

Comprehension and production of Kinyarwanda verbs in the Discriminative Lexicon

Ruben van de Vijver
Heinrich-Heine-Universität
Düsseldorf, Germany
ruben.vijver@hhu.de

Emmanuel Uwambayinema
Heinrich-Heine-Universität
Düsseldorf, Germany
emuwa100@hhu.de

Yu-Ying Chuang
Eberhard-Karls-Universität
Tübingen, Germany
yu-ying.chuang@uni-tuebingen.de

Abstract How do speakers comprehend and produce complex words? In the theory of the Discriminative Lexicon this is hypothesized to be the results of mapping the phonology of whole word forms onto their semantics and vice versa, without recourse to morphemes. This raises the question whether this hypothesis also holds true in highly agglutinative languages, which are often seen to exemplify the compositional nature of morphology. On the one hand, one could expect that the hypothesis for agglutinative languages is correct, since it remains unclear whether speakers are able to isolate the morphemes they need to achieve this. On the other hand, agglutinative languages have so many different words that it is not obvious how speakers can use their knowledge of words to comprehend and produce them.

In this paper, we investigate comprehension and production of verbs in Kinyarwanda, an agglutinative Bantu language, by means of computational modeling within the the Discriminative Lexicon, a theory of the mental lexicon, which is grounded in word and paradigm morphology, distributional semantics, error-driven learning, and uses insights of psycholinguistic theories, and is implemented mathematically and computationally as a shallow, two-layered network.

In order to do this, we compiled a data set of 11528 verb forms and annotated for each verb form its meaning and grammatical functions, and, additionally, we used our data set to extract 573 verbs that are present in our full data set and for which meanings of verbs are based on word embeddings. In order to assess comprehension and production of Kinyarwanda verbs, we fed both data sets into the Linear Discriminative Learning algorithm, a two-layered, fully connected network. One layer represent the phonological form and the layer represents meaning.

Comprehension is modeled as a mapping from phonology to meaning and production is modeled as a mapping from meaning to phonology. Both comprehension and production is learned with high accuracy in all data and in held-out data, both for the full data set, with manually annotated semantic features, and for the data set with meanings derived from word embeddings.

Our findings provide support for the various hypotheses of the Discriminative Lexicon: Words are stored as wholes, meanings are a result of the distribution of words in utterances, comprehension and production can be successfully modeled from mappings from form to meaning and vice versa, which can be modeled in a shallow two-layered network, and these mappings are learned in by minimizing errors.

Keywords: Kinyarwanda; verbs; Discriminative Lexicon; Word and Paradigm; word embeddings; Linear Discriminative Learning; computational modeling; comprehension; production

1 Introduction

How do we figure out what people mean when they speak, and how do we figure out how to say something which can be understood? These are two daunting but fundamental tasks for any competent speaker of any language. Many morphological and psycholinguistic theories explain this by appealing to compositionality: complex words are made up of simpler parts that themselves cannot be further analyzed into smaller units. Most morphological theories share this assumption (Halle & Marantz 1993; Booij 2010; 2016; Stump 2001; 2016; 2018; Zwitserlood 2018), even though they have different views on what these simpler parts are.

In theories that are based on the construct of the morpheme—the smallest unit with form and meaning (Bauer 2016; Halle & Marantz 1993; Booij 2012; Haspelmath 2020)—it is commonly assumed that the meaning of a complex word is the sum of the meaning of its morphemes. A speaker's knowledge to comprehend and produce complex words consists of rules that formalize how to combine morphemes; for comprehension a speaker needs rules that formalize how to parse words into their constituent morphemes (Zwitserlood 2018).

Other theories deny the existence of morphemes, but nevertheless rely on composition of complex words by means of rules or schemas. One such theory is Paradigm Function Morphology (Stump 2001; 2016), in which complex words are induced from simpler stems or words (Stump 2018). In Construction Morphology (Booij 2010; 2016) complex words are so-called constructions, which are pairings of form and meaning. Systematic aspects of this pairing are formalized in schemas, which, in turn, can be used to produce and comprehend novel words.

But what if the analysis of complex words into simpler parts proves to be difficult, and, if possible, rather arbitrary (Katamba 1978; Hockett 1954; Blevins 2016a)? In that case speakers could not work out the necessary rules or schemas, which then raises the question by what mechanisms speakers can comprehend and produce complex words.

There is a great deal of evidence that it is indeed the case that complex words cannot easily be analyzed into simpler parts, even for agglutinative languages that are often cited as paradigm examples of the idea that complex words are made up of simpler parts (Bauer 2016; Goldsmith et al. 2016; Katamba 1978), in the same way a pearl necklace is made up of a sequence of individual pearls.

After illustrating these difficulties in the next paragraphs for Kinyarwanda, an agglutinative Bantu language and the empirical focus of this paper and other Bantu languages, we will present an alternative approach to comprehension and production of complex words within the framework the Discriminative Lexicon (Baayen et al. 2018; 2019b), a theory of the mental lexicon.

