalization of Word and Paradigm Morphology is presented in Baayen et al. (2018, 2019b). In this study, we apply their approach to the declensional system of Estonian. We will show that Estonian nominal inflection can be modeled with high precision without requiring form units for stems and exponents, and without needing to define inflectional classes. With respect to words' semantics, however, our approach remains analytical. Like Bauer (1983), we posit minimal units of grammatical analysis (which Bauer calls morphemes, but for which we use the term lexome, for terminological clarity) which we operationalize using methods from distributional semantics. In what follows, we first provide further information on Estonian inflection, and then introduce our computational model.

2 Estonian inflection

The Estonian declensional system appears at first sight to exhibit a bewildering number of forms and an extreme degree of variation. However, patterns of variation are highly systematic, and define a tight implicational organization at the level of paradigms and classes. The 'intramorphological

information' expressed by form variation allows Estonian to approach the theoretical limit in which a single form of a noun can identify the remaining (27) regularly inflected forms. Unlike in Indo-European languages, this information is not mediated by an association between classes and gender (which the language lacks altogether), or, for the most part, expressed by distinctive patterns of affixation. Instead, the organization of the Estonian declensional system is determined largely by variation in the shape and prosodic structure of the stems underlying one or more diagnostic case forms. The role of these 'principal parts' has long been recognized in traditional descriptions, which tend to distinguish three basic forms: the nominative, genitive and partitive singular (though the class of most nouns can be identified from the genitive or partitive singular alone). The treatment of the genitive and partitive singular forms as basic rests ultimately on the fact that the remaining forms of a nominal are transparently related to, hence deducible from, the genitive or partitive singular. Conversely, given the limited variation in endings, the genitive or partitive singular forms can themselves be deduced from just about any other form of a noun. The result is a system in which almost any case form identifies one of the class-defining principal parts.

The word-based organization of the Estonian declensional system can be illustrated with reference to the paradigm of LAGI 'ceiling'. As noted above in section 1, the eleven locative case forms (illative, inessive, elative, allative, adessive, ablative, translative, terminative, essive, abessive and comitative) are all based on the corresponding genitive. The genitive singular *lae* directly implies the shape of the eleven singular locative forms (and the nominative plural *laed*). Given that locative (and nominative plural) case endings are invariant in Estonian, the singular locative case forms and the nominative plural all imply the shape of the genitive singular, and thereby identify the class of LAGI. The genitive plural *lagede* likewise implies – and is implied by – each of the plural locative case forms. In nouns like LAGI that exhibit 'weakening' gradation, the genitive plural *lagede* also predicts the form of the partitive singular *lage*. The partitive singular, in turn, predicts the partitive plural *lagesid*. The form of the genitive singular *lae* in this paradigm illustrates a pattern of 'qualitative' gradation that is no longer productive in modern Estonian. In paradigms of nouns that exhibit 'quantitative' gradation, such as LUKK 'lock', just the genitive singular *luku* will suffice, given that the 'long' stem consonant -k- can only occur in a paradigm where it alternates with an 'overlong' -kk-, here in the partitive singular *lukku*. Hence at least two forms, i.e., forms that identify the the genitive and partitive singular, are required to reliably predict the full paradigm of *lagi*. Given this pair of forms, the remaining 26 regular case forms are predictable.

Like many first declension nouns (i.e., nouns with trochaic genitive singulars and partitive singulars ending in a vowel), LAGI has an alternative 'short' illative singular form. In nouns that exhibit quantitative gradation, such as LUKK, the short illative is identical to the partitive singular, *lukku*. In other first declension nouns, the short illative is an overlong counterpart of the partitive singular. Thus in the paradigm of LAGI, the short illative is *lakke*, in which -kk- is the overlong counterpart of the short -g- in *lage*. The same pattern holds for first declension nouns that do not exhibit gradation. For example, the paradigm of MAJA 'house' contains the partitive singular *maja* and the short illative *majja*.

The paradigms of LUKK and MAJA contain a further alternate form based on the partitive singular, namely a fusional 'stem' partitive plural. In first declension paradigms containing this form, it usually preserves the stem of the partitive singular and exhibits one of a restricted number of vowel alternations. Hence the stem partitive plural of LUKK is *lukke*, in which the theme vowel -u alternates with -e, while the stem partitive plural of MAJA is *maju*, in which the theme vowel -a alternates with -u.

Second declension nouns (i.e., nouns with non-trochaic genitive singulars and partitive singulars ending in a consonant -t/-d) contain a different set of alternate forms, again related to the partitive. Whereas regular partitive plurals end in -sid in the first declension, they end in -id in the second

Table 1: Case and number inflections for three word forms from the paradigm LAGI 'ceiling'.

	Lexeme	Case	Number
lagi	LAGI	NOMINATIVE	SINGULAR
lage	LAGI	PARTITIVE	SINGULAR
laed	LAGI	NOMINATIVE	PLURAL

declension. For example, the partitive plural of the noun RAAMAT 'book' is *raamatuid*. This form in turn predicts alternative 'i-plural' forms of 7 of the 11 the locative cases (illative, inessive, elative, allative, adessive, ablative and translative). Alongside the regular elative plural *raamatutest*, the paradigm of RAAMAT contains an alternative *raamatuist*, and similarly for the other 6 locative cases. Like the regular plurals, these alternative forms are predictable from the shape of grammatical case forms, in this instance, the form of the regular partitive plural.