On why it is difficult to isolate morphemes, stems or exponents in Bantu languages in general, and Kinyarwanda in particular. Verbs in Kinyarwanda are highly inflectional, and are commonly analyzed as consisting of a string of morphemes, the order of which is determined by a template (Banerjee 2019; Creissels 2019; Hyman 2003; Hyman & Inkelas 2017; Wal 2015). In his analysis of Kinyarwanda, Banerjee (2019) follows most literature on Bantu (Hyman 2003; Hyman & Inkelas 2017; Wal 2015) and specifies that this template determines the order of the morphemes as follows: subject - tense - object - verbal radical - extensions - aspect - final vowel. Extension is a collective term for a number of valency changing morphemes (Banerjee 2019), and the final vowel is sometimes analyzed as a mood marker (Goldsmith & Mpiranya 2010). Such a templatic analysis is very insightful diachronically, and from the point of view of typology, but it is not clear whether native

speakers of Kinyarwanda comprehend and produce complex words in terms of their morphemes, because there are many arguments to show that such morphemes can hardly be isolated in a non-arbitrary way.

The observation that it is difficult to isolate morphemes in Bantu languages is not new, nor restricted to a handful of them (Katamba 1978). One of these difficulties is the existence of multiple exponence, in which one meaning is expressed more than once in a word. An example can be found in the Bantu language IsiNdebele, spoken in Zimbabwe, which exhibits multiple exponence, amongst others, in the locative (Ndlovu & Dube 2019). The locative is expressed by a prefix and suffix: *e-gwalw-eni* LOC-book-LOC ‘in the book’. Multiple exponence is also reported for verbs in Lusonga, a Bantu language spoken in Uganda (Hyman & Inkelas 2017). In Lusonga there are two causatives which are both expressed in verb forms (Hyman & Inkelas 2017). Multiple exponence is problematic for the morphemic view in which each morpheme is associated with one meaning (Chuang et al. 2020b).

A further difficulty is the existence of cases of fusion, in which a meaning is spread out over several places in the complex word, thereby making it difficult to decide with which morpheme the meaning is associated. An example is found in Luganda, another Bantu language, related to Lusonga and also spoken in Uganda. The infinitive of ‘to bring’ is *kuleeta*, and its perfective is *aleese* ‘he has brought’ (Katamba 1978). The perfective is expressed by both the final [e], analyzed as a morpheme called *final vowel* and also by the [s], analyzed as part of the verbal stem.

Kinyarwanda (Banerjee 2019; Goldsmith & Mpiranya 2010; Nurse & Philippson 2006) exhibits yet another difficulty for a morphemic analysis, namely allomorphy. Negation in Kinyarwanda is expressed differently in infinitives than in non-infinitives. The negated form of the infinitive *gusoma* ‘to read’ is *gutásomá* (high tones are indicated by acute accents), whereas the negative of the first person singular present tense of ‘to read’ *nda-soma*, is *sínsoma* (Goldsmith & Mpiranya 2010). A morphemic analysis would have to assume that negation has two allomorphs, one for the infinitive and one for the non-infinitive forms, which means that their meaning is not just negation, but negation-for-infinitives and negation-for-non-infinitives. Such analyses fly in the face of the central idea of morphemic theories that the meaning of a complex word is the sum of its constituent morphemes; the meaning of the morphemes, in this case, is dependent on their context.

Tones in Kinyarwanda verbs pose further serious analytical problems for a morpheme-based analysis of complex verbs in Kinyarwanda (Goldsmith & Mpiranya 2010). Verbs either carry a high tone or no tone (Goldsmith & Mpiranya 2010). This can be seen in the contrast between *ndabóna* ‘I see’, with high tone, versus *ndarima* ‘I cultivate’ without a tone. However, if there are object markers in the verb form, these carry the high tone, and the verbal radical remains toneless (Goldsmith & Mpiranya 2010), as is illustrated in the verb form *ndamúbona* ‘I see him/her’. Part of the morpheme of the radical is realized elsewhere, which is difficult to reconcile with the definition of a morpheme.

Finally, Kinyarwanda verb forms also show homophony. Verb forms that express past, perfective, and those that express the subjunctive both end in *e*: *yarashe* ‘he/she/it shot’, and *arase* ‘that he/she/it shoots’. Such homophony is difficult to reconcile with assumptions from morphemic-based theories. In addition, Kinyarwanda also appears to have multiple exponence: Comparing the past, perfective form *yarashe* with the present, perfective form *arasa* ‘he/she/it shoots’ brings to light that the grammatical function past is expressed by the initial *y* as well as the final *e*.

Individually none of these arguments against the construct of the morpheme is fatal to it. But the accumulation of evidence against it, both in terms of the number of analytic problems, and in terms of the enormous amount of languages in which the problems crop up, seriously undermines its usefulness as a linguistic construct.

Word-based theories of morphology are not based on morphemes, and therefore avoid the problems associated with the construct of the morpheme. Yet, these theories rely on other units, for example stems, or exponents, and therefore some word based theories, we will argue, do not solve the problems for Kinyarwanda speakers.

In Paradigm Function Morphology (Stump 2001; 2016) the lexicon is hypothesized to consist of paradigm functions which pair lexemes with morphosyntactic property sets. The functions are themselves associated with rules of exponence—these rules are further functions—that allow a Kinyarwanda speaker to pronounce these words. These rules associate the lexeme and the morphosyntactic property set with a phonological form. Kinyarwanda, like all Bantu languages, has an intricate class system for nouns. In Proto-Bantu the noun class system was semantically productive, but synchronically it no longer is. For each noun there is a class for the singular and a different one for the plural (Demuth 2000; Nurse & Philippson 2006; Wal 2015). For example the Kinyarwanda noun *umupfayongo* “absent minded person” is of class one and can be represented as follows: the lexeme PFAYONGO and its morphosyntactic property set {class one}: <PFAYONGO, {class one}>. Its plural, *abapfayongo*, can be represented as <PFAYONGO, {class two}>. For these two words it is easy to understand how a child would arrive at an analysis in terms of paradigm functions.