Moreover, as documented in the pedagogical, philological and theoretical literature on Estonian declensions, see, e.g., Erelt (2003) and Blevins (2008), the types of implicational patterns that hold within the paradigm of LAGI apply across the entire inflectional system of Estonian. Thus Estonian provides an instructive example of a system that is not organized in such a way as to facilitate the disassembly of individual forms into smaller recurrent parts. Instead, the system is organized in a way that facilitates the deduction of full paradigms from a small number of forms, rarely more than two, and often just one.

In what follows, we examine whether the approach to the mental lexicon proposed by Baayen et al. (2019b, 2018) makes the right kind of predictions not only for English and Latin, but also for Estonian noun declensions. The next section introduces the mathematical engine of this approach, Linear Discriminative Learning, after which we present a series of computational experiments based on a set of 232 nouns from 26 inflectional classes.

3   Linear discriminative learning

To computationally implement Word and Paradigm morphology, Baayen et al. (2019b) proposed a new model, Linear Discriminative Learning (LDL). Unlike connectionist models that use multiple hidden layers (e.g., the triangle model, Harm and Seidenberg, 2004), LDL works with only two levels, a word form level and a word meaning level, both of which are represented as numeric vectors. Two networks are of specific importance for the present study, one for comprehension, mapping word form to word meaning, and the other for production, mapping word meaning to word form. By way of example, consider three Estonian word forms from the paradigm LAGI 'ceiling', whose case and number inflections are specified in Table 1.

For representing word forms, we made use of sets of features (henceforth cues) consisting of three phones (henceforth triphones). In these triphones, a word boundary is marked by the symbol #. For each word, cues that are present are coded as 1 whereas cues that are absent are coded as 0. The matrix bringing together words' form vectors is referred to as the cue matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \begin{array}{c} \\ lagi \\ lage \\ laed \end{array} \begin{array}{cccccccccc} \text{\#la} & \text{lag} & \text{agi} & \text{gi\#} & \text{age} & \text{ge\#} & \text{lae} & \text{aed} & \text{ed\#} \\ \left(\begin{array}{ccccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{array} \right) \end{array}.$$

With respect to the representation of words' meanings, we followed the general approach of distributional semantics (Landauer and Dumais, 1997) and constructed a real-valued semantic vector

for each word. Ideally, we would calculate semantic vectors from corpora, as outlined in Baayen et al. (2019b). At the time of writing, such vectors are not available to us for Estonian. Therefore, we simulated semantic vectors. Both content lexomes and lexomes for inflectional functions were coupled with vectors of normally distributed random numbers. In the current example, a random vector of length nine is created for each of the five lexomes. These five semantic vectors constitute the row vectors of a matrix $\boldsymbol{Q}$:

$$\boldsymbol{Q} = \begin{array}{c} \\ \text{LAGI} \\ \text{NOMINATIVE} \\ \text{PARTITIVE} \\ \text{SINGULAR} \\ \text{PLURAL} \end{array} \begin{array}{ccccccccc} \text{S1} & \text{S2} & \text{S3} & \text{S4} & \text{S5} & \text{S6} & \text{S7} & \text{S8} & \text{S9} \\ \left(\begin{array}{ccccccccc} -6.02 & 5.04 & -1.54 & 0.84 & 4.51 & -3.18 & 1.66 & -2.98 & 4.01 \\ -3.65 & 0.56 & -3.08 & -0.28 & -2.01 & -1.05 & 1.94 & -0.26 & -5.17 \\ -8.01 & 3.98 & -0.34 & -2.65 & 2.30 & -2.46 & 5.32 & -2.34 & -0.62 \\ 0.62 & 1.37 & 0.41 & 1.61 & 7.80 & 0.86 & 1.99 & 2.34 & 7.97 \\ 7.04 & 4.38 & 2.77 & 3.01 & -4.36 & -4.14 & 0.24 & -1.67 & -3.35 \end{array} \right) \end{array}.$$

In this example, the semantic dimensions S1, S2, ..., S9 are uninterpretable, but when semantic vectors are derived from corpora using the method of Baayen et al. (2019b), each dimension is linked to a specific lexome and the values represent collocational strengths with these lexomes.

In LDL, the semantics of a complex word is dealt with analytically: it is built up by summing the semantic vectors of its constituent lexomes. Take the word *lagi*, for example. To construct the semantic vector of *lagi*, we take the semantic vectors of the lexomes LAGI, NOMINATIVE, and SINGULAR in $\boldsymbol{Q}$ and add them up:

$$\overrightarrow{\text{lagi}} = \overrightarrow{\text{LAGI}} + \overrightarrow{\text{NOMINATIVE}} + \overrightarrow{\text{SINGULAR}}. \tag{1}$$

Repeating this procedure for each word gives us the semantic matrix $\boldsymbol{S}$:

$$\boldsymbol{S} = \begin{array}{c} \\ \textit{lagi} \\ \textit{lage} \\ \textit{laed} \end{array} \begin{array}{ccccccccc} \text{S1} & \text{S2} & \text{S3} & \text{S4} & \text{S5} & \text{S6} & \text{S7} & \text{S8} & \text{S9} \\ \left(\begin{array}{ccccccccc} -9.05 & 6.97 & -4.22 & 2.18 & 10.30 & -3.37 & 5.59 & -0.90 & 6.81 \\ -13.41 & 10.39 & -1.47 & -0.20 & 14.61 & -4.78 & 8.97 & -2.98 & 11.36 \\ -2.63 & 9.98 & -1.85 & 3.58 & -1.87 & -8.37 & 3.84 & -4.91 & -4.51 \end{array} \right) \end{array}.$$

Given $\boldsymbol{C}$ and $\boldsymbol{S}$, we can map the form vectors of $\boldsymbol{C}$ onto the semantic vectors of $\boldsymbol{S}$ by means of a transformation matrix $\boldsymbol{F}$. This mapping represents a central part of the comprehension process. For production, we reverse the process and use a transformation matrix $\boldsymbol{G}$ to move from the semantic vectors in $\boldsymbol{S}$ to the form vectors of $\boldsymbol{C}$. The two transformations $\boldsymbol{F}$ and $\boldsymbol{G}$ are estimated by solving the equations:

$$\boldsymbol{C}\boldsymbol{F} = \boldsymbol{S}, \tag{2}$$
$$\boldsymbol{S}\boldsymbol{G} = \boldsymbol{C}. \tag{3}$$

Mathematical details about how to estimate $\boldsymbol{F}$ and $\boldsymbol{G}$ can be found in Baayen et al. (2019b, 2018). Here, we illustrate how comprehension equation 2 works. The matrix $\boldsymbol{F}$, rounded to two decimal digits, is:

1	1	1	1	0	0	0	0	0	
-3.54	-3.84	-0.83	-0.83	-3.01	-3.01	0.30	0.30	0.30	
-3.54	-3.84	-0.83	-0.83	0	0	0	0	0	**-9.05**

$$\boldsymbol{F} = \begin{array}{c} \\ \text{\#la} \\ \text{lag} \\ \text{agi} \\ \text{gi\#} \\ \text{age} \\ \text{ge\#} \\ \text{lae} \\ \text{aed} \\ \text{ed\#} \end{array} \begin{pmatrix} \text{S1} & \text{S2} & \text{S3} & \text{S4} & \text{S5} & \text{S6} & \text{S7} & \text{S8} & \text{S9} \\ -3.54 & 4.18 & -1.11 & 0.92 & 3.06 & -2.63 & 2.68 & -1.42 & 1.66 \\ -3.84 & 2.25 & -0.87 & 0.03 & 4.70 & -0.72 & 2.30 & -0.26 & 3.71 \\ -0.83 & 0.27 & -1.12 & 0.61 & 1.27 & -0.01 & 0.30 & 0.39 & 0.72 \\ -0.83 & 0.27 & -1.12 & 0.61 & 1.27 & -0.01 & 0.30 & 0.39 & 0.72 \\ -3.01 & 1.98 & 0.25 & -0.58 & 3.43 & -0.71 & 1.99 & -0.65 & 3.00 \\ -3.01 & 1.98 & 0.25 & -0.58 & 3.43 & -0.71 & 1.99 & -0.65 & 3.00 \\ 0.30 & 1.93 & -0.25 & 0.89 & -1.64 & -1.91 & 0.39 & -1.16 & -2.06 \\ 0.30 & 1.93 & -0.25 & 0.89 & -1.64 & -1.91 & 0.39 & -1.16 & -2.06 \\ 0.30 & 1.93 & -0.25 & 0.89 & -1.64 & -1.91 & 0.39 & -1.16 & -2.06 \end{pmatrix}.$$

This matrix summarizes the weight on the connections from triphone cues to the semantic outcomes in the network shown in Figure 1. To see how the matrix multiplication in equation 2 works, we zoom in on how the first element of the semantic vector for the word *lagi* is derived. As illustrated in Table 2, we take the form vector of *lagi* (the first row in $\boldsymbol{S}$) and the first column vector of $\boldsymbol{F}$, and pairwise multiply. The resulting values are then summed. When the unrounded values of $\boldsymbol{F}$ are used, the resulting sum is -9.05, exactly the value of the first element of the semantic vector of *lagi*. The other elements of the semantic vector of this word are calculated likewise by taking the corresponding columns of $\boldsymbol{F}$. Returning to the network representation of $\boldsymbol{F}$, note that the form vector of *lagi* specifies which triphones are present and which are absent. The triphones that are absent are inert, they do not contribute to the network's predictions. For the triphones that are actually present, the weights on their connections to a given semantic outcome, in the present example S1, are summed, yielding the value -9.05.

For this simple example, the semantic vectors calculated by multiplying $\boldsymbol{C}$ with $\boldsymbol{F}$ are identical to those in $\boldsymbol{S}$ up to six decimal digits. For larger more realistic datasets, this will not be the case, but they will be very similar. Below, we discuss in more detail how we evaluate the accuracy and the predictions of the networks. We used the R package WpmWithLdl (Baayen et al., 2019a) to run the simulations reported below. The supplementary material provides the code for running these simulations.

It is important to note that in the LDL framework, no attempt is made to isolate morphs, stems, exponents, or inflectional classes at the level of words' forms. LDL is non-decompositional at the form level, but analytical at the semantic level. In previous work, we showed that this simple setup works well for English inflection and derivation (Baayen et al., 2019b), and we also provided a proof of concept that LDL can handle Latin verb conjugations (Baayen et al., 2018). In what follows, we extend this approach to Estonian nominal inflection.
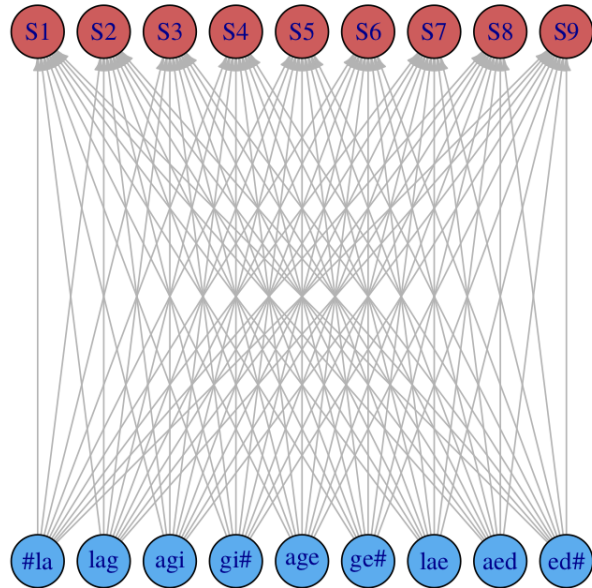
Figure 1: The network corresponding to the transformation matrix $\boldsymbol{F}$.