It is, however, not always the case that a lexeme can be easily identified. For example, the Kinyarwanda lexeme RAS “to shoot” can be associated with the morphosyntactic property set {infinitive}: <RAS, {infinitive}>. A paradigm function will realize this configuration as *kurasa*. It is unclear how native speakers of Kinyarwanda will arrive at an analysis in which they have identified RAS as a unit. In the infinitive the final vowel is an obligatory part of the verb, but it is not clear which function it has. As far as a speaker is concerned it could also be part of the lexeme.

The same problem arises in construction morphology. Constructions are form-meaning correlations. In a discussion of English agentive constructions, such as for example *baker*, Booij (2016) gives the following example:

$$(1) \quad \langle [X]_{V_i}er \rangle N_j \leftrightarrow [Agent\ of\ SEM_i]_j \rangle$$

In the construction in 1 the double arrow is the correlation between form and meaning, the index i shows that the meaning also appears in the verb, and the index j shows that the meaning of the whole is also the meaning of the whole word. The X is a variable representing the phonological content of the base word. Kinyarwanda infinitives, as for example *kurasa*, can be represented as a construction, as is illustrated in 2.

$$(2) \quad \langle ku[X]a \rangle V_i \leftrightarrow [Infinitive\ of\ SEM]_i \rangle$$

In 2 the x represents the phonological content of the base verb. In this case, too, it is unclear whether a native speaker would arrive also at this conclusion.

In short it is not clear how speakers can isolate the morphemes, lexemes or other purported units of meaning of a Kinyarwanda verb. This, in turn, makes it difficult to see how speakers can produce complex words by compositionally, or comprehend complex words by parsing them into smaller units, and deriving their meaning compositionally

(Zwitserslood 2018).

After this our discussion of problems, it is time to focus on our proposed solution. The Discriminative Lexicon is a theory of the mental lexicon that is word based and eschews morphemes, stems and exponents. In section 2.1 we will introduce the mathematical concepts behind the Discriminative Lexicon, here we will introduce the central theoretical tenets.

The Discriminative Lexicon is a comprehensive theory of the mental lexicon Baayen et al. (2018; 2019b); Chuang et al. (2020b; c; 2021) that brings together several strands from independent theories: With word and paradigm theory it shares the hypothesis that words, not morphemes, stems or exponents are the relevant cognitive units (Blevins 2006; 2016a);¹ with distributional semantics it shares the hypothesis that words get their meaning in utterances (Firth 1957; Landauer & Dumais 1997; Sahlgren 2008; Weaver 1955); from error-driven learning it implements the hypothesis learning is the result of minimizing prediction errors (Rescorla & Wagner 1972; Widrow & Hoff 1960).² From machine learning it incorporates the insight that fully connected neural networks are very successful at language learning (Boersma et al. 2020; Magnuson et al. 2020; Malouf 2017; Pater 2019; Prickett et al. 2018).

This latter insight is adapted to modeling comprehension and production by representing both the phonological forms in a lexicon and their semantics as points in high-dimensional vector spaces. In order to model comprehension the form representations are mapped onto the meaning representations, and in order to model production the meaning representations are mapped onto the form representations (Baayen et al. 2019b). The mappings are fully connected networks, which can be calculated by means of the mathematics of multivariate multiple regression (Baayen et al. 2018; 2019b) (see section 2.1 for more details).

We will illustrate how the insights of various theories come together in the Discriminative Lexicon, by the example of the concept of proportional analogies from word and paradigm morphology and the idea that meanings are the result of the distribution of words in utterances from distributional semantics. Inflected words stand in a proportional analogical relation to other words in the language (Matthews & Matthews 1972). For example, the first person, singular, active, indicative, present tense, imperfective of the verb “to cut”, *ndaca*, is to the second person, singular, active, indicative, present tense, imperfective, *uraca*, so is first person, singular, active, indicative, present tense, imperfective, *x*, to the second person, singular, active, indicative, present tense, imperfective of the verb “to calm”, *urahora*. In this analogy *x* must be *ndahora*. The morphosyntactic features that these forms express, for example, the features first person, singular, active, indicative, present tense and imperfective, are given their interpretation on the basis of their distribution in utterances in which they occur (Blevins 2016a; Firth 1957; Landauer & Dumais 1997; Sahlgren 2008; Weaver 1955).

¹ We are aware of the fact that the construct word has its own problems in typology (Haspelmath 2011). Yet, in individual languages words can be isolated. This can be done, for example, on the basis of entropy, which is relatively high at word boundaries and low in the middle of a word (Goldwater et al. 2009). As we consider only one language, and do not compare languages, we will set the problem of isolating words in Kinyarwanda aside for future research.

² In phonological theory, error-driven learning theories explain the learning of constraint ranking (Boersma & Hayes 2001; Tesar & Smolensky 1998; Tessier 2019). The research in this line of work does not cite Rescorla & Wagner or related work, and most likely has been conceived independently. It is not the case that error-driven learning is considered a general theory for learning phonology, as other work in phonological theory relies on the conception of learning as association between cues and outcomes (Adriaans & Kager 2010; Adriaans 2011), without a role for minimizing errors.