4 Modeling Estonian case inflections

4.1 Dataset

Our Estonian dataset contains 232 nouns, and for each noun 28 case inflections (14 cases $\times$ 2 numbers), adding up to a total of 6496 word forms. These nouns are distributed across the 26 inflectional classes suggested by the online dictionary of standard Estonian.[2] The word forms were declined by the second author, a native speaker of Estonian, who followed the examples provided by the dictionary. In cases where more than one word form is possible for a given paradigm slot (i.e., it has alternative or parallel forms), the most natural-sounding and frequent form was selected.

In what follows, we first consider this dataset with complete paradigms of exactly 28 case forms. Parallel forms will be discussed later in Section 4.4. As for many nouns only a subset of paradigm cells is in actual use, we below also consider incomplete paradigms (Section 4.3).

4.2 Comprehension and production

4.2.1 Comprehension accuracy

To evaluate model performance for comprehension, we generated the predicted semantic vectors for the same words as used for training. Given the cue matrix $\boldsymbol{C}$ and the mapping $\boldsymbol{F}$, we obtained the predicted semantic matrix $\hat{\boldsymbol{S}}$. For each predicted semantic vector ($\hat{\boldsymbol{s}}$), we calculated its correlation with all the semantic vectors in $\boldsymbol{S}$, and the word whose semantic vector had the highest correlation with $\hat{\boldsymbol{s}}$ was selected as recognized. If the recognized word is identical to the target word, word recognition is successful. Comprehension accuracy was 99.2%. A closer examination of the 52 errors revealed that content lexomes are seldom misunderstood. Most errors occurred for cases, especially those of the nominative and partitive. However, for 41 out of 52 errors, the targeted semantic vector was ranked among the top five.
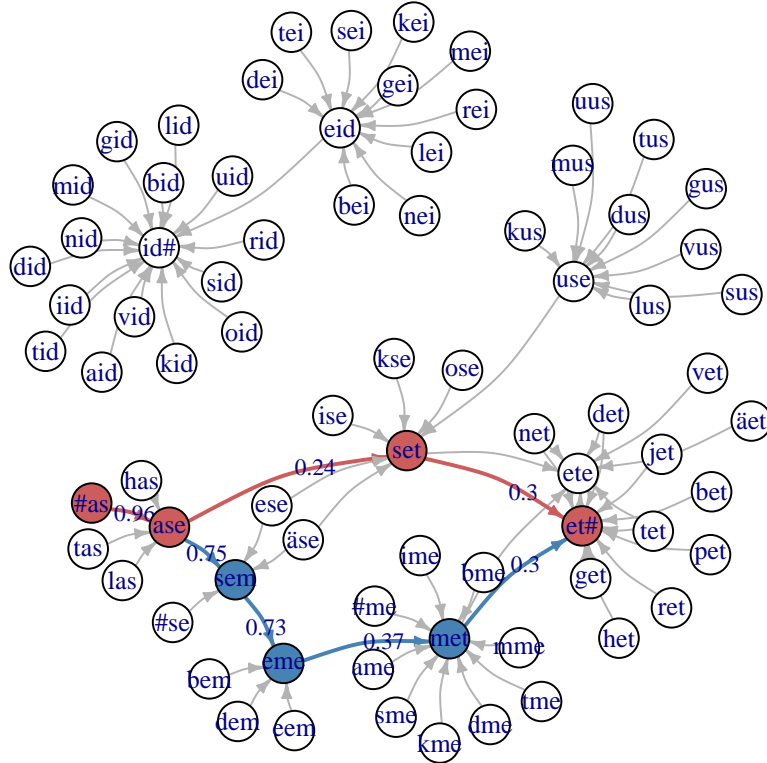
---

[2]https://www.eki.ee/dict/qs/

Figure 2: The paths in the thinned triphone graph of the word *aset* 'stead' PARTITIVE SINGULAR (marked in red) and its predicted pronunciation form *asemet*. The divergent parts between the observed and predicted forms are marked in blue. The numbers indicate activation strength.

4.2.2 Production accuracy

Given the transformation matrix (network) G and the semantic matrix S, the predicted triphone vector for each word ($\hat{c}$) can be generated. A predicted triphone vector specifies which triphones are best supported by a word's semantic vector. However, these activated triphones are not ordered, and hence are not specific enough to drive production. LDL implements an algorithm that arranges triphones into best supported sequences, and selects that sequence of triphones for articulation that when presented to the comprehension network most closely approximates the intended semantic vector (synthesis-by-analysis, Baayen et al., 2018). An example of how triphones are sequenced is shown in Figure 2. For the targeted form *aset*, the partitive singular of the noun *ase* 'stead', the model produces *asemet*, a non-existent form that builds on the strong support for the triphones *sem* and *eme*, which occur in most of this word's inflected forms.

Production accuracy was 91.6%, with 543 production errors. For 93.4% of the errors, the targeted forms are also suggested by the model as candidate forms with high rankings (for 70% of these errors, the targeted form is the model's second best choice). About 93% of the errors are

Table 3: Production errors made by the model.

Error type	Target form	Predicted form	Gloss
Substitution	ar*m*sast	ar*n*sast	'sweet' ELATIVE SINGULAR
	uu*t*eks	uu*r*eks	'novel' TRANSLATIVE PLURAL
Deletion	kub*e*metel	kubmetel	'groin' ADESSIVE PLURAL
	kas*te*	kas	'sauce' NOMINATIVESINGULAR
Insertion	katust	katus*et*	'roof' PARTITIVE SINGULAR
	taevasse	taeva*sse*sse	'heaven' ILLATIVE SINGULAR
	juustes	juus*testes*	'hair' INESSIVE PLURAL
Transposition	va*l*gustelt	va*gl*ustelt	'protein-rich' ABLATIVE PLURAL
	vaese*kes*esse	vaese*sek*esse	'poor-thing' ILLATIVE SINGULAR

novel forms.

4.2.3 Production errors

Upon closer inspection, 16% of the mispronunciations should actually not be counted as errors, since, as judged by the second author, they are alternative forms that are also acceptable and can be understood by native speakers.

Table 3 groups some common errors into four categories: substitution, deletion, insertion, and transposition errors. Although whether these speech errors are also made by native speakers remains an empirical question, they at least do not seem unexpected or unnatural. Often all the necessary triphones have obtained sufficient semantic support, but they have been ordered wrongly. For instance, *vaesesekesse* is produced instead of *vaesekesesse*, a mistake similar to what native speaker students in Estonian grammar classes might produce.

Some errors suggest that the model is regularizing paradigms. Take the production error shown in Figure 2. The model mistakenly inserts *me* into the correct form *aset*. Interestingly, *aseme*, the genitive singular case, is the base form from which all the other cases are derived (e.g., *asemes* IN-ESSIVE SINGULAR, *asemele* ALLATIVE SINGULAR, *asemena* ESSIVE PLURAL). The model apparently predicts the partitive singular to have the same base form, regularizing *aset* to *asemet*.

A change that is valid for certain inflectional classes is sometimes mistakenly carried over to another inflectional class. For instance, for many *i*-ending lexemes, *i* becomes *e* in declension (e.g., *suvi* 'summer', NOMINATIVE SINGULAR, *suv_e_sse* ILLATIVE SINGULAR, *suv_e_s* INESSIVE SINGULAR, *suv_e_st* ELATIVE SINGULAR). While this change does not apply to *voodi* 'bed', the model nonetheless produces *vood_e_sse* for the illative singular, *vood_e_s* for the inessive singular, and *vood_e_st* for the elative singular, the correct forms of which are *vood_i_sse*, *vood_i_s*, and *vood_i_st* respectively.

Plural marking is at times also problematic for the model to predict. Typically, plural is marked by either -*te* or -*de*. The choice is lexeme-dependent, and can confuse the model. Thus, while *ahneteks* 'greedy' ABESSIVE PLURAL is the correct form, the predicted form is *ahne_de_ks*. Conversely, for *ida_de_le* 'east' ALLATIVE PLURAL, the model predicts *ida_te_le*.

Thus, although LDL is able to capture analogical relations between form and meaning, it can run into trouble with lexeme idiosyncracy. One possible solution to this problem is to increase the semantic distance between word forms. This can be achieved by assigning an additional random vector ($\overrightarrow{v}$) to each individual word form. Thus, the semantic vector for the word *lagi*, previously constructed by using 1, is now given by:

$$\overrightarrow{\text{lagi}} = \overrightarrow{\text{LAGI}} + \overrightarrow{\text{NOMINATIVE}} + \overrightarrow{\text{SINGULAR}} + \overrightarrow{v}. \tag{4}$$

The addition of a random vector to the semantic vector of each word implements the hypothesis of 'inflectional islands' proposed by Newman and Rice (2006). They argue that each inflected verb form has its own preferred argument structure and collocational pattern. Random vectors $\overrightarrow{v}$ represent such inflectional islands. With this design built into the model, production accuracy is substantially improved, rising to 99%. Given this improves model performance, in the following simulations, a random vector will be added to the semantic vector of every word form by default. An alternative that we do not further pursue in this study, is to work not with triphones, but with quadraphones. Production accuracy now increases dramatically, with only one remaining error. The use of quadraphones also reduces the number of comprehension errors by two-thirds.

4.3   Incomplete paradigms

Although an Estonian noun has 14 case forms for two numbers, not all members of its inflectional paradigm are in regular use. As shown by Lõo et al. (2018b,a), lexical processing is sensitive to words' inflectional paradigm size, i.e., the number of inflected forms in a paradigm that are actually used by native speakers. In this section, we examine how model performance is affected when trained on incomplete paradigms. We first present results of testing the same words as in training (i.e., seen forms). Next we test our model's productivity by presenting the model with unseen word forms. Finally, the role of principal parts is examined.

4.3.1   Trained and tested on the same forms

For each of the 232 nouns in our dataset, we consulted the Balanced Corpus of Estonian[3] to determine which forms are in actual use. Most paradigms (92%) have numbers of inflected forms ranging from 5 to 25. Only for three nouns, all 28 forms are attested. A cross-tabulation by case and number of attested forms is presented in Table 4. As expected, NOMINATIVE, GENITIVE, and PARTITIVE are used for almost all paradigms, whereas ILLATIVE, TERMINATIVE, and ABESSIVE are more likely to be absent. For two paradigms (*valgune* 'protein-rich' and *šokolaadine* 'chocolaty'), no forms are available in the corpus, they are thus excluded from the analyses below.

Removing unused inflectional members leaves us with 3732 word forms. For evaluation, we first tested the model's predictions for the word forms on which it was trained. Accuracy was high for both comprehension and production, 99.2% and 99.1% respectively.