The proportional analogies in the Discriminative Lexicon are used to assess comprehension and production of new words. To do this the word forms in the lexicon have to be represented in phonological units (see section 2.1 for more details on how the phonology of a word form is represented.) In order to comprehend a new word speakers assess what meaning the combination of phonological units best discriminates, and in order to produce a new word speakers assess what combination of phonological units best discriminates the meaning the speakers intend to convey.

There are many possible ways to represent phonological units of a word form: single segments, or bi- or triphones, or syllables. Research on lexical access [McQueen \(2007\)](#) provides an overview of theories of lexical access, and argues that lexical access is difficult because of the variability in the signal, and suggests that the detrimental effects of variability would be mitigated if language users would rely on prelexical information. Yet, the exact nature of these prelexical information is controversial ([McQueen 2007](#)). In work on perceptual learning it has been shown that listeners make use of allophonic information to adjust to peculiarities of individual speakers ([Mitterer et al. 2018](#)). In work in the Discriminative Lexicon these prelexical units are conceived of as sublexical units. Such units encode contextual information, and therefore allophonic information. In languages that has a limited set of syllable structures, such as Kinyarwanda, it is possible that other sublexical units, for example syllables, are more useful ([Pham & Baayen 2015](#)).

A final important property of the Discriminative Lexicon is discriminative learning, ([Rescorla & Wagner 1972](#); [Rescorla 1988](#)), a general theory of learning, which is very successful in animal learning ([Heyes 2012](#)), and has been successfully applied to a great deal of topics in language learning ([Baayen 2011](#); [Baayen et al. 2016](#); [2018](#); [2019b](#); [Chuang et al. 2020b](#); [2021](#); [Milin et al. 2017](#); [Ramscar & Gitcho 2007](#); [Ramscar & Yarlett 2007](#); [Ramscar et al. 2013a](#); [Ramscar 2019](#)). In discriminative learning an association between a cue and an outcome is established, the weight of which changes as a result of minimizing prediction errors from cue to outcome ([Rescorla & Wagner 1972](#)). In error-driven learning the strength of an association between cue and outcome is not related to mere co-occurrence, as is often assumed in language learning ([Saffran et al. 1996](#); [Adriaans & Kager 2010](#)), but to predictivity, and cues to infrequent outcomes may have an outsized effect on the association strength between them ([Olejarczuk et al. 2018](#)). Error-driven learning has been incorporated in Linear Discriminative Learning (LDL) ([Baayen 2011](#); [Baayen et al. 2018](#); [Chuang et al. 2020b](#)), a computational implementation of the Discriminative Lexicon as the Widrow-Hoff learning rule ([Shafaei-Bajestan et al. 2020](#)). In order to incrementally minimize the difference between the predicted outcomes and the outcomes given by the data, the Widrow-Hoff rule uses gradient descent. As incremental learning models require frequency data and usually take long to train, in this study we target at the endstate of learning, a theoretical equilibrium state in which the effect of further learning is negligible. Details of the mathematics to estimate the networks at the endstate of learning will be provided in section 2.1 below.

In order to evaluate the various hypotheses of the Discriminative Lexicon we will analyze comprehension and production of verbs from Kinyarwanda ([Banerjee 2019](#); [Nurse & Philippson 2006](#)) in Linear Discriminative Learning. We will do so by modeling comprehension and production of verbs from two datasets and assessing how well LDL predicts their comprehension and production. One data set consists of 11528 manually annotated verb forms, and the other data set consists of 573 verbs for which word embeddings are available.

This paper is organized as follows. In section 2 we introduce one mathematics and the computational implementation of the Discriminative Lexicon, Linear Discriminative Learning (LDL), and discuss previous work in LDL in which comprehension and production has been modeled for Latin (Baayen et al. 2018) and Estonian (Chuang et al. 2020b). A comparison of the phonotactics of Latin and Estonian and the way in which the phonology was represented in these modelings on the one hand, and Kinyarwanda on the other hand, will result in additional questions as to how to represent the phonology of Kinyarwanda in our modeling. In section 3 we will describe the verbal system of Kinyarwanda, the phonotactics of Kinyarwanda, and describe our dataset. In section 4 we discuss the results of our modeling, and we conclude the paper in section 5.

2 Linear Discriminative Learning

Linear Discriminative Learning is one of the two computational implementation of the theory of the Discriminative Lexicon, the other one being Naive Discriminative Learning (Baayen 2011; Baayen et al. 2016; 2018; 2019b; Chuang et al. 2020b; 2021; Tomaschek et al. 2019).

In this section we will explain how Linear Discriminative Learning represents word forms and meanings, the mathematics and the computational implementation and how LDL models comprehension and production. The mathematics are described in great detail in Baayen et al. (2019b) and in (Shafaei-Bajestan et al. 2020). In section 2.2 we will review the results of modeling comprehension and production in Latin (Baayen et al. 2018) and Estonian (Chuang et al. 2020b).

2.1 Introduction to LDL

LDL models comprehension and production by setting up a fully connected two-layered network, in which one layer represents the phonology of the word forms, and the other layer represents their meaning, and mapping them onto each other (Baayen et al. 2018; 2019b; Chuang et al. 2020b) (see figure 1). The mapping of the meaning onto phonology represents production, while the mapping of phonology onto meaning represents comprehension.