4.3.2   Trained and tested on different forms

We then presented this model with the 2708 unseen forms of words that are not in actual use. With the default settings of the `WpmWithLdl` package, comprehension accuracy remained high (97.5%), but production accuracy plummeted to 45.2%. The crucial parameter is the threshold value chosen for thinning the Estonian triphone matrix, by default 0.1. The cognitive interpretation is the minimum from the semantics required for a word to be producible. The higher this threshold is set, the less conservative the speaker will be. In what follows, we first consider a conservative speaker using the default value. Later we consider a less conservative speaker with a threshold value of 0.01.

One source of production errors is that mispronounced forms may contain novel triphones (here ranging from one to four) that are absent in the training data. For comprehension, encountering novel cues is relatively unproblematic, as long as the existent cues can together generate semantics that is close enough to the targeted meaning. However, when a triphone is missing, obviously there

---

[3]http://www.cl.ut.ee/korpused/grammatikakorpus/, this corpus consists of 15 million word tokens collected from newspaper, fictions, and scientific texts.

Table 4: Cross-tabulation by case and number of attested forms in the Balanced Corpus of Estonian. The total number of nouns in our study is 232, this is the maximum number of forms that could have been found in the corpus.

	Singular	Plural
Nominative	229	204
Genitive	225	178
Partitive	220	182
Illative	93	95
Inessive	175	115
Elative	193	144
Allative	180	122
Adessive	170	123
Ablative	118	55
Translative	161	82
Terminative	62	33
Essive	110	53
Abessive	77	38
Comitative	175	120

is no path for the word in the triphone graph, and the correct form can never be in the set of paths considered by the model. Discounting such errors increases production accuracy to 50.2%.

Could it be the case that the unused paradigm members are especially difficult words? To test this, we randomly left out the same number of 2708 words from the training data and used them for testing. Comprehension and production accuracies, similar to results of corpus-based data, were 97% and 53.5%[4] respectively. This suggests that the model exhibits a fundamental asymmetry in accuracy for the comprehension and production of unseen forms. This fits well with the claim that in language acquisition comprehension usually precedes production (Ingram, 1974; Clark, 1993), as well as with the observation that we typically understand far more words than we actually actively produce ourselves (Morgan and Oberdeck, 1930; Gershkoff-Stowe and Hahn, 2013).

Despite similar accuracy results, real and random unseen forms, however, differ crucially in terms of their sensitivity to paradigm size. As reported by Lõo et al. (2018b,a), words from larger paradigms are responded to faster in the lexical decision and naming tasks. A mirror image effect is observed in our simulation results: unseen words from paradigms of larger size are more likely to be produced correctly ($r = 0.43$, $p < 0.001$), as shown in Figure 3. Such a correlation is absent for randomly sampled unseen forms ($r = 0.04$, $p = 0.6$). This suggests that there is some optimization in the language ensuring a higher degree of within-paradigm predictivity for larger paradigms.

4.3.3 The role of principal parts

Traditional Estonian grammar regards the singular cases of nominative, genitive, and partitive as the pivotal forms for noun declension. The importance of these three cases can also be inferred from Table 4, which shows that in actual language use they are present in almost all paradigms. We therefore investigated whether the principal parts (henceforth PPs) are also crucial for our model. Does producing the correct forms for other paradigm cells become more difficult if one or more of the PPs are not present during training? To test this, we randomly sampled 800 words, none of which belonged to the PPs, for testing. Seven training sessions were simulated, in which one,

---

[4]Errors resulting from having novel triphones had been discounted here.
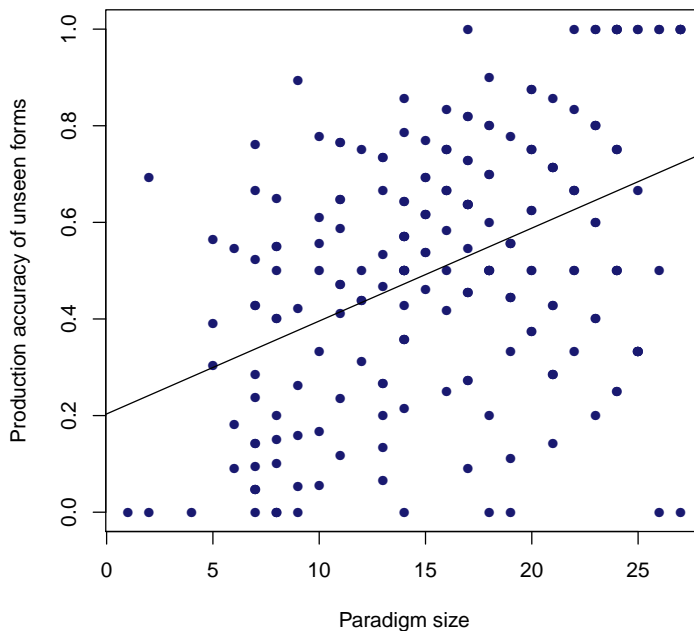
Figure 3: Scatter plot for production accuracy by paradigm size for unseen forms that are not in use.

Table 5: Comprehension and production accuracy (%) for sample unseen words when one or more principal parts are excluded from training. "None" indicates that all principal parts are included, whereas "all" indicates that all are excluded.

	none	NOM	GEN	PART	NOM GEN	NOM PART	GEN PART	all
Comprehension	98.4	98.4	98.6	98.2	98.5	98.2	98.9	98.8
Production	54.0	56.0	54.9	54.5	56.6	56.6	55.6	58.2

two, or all PPs were excluded from the training data. We then examined the comprehension and production accuracy on the same set of 800 unseen words. As shown in Table 5, when all PPs are included in the training dataset (the column "none"), production accuracy was 54.0%, comparable to the results reported above. However, no matter which PPs are removed from training, production accuracy remained basically the same.

At first sight, results are at odds with traditional descriptions, since the absence of PPs does not lead to more mispronunciations. However, unseen PPs appear difficult to produce. As shown in Figure 4, among the 28 paradigm forms, PPs that are not attested in the corpus have lower production accuracy. They are also more likely to contain novel triphones. The only exception is the nominative singular, which has only one unseen token that happens to contain no novel triphones yet is pronounced wrongly. Interestingly, all plural forms, except for plural PPs, rarely contain novel triphones and usually have high production accuracy. Apparently PPs are phonologically more distinctive, but this makes them more difficult to produce.

Taken together, the current results only partially confirm the importance of PPs in Estonian
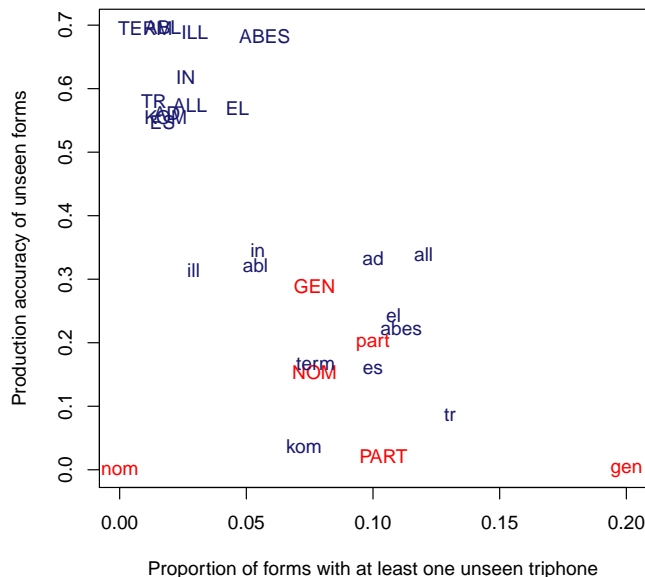
14

Figure 4: Production accuracy by paradigm size for the unseen forms of each case-number combination. Principal parts, both singular and plural, are marked in red, and plural paradigm forms are in capital letters.

noun declension. On the one hand, PPs are more discriminative at the form level, which makes them more difficult to be predicted based on the form-meaning relation learned from other paradigms. In this sense, PPs are indeed more crucial than the other paradigm forms. On the other hand, deducing non-PP forms nonetheless does not always depend on PPs. In fact, due to the high regularlity among non-PPs, they can potentially be deduced from any other paradigm forms (Viitso, 2003, 38). This is consistent with the idea laid out previously in Section 2 that the declensional system is organized in such a way that creates a high co-predictability environment within paradigms.

Thus far we have considered a conservative speaker with a high threshold. When we lower the threshold to 0.01 in order to model a less conservative and more productive speaker, production accuracy increases to 69.5%. Of the 826 errors, 8.7% are genuine errors where the targeted form is outranked by a rival form. In 68% of these cases, the target form is the runner-up.

In order to obtain an estimate of production accuracy when the threshold is lowered even further, we proceeded as follows. The 754 errors that remain are cases for which the targeted form is not found due to the threshold being still too high. When we calculate the correlation between the semantic vector predicted by the targeted form, and the semantic vector to be articulated, in 86.8% of the cases this correlation is stronger than any of the competitors' correlations. This suggests informally an overall production accuracy of 93.7%. This estimate is informal and an upper bound because with even lower thresholds it is possible, although unlikely, that alternative paths with very low triphone support turn out to provide superior correlations.

We note that there is still one disadvantage to the current simulations, namely, that frequency of occurrence does not play a role. To model the effect of frequency, incremental discriminative learning is required, which takes us beyond the scope of this study. Since presumably PPs are more frequently used than non-PPs (and singular forms are more frequent than plurals), higher fre-

quency of occurrence may offset the discriminative vulnerability of PPs (and singulars). Conversely, non-PPs (and particularly plurals) require less protection from frequency, given their already high regularity.

4.4   Parallel forms

As mentioned previously, some paradigm slots may be filled by more than one form. For example, both *akendest* and *aknaist* 'window' are eligible forms for the elative plural. To examine the impact of parallel forms on model performance, we added to the dataset the 144 parallel forms that occur in the Balanced Corpus of Estonian. For terminological clarity, the forms added will be referred to as parallel forms, whereas the forms that are already in the dataset will be referred to as canonical forms. The number of parallel forms per paradigm ranged from one to seven. Forty-eight paradigms had one parallel form added, typically for the partitive, and only one noun (*mõte* 'thought') exhibited seven parallel forms.

Despite the addition of parallel forms, model performance remained robust. For comprehension, accuracy was at 99.26%. Six canonical forms were recognized as their parallel counterparts, which differed from the targeted forms by the random vector $\overrightarrow{v}$ (cf. equation 4). When the model is run without adding $\overrightarrow{v}$, overall accuracy dropped slightly to 99.19%, and only five out of the 144 parallel forms were recognized wrongly.

With regards to production, accuracy was at 98.1% with random vectors $\overrightarrow{v}$. Accuracy dropped to 89% without these vectors. Only a small portion of the errors comes from parallel forms. Given exactly the same semantic vectors for parallel and canonical forms, predicted pronunciations are however more likely to be canonical than parallel forms. This suggests that canonical forms are more consistent and regular than their parallel counterparts. However, we cannot exclude the possibility that this result reflects a bias in our choice of canonical forms.