The layer of the network in which the phonology is represented is a matrix. The rows of the matrix consist of form vectors, one for each word, in which the presence or absence of the ngrams of sounds—in our models letters or syllables—in the word form is encoded. If an ngram is present in the word it is encoded as 1 in the vector, and if it is absent it is encoded as 0.

To give a concrete example of the phonological layer, let us consider three word forms from our data set: *ndarasa* 'I am shooting', *urarasa* 'you are shooting' and *bararasa* 'they are shooting'. To represent these with trigrams of letters, a matrix of vectors of trigrams of letters is set up: #nd, nda, dar, ara, ras, asa, sa#, #ur, ura, rar, ara, ras, asa, sa#, #ba, bar, ara, rar, ara, ras, asa, sa#, and for each a 1 or a 0 indicates whether it is present or absent. The name of the matrix, *C*, is short for cue. The vectors of all words are stored in a matrix, called the *C* matrix (Baayen et al. 2018; 2019b).

$$(3) \quad C = \begin{matrix} & \#nd & nda & dar & ara & ras & asa & sa\# & \#ur & ura & \#ba & bar & rar \\ \begin{matrix} ndarasa \\ urarasa \\ bararasa \end{matrix} & \left(\begin{matrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \end{matrix} \right) \end{matrix}$$

In our modeling we represented the phonology of each word in one of three ways. Words are represented either as triphones of letters (phonemes), as bigrams of syllables, or as trigrams of syllables.

We can now turn to the layer of the network in which meaning is represented. This layer, too, is a matrix, which consists of vectors specifying the meaning of a verb form. In order to do this meaning has to be expressed by numbers. This can be achieved in two ways. One method is to use numbers derived from methods of distributional semantics (also known as *word embeddings*) (e.g. Landauer & Dumais 1997). The other method is to use simulated vectors that are constructed based on words' base and inflectional meanings. There are word embeddings for Kinyarwanda and for Kirundi, a Bantu language very similar to Kinyarwanda and spoken mainly in Burundi (Niyongabo et al. 2020). Even though only 537 of the 11528 verbs in our dataset are included in the word embeddings, we decided to use both full data set of 11528 verbs with their simulated meanings and the smaller data set of 537 verbs with meanings derived from word embeddings.

We simulated the meanings of the verbs in the following way. Each lexeme and each grammatical function is assigned a vector of random numbers, which are drawn from a normal distribution with the same mean and variance. The meaning of a word is then represented as the sum of the vectors of its lexeme and its grammatical functions Baayen et al. (2018); Chuang et al. (2020b). The semantic vector of, for example, *ndarasa* is the sum of the vectors of its lexeme and of its grammatical functions, as presented in (4). The content of each vector is described above in section 3.3.

$$(4) \quad \overrightarrow{ndarasa} = \overrightarrow{Kurasa} + \overrightarrow{Person} + \overrightarrow{Number} + \overrightarrow{Tense} + \overrightarrow{Voice} + \overrightarrow{Mood} + \overrightarrow{Extension}$$

In this way the semantic vector of each word form is established. The vectors with semantic information are stored in another matrix called *S* matrix. The *S* matrix of the example of three words is shown in (5).

$$(5) \quad S = \begin{matrix} & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & S12 \\ \begin{matrix} ndarasa \\ urarasa \\ bararasa \end{matrix} & \begin{pmatrix} 13.99 & 0.63 & 12.13 & -4.22 & 16.17 & 15.13 & 1.26 & 26.50 & -4.07 & 4.18 & 0.77 & 4.53 \\ 27.42 & -3.04 & 1.41 & 8.85 & 17.04 & 14.14 & -0.53 & 22.03 & -6.60 & 13.43 & 10.92 & 3.83 \\ -1.97 & -8.90 & -0.19 & -0.75 & 11.16 & 7.59 & 4.21 & 12.93 & 2.32 & 9.06 & -4.50 & 10.60 \end{pmatrix} \end{matrix}$$

As word meanings are now represented by a vector of real numbers, the columns in *S*, unlike the columns in *C*, are not associated with any specific semantic features, and are therefore not directly interpretable. This is not always the case: If semantic vectors are based on word embeddings, the vectors reflect the associations of the words in the data set and their presence in a corpus (Baayen et al. 2018).

The semantic relation between words in our simulations is determined by the similarity between vectors, commonly assessed by correlation. For example, the correlation between *ndarasa* and *urarasa* is 0.70, which is larger than that between *ndarasa* and *bararasa*, 0.53. This is because *urarasa* is closer to *ndarasa* in meaning than *bararasa* is: while *ndarasa* and *urarasa* only differ in Person ('I shoot' vs. 'you (sg.) shoot'), *bararasa* differs from *ndarasa* in both Person and Number ('I shoot' vs. 'they shoot'). Although the use of simulated vectors is far from satisfactory, with the current set-up, these simulated vectors at least can still capture inter-word semantic and inflectional similarity to a certain extent.

The matrices *C* and *S* are used to model comprehension and production. Comprehension is an answer to the question: given a word form, which meaning is predicted, and production is an answer to the question: given a meaning, which word form is predicted.