5   Concluding remarks

From the perspective of decompositional theories, the declensional system of Estonian is complex and irregular. However, we argued that these theories not only miss out on the systematicity and predictability within and across paradigms, they also unnecessarily complicate the system. As demonstrated in the present study, the implementation of LDL, a computational model that completely eschews morphemes, yields highly accurate results for both word comprehension and production.

There are several issues that need to be further addressed. First, although the triphone representations are surprisingly effective, they do not take prosodic features such as syllable structure into account. Second, semantic vectors were simulated, and should be compared with vectors trained on corpora. Third, frequency of occurrence can potentially have a robust impact on production, particularly for the principal parts. This issue can be pursued with an incremental LDL model, using the Widrow-Hoff (Widrow and Hoff, 1960) learning rule. Finally, given that LDL is a cognitively motivated model, its predictions for Estonian lexical processing have to be pitted against human behavioral data such as reaction times and acoustic durations (see Baayen et al., 2019b).

6   Acknowledgement

References

Baayen, R. H., Chuang, Y.-Y., and Blevins, J. P. (2018). Inflectional morphology with linear mappings. *The Mental Lexicon*, pages 232–270.

Baayen, R. H., Chuang, Y.-Y., and Heitmeier, M. (2019a). *WpmWithLdl: Implementation of Word and Paradigm Morphology with Linear Discriminative Learning*. R package version 1.3.1.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019b). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, pages 1–39.

Bauer, L. (1983). *English Word Formation*. CUP, Cambridge.

Blevins, J. P. (2008). Declension classes in Estonian. *Linguistica Uralica*, 44(4):241–267.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press, Oxford.

Butz, M. V. and Kutter, E. F. (2016). *How the mind comes into being: Introducing cognitive science from a functional and computational perspective*. Oxford University Press.

Clark, E. V. (1993). *The lexicon in acquisition*, volume 65. Cambridge University Press, Cambridge.

Erelt, M., editor (2003). *Estonian language*. Estonian academy publishers, Tallinn.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496.

Gershkoff-Stowe, L. and Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of experimental child psychology*, 114(4):489–509.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. and Keyser, S. J., editors, *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, volume 24 of *Current Studies in Linguistics*, pages 111–176. MIT Press, Cambridge, Mass.

Harm, M. W. and Seidenberg, M. S. (2004). Computing the meanings of words in reading: Co-operative division of labor between visual and phonological processes. *Psychological Review*, 111:662–720.

Harris, A. C. (2009). Exuberant exponence in batsbi. *Natural Language & Linguistic Theory*, 27(2):267–303.

Harris, A. C. (2017). *Multiple exponence*. Oxford University Press.

Ingram, D. (1974). The relation between comprehension and production. In Schiefelbusch, R. L. and Lloyd, L. L., editors, *Language perspectives—Acquisition, Retardation, and Intervention*, pages 313–334. University Park Press, Baltimore.

Lõo, K., Järvikivi, J., and Baayen, R. (2018a). Whole-word frequency and inflectional paradigm size facilitate estonian case-inflected noun processing. *Cognition*, 175:20–25.

Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B., and Baayen, R. (2018b). Production of estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, 1(28):71–97.

Landauer, T. and Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

Levinson, S. C. (2006). The language of space in yélî dnye. In *Grammars of space: Explorations in cognitive diversity*, pages 157–203. Cambridge University Press.

Levinson, S. C. and Majid, A. (2013). The island of time: Yélî Dnye, the language of Rossel island. *Frontiers in psychology*, 4.

Lieber, R. (2010). *Introducing Morphology*. Cambridge University Press, Cambridge, UK.

Majid, A., Bowerman, M., Kita, S., Haun, D., and Levinson, S. C. (2004). Can language restructure cognition? the case for space. *Trends in Cognitive Sciences*, 8:108–114.

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

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. Cambridge University Press, Cambridge.

McCarthy, J. J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry*, 12:373–418.

Morgan, B. Q. and Oberdeck, L. M. (1930). Active and passive vocabulary. In Bagster-Collins, E. W., editor, *Studies in modern language teaching*, pages 213–221. Macmillan, London.

Newman, J. and Rice, S. (2006). Transitivity schemas of english eat and drink in the bnc. In Gries, S. T. and Stefanowitsch, A., editors, *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*, pages 225–260. Mouton de Gruyter, Berlin.

O'Neill, G. (2014). Humming, whistling, singing, and yelling in pirahã context and channels of communication in fdg. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 24(2):349–375.

Plag, I. (2003). *Word Formation in English*. Cambridge University Press, Cambridge, UK.

Plag, I., Homann, J., and Kunter, G. (2015). Homophony and morphology: The acoustics of word-final s in english. *Journal of Linguistics*, pages 1–36.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: The acoustics of word-final s in english 1. *Journal of Linguistics*, 53(1):181–216.

Stump, G. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge University Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. *Language typology and syntactic description*, 3:57–149.

Ussishkin, A. (2005). A Fixed Prosodic Theory of Nonconcatenative Templaticmorphology. *Natural Language & Linguistic Theory*, 23(1):169–218.

Viitso, T.-R. (2003). Structure of the Estonian language: Phonology, morphology and word formation. In Erelt, M., editor, *Estonian Language*, pages 9–129. Estonian Academy Publishers, Tallinn.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits. *1960 WESCON Convention Record Part IV*, pages 96–104.

Young, R. and Morgan, W. (1980). *The Navajo language: A grammar and colloquial dictionary.* University of New Mexico Press.

Zwitserlood, P. (2018). Processing and representation of morphological complexity in native language comprehension and production. In Booij, G. E., editor, *The construction of words. Advances in construction morphology*, pages 583–602. Springer.