These two questions are answered by mapping the C matrix onto the S matrix (comprehension), and by mapping the S matrix onto the C matrix (production), respectively.

The mappings (or transformation matrices), represented by F and G , can be obtained by solving the following matrix equations (Baayen et al. 2018; 2019b):

$$CF = S$$

$$SG = C$$

As C and S are known they can be used to derive F and G . In LDL, by representing word forms and meanings with vectors, we are treating them as points in high-dimensional form and meaning spaces. What the mappings of F and G do is to build connections between the two spaces, so that the points in one space can be mapped onto points in the other. That is, the F matrix maps points from C onto points in S , and the G matrix maps points from S onto C . In doing so these mappings can be used to generate predictions for comprehension and production.

For each ngram of the C matrix, the F matrix specifies the weight for each column in the S matrix, thus creating the matrix \hat{S} , which is the predicted semantic matrix. The weights of the ngrams present in a word to a given semantic dimension (e.g. S_1) are summed to give the value of that column for the word in \hat{S} . This is repeated for the ngrams of each word and each column of the S matrix. This procedure is reversed for the predicted production, although in this case, every semantic dimension will contribute to the values in \hat{C} because unlike C , S is not binary-coded. In order to model production the transformation matrix G maps points from S onto points in C , thus creating the matrix \hat{C} for the predicted form matrix. The mathematics is explained in great detail in Baayen et al. (2018).

Let us now describe how the accuracy of the modeling is calculated. The accuracy of comprehension is calculated by correlating the predicted semantic vector with all the gold standard semantic vectors of the dataset. The word meaning with the highest correlation to the predicted semantic vector is selected as the recognized meaning. If the recognized meaning is identical to the targeted meaning, comprehension is considered accurate.

The accuracy of production is assessed by two algorithms. Both search for a path from the first ngram of the word to the last ngram of the word. The first algorithm does this by considering the closest neighbors of the target in the \hat{C} matrix. In our modeling the number of neighbors was set at 15. The algorithm considers all ngrams in these neighbors, and finds all possible orderings of them such that they begin at the beginning and end at the end. Then it checks the semantic vectors these candidate word forms internally evoke, and selects as prediction the word with the semantic vector that has the best correlation with the semantic vector of the target (referred to as ‘synthesis-by-analysis’ in Baayen et al. 2018). If the predicted word form is identical to the target form the prediction is correct. In section 4 this production measure is called production (build).

The second measure of production accuracy is learning-based. The algorithm also finds a path from the first ngram of the word to the last ngram of the word, but it takes into account how much semantic support ngrams receive in each position. To be more specific, again take the word *ndarasa* for example. The first cue (letter trigram) of this word is #nd, the second is nda, and the third is dar, etc. This algorithm, when searching for paths, considers whether a cue fits a given position. That is, the fact that nda is a valid continuation for #nd is not sufficient, in order to be considered by this path-finding algorithm, nda also has to be a well-supported cue in position two. Such positional

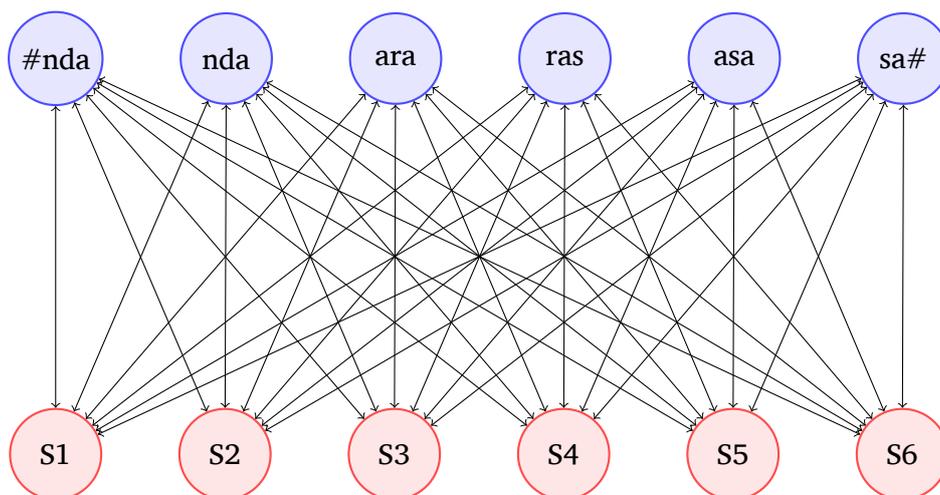


Figure 1: The network representing the mappings of verb forms onto meaning and vice versa, as mediated by the transformation matrices (arrows). These arrows map cues in the word form matrix C (the blue nodes) to semantic dimensions in the meaning matrix S (the red nodes), and vice versa.

learning is achieved by setting up further networks, one for each ngram position. This provides information about the predicted learned strengths for ngrams at a given position for every word. The ngrams are then combined into complete words, and similar to the first algorithm, usually more than one candidate forms can be found. To decide on the best (predicted) form, the same procedure of selection via the internal semantic loop (i.e. ‘synthesis-by-analysis’) is adopted. In section 4 this production measure is called production (learn).

We can now turn to the two parameter settings that were kept constant in all simulations. One concerns the first path-finding algorithm of production (build). In order to produce a word, ngrams must be combined in such a way that their associated semantics correlate with the intended semantics. As each predicted word contains many ngrams, the possible combinations of ngrams quickly becomes very large. In order to reduce the number of possible candidate ngrams, the search is limited to a number of closest phonological form neighbors. The number of form neighbors that provided the ngrams that are used in building the word form to be produced was set at 15.

The other parameter is relevant for the second path-finding algorithm of production (learn). As explained above, the ngram cues at a given position have to receive sufficient semantic support in order to be considered by the path-finding algorithm. The amount of semantic support, controlled by a parameter called `threshold`, thus determines which ngram cues to include. We set this parameter such that ngrams with support that falls short of 0.01 were excluded. In practice this means that very few ngrams were excluded, which can be interpreted as a speaker who would consider many forms, and as a result may make a few mistakes. As Alderete & Davies (2019) show this is not an implausible assumption on our part.

LDL is computationally implemented as the package `JudiLing` (Luo 2021; Luo et al. 2021), for the programming language `julia`.³ The implementation of LDL in `julia` offers a better algorithm for production and testing of held-out data, has an extra function

³ A manual for the `JudiLing` package is available at: <https://megamindhenry.github.io/JudiLing.jl/stable/>. Upon publication our data sets and scripts will be made available through OSF.

to assess the accuracy of production, and is more efficient than the one in R (Baayen et al. 2019a), and therefore reduces the carbon footprint of our modeling. These reasons convinced us to use Judiling.

2.2 Previous work using LDL to model comprehension and production in Latin and Estonian

LDL has been used to model comprehension and production in Latin and Estonian. For Latin Baayen et al. (2018) investigated comprehension and production of 672 different verb forms of 8 verbs, 2 verbs of each of the four conjugation classes. Each verb was inflected for PERSON and NUMBER, for TENSE, ASPECT, MOOD and VOICE. The verb forms were used to create vectors of the word forms by recording for each word form which of the triphone combinations of Latin were present, and which ones were absent. Baayen et al. choose triphones, because they are inherently context sensitive, and would capture information about coarticulations; they convey information about ordering. The semantic vectors were simulated from information about the lexeme and the grammatical function of the verb form. The *C* matrix was mapped onto the *S* matrix to assess comprehension, and this procedure achieved an accuracy of 100%. The *S* matrix was mapped onto the *C* matrix to assess production, and this procedure achieved an accuracy of 99.7%.

For Estonian Chuang et al. (2020b) modeled the comprehension and production of 232 nouns, each inflected for 14 cases and two numbers, yielding a dataset of 6496 nouns. The noun forms were used to create a vector for each word form, recording which triphones from Estonian were present and which ones were absent. The semantic vectors were simulated in the same way as they were for Latin. The *C* matrix was mapped onto the *S* matrix to assess comprehension, and this procedure achieved an accuracy of 99.2%. The *S* matrix was mapped onto the *C* matrix to assess production, and this procedure achieved an accuracy of 91.6%.

Baayen et al. (2018) and Chuang et al. (2020b) conclude that Latin verb forms and Estonian noun forms can be comprehended and produced without recourse to morphemes, providing computational support for the word and paradigm model of morphology (Blevins 2006; 2016b). Yet, an open question concerns the way in which the word forms have been vectorized, and whether this type of triphone vectorization always results in enough different phonological ngrams to distinguish among all meanings in any language. For Latin and Estonian, Baayen et al. (2018) and Chuang et al. (2020b) vectorized the word forms by recording present and absent triphones. Since both Latin and Estonian allow complex onsets and complex codas, there are many different possible triphones and each of them is theoretically able to discriminate among different meanings. Kinyarwanda, though, is different in two respects. First, it has complex segments, prenasalized consonants and affricates, that are written with two letters, but that are one segment (see section 3.3 below). The second difference concerns its syllable structure. All syllables in Kinyarwanda end in a vowel, and only velar glides are allowed in complex onsets. As a result there are fewer different letter triphones to distinguish among meanings. This raises the question as to what is the best phonological representation for the Kinyarwanda verb forms to successfully comprehend and produce them?

3 Kinyarwanda

3.1 Kinyarwanda verbs

Kinyarwanda verbs provide a wealth of lexical and grammatical information. They consist of a verbal radical, which encodes the lexical meaning of a verb, preceded by information about subjects, tense, aspect and mood, objects and followed by information about extra differentiation in the meaning of the verb (such as causative, applicative, frequentative, iterative, comative, reversive), aspect, and verbs end in a vowel (Banerjee 2019). As was reviewed in section 1, the order of this information is fixed and historically determined (Hyman 2003; Wal 2015).

The verb forms in table 1, taken from our dataset, provide only a faint glimpse of the wealth of Kinyarwanda verb forms. They are introduced here to illustrate the flow of information from left to right: The beginning of each verb form co-varies with the grammatical function person, the middle with the lemma meaning and the end with aspect. The order of information is in line with the proposed order in the general Bantu verbal template (Banerjee 2019).

The verb forms further illustrate the difficulty of segmenting the forms into segmental strings which each denote a discrete grammatical function or lemma, while at the same time illustrating how parts of a word can discriminate the meaning of the verb form from other verb forms. For example, all first person forms start with a nasal or a prenasalized stop. The prenasal part of the stop is not a segment by itself (see section 3.2) and hence cannot be a morpheme, but a nasal or nasalized segment at the beginning of a word discriminates the verbs in the first person very well from verb forms of other persons.

3.2 Kinyarwanda phonotactics

Since we are addressing the question as to how meaning is encoded and decoded in word forms, we need to briefly acquaint ourselves with the phonotactics of Kinyarwanda. Kinyarwanda's syllables are of the form C(C)V (Kimenyi 1979). The C in brackets can be filled with glides, and if they are they form the only possible complex onsets. There are five vowels, two high ones: i, u; two mid ones: e, o; and a low one a. Each of these can be short or long (Myers 2005). A source of variation in Kinyarwanda verbs concerns [i, e] and [u, o]. When these are followed by a consonant in the word, they are realized as vowels, but when they are followed by a vowel, they are realized as glides, [j] and [w] respectively (Zorc & Nibagwire 2007).

The consonants of Kinyarwanda are given in table 2, based on Walker et al. (2008) and references therein. There are also a number of prenasalized consonants: ^mb, ^mf, ^mv, ⁿd, ⁿt, ⁿs, ⁿz, ⁿɗ, ⁿʒ, ⁿz̥, ^ɲj, ^ɲc, ^ŋg, ^ŋk.

Since syllables are always vowel final, the verb form *nkubitagurwa* 'I am being frequently beaten' is syllabified as *nku.bi.ta.gu.rwa*. The prenasalized consonant and the consonant-glide sequence are syllabified as onsets (Kimenyi 1979), as they are in many Bantu languages (Kimenyi 1979).

There is a minor controversy about the syllabification of prenasalized consonants (Myers 2005). According to Kimenyi (1979) prenasalized consonants are syllabified as onsets, based on distributional evidence (prenasalized consonants may occur word-initially), and evidence from language games (prenasalized consonants are transposed as units). Myers (2005), in contrast, argues that prenasalized consonants are split in syllabification, with the nasals in the coda and the consonants in the following syllable. The justification for this syllabification, according to Myers, is based on the distribution of long vowels

Verb form	Person	Tense	Lemma	Number	Aspect	Gloss
ndaca	1	present	guca	sing	imperfective	to cut
uraca	2	present	guca	sing	imperfective	to cut
araca	3	present	guca	sing	imperfective	to cut
nca	1	present	guca	sing	perfective	to cut
uca	2	present	guca	sing	perfective	to cut
aca	3	present	guca	sing	perfective	to cut
nacaga	1	past	guca	sing	imperfective	to cut
wacaga	2	past	guca	sing	imperfective	to cut
yacaga	3	past	guca	sing	imperfective	to cut
naciye	1	past	guca	sing	perfective	to cut
waciye	2	past	guca	sing	perfective	to cut
yaciye	3	past	guca	sing	perfective	to cut
ndasoma	1	present	gusoma	sing	imperfective	to read
urasoma	2	present	gusoma	sing	imperfective	to read
arasoma	3	present	gusoma	sing	imperfective	to read
nsoma	1	present	gusoma	sing	perfective	to read
usoma	2	present	gusoma	sing	perfective	to read
asoma	3	present	gusoma	sing	perfective	to read
nasomaga	1	past	gusoma	sing	imperfective	to read
wasomaga	2	past	gusoma	sing	imperfective	to read
yasomaga	3	past	gusoma	sing	imperfective	to read
nasomye	1	past	gusoma	sing	perfective	to read
wasomye	2	past	gusoma	sing	perfective	to read
yasomye	3	past	gusoma	sing	perfective	to read

Table 1: A few of the verbs forms of the the verbs “guca” to cut and “gusoma” to read from the data set. Not shown are verb forms expressing subjunctive and imperative, verbs forms expressing any of the extensions, and verb forms expressing future and passive.

in Kinyarwanda. Vowels preceding prenasalized consonants are intermediate in length between short and long vowels. Myers this is because vowels preceding a prenasalized consonant are in a closed syllable, and are shortened slightly as a consequence.

However, our alternative explanation would be that the length of the vowels before prenasalized is the consequence of lengthening before voiced consonants (Keating 1985), and the prenasalized consonants contain a voiced nasal part. This provides a phonetic basis for lengthening of the preceding vowel, and does not force us into theoretical contortions to explain away the presence of word-initial prenasalized consonants. In line with the analysis of Kimenyi (1979), we will assume that Kinyarwanda syllables have the structure C(C)V.

3.3 Kinyarwanda dataset

The groundwork of our investigation is laid by a dataset created by the second author, a native speaker of Kinyarwanda. We decided to do this, because there are no ready-made datasets for Kinyarwanda that we could use. He inflected a total of 19 different lexemes, for Person (first, second, third), Number (singular, plural), Tense (present, past, future), Voice (active, passive), Mood (imperative, indicative, subjunctive), and each of 5 possi-

	labial	labio-dental	alveolar	postalveolar	palatal	velar	glottal
plosive	p		t, d	c, ɟ		k, g	
nasal	m		n		ɲ		
fricative		f, v	s, z	ʃ, ʒ			h
affricate	pf		ts	tʃ			
flap				ɾ			
approximant	w				j		

Table 2: Consonants of Kinyarwanda.

ble Extensions: applicative, causative, comitative, frequentative, iterative and reversible. This resulted in 11528 verb forms. Syllable structure was added to the word forms, by automatically adding a boundary after each vowel. This data set, then contains the word forms and the meanings (the lexical meaning of each verb and the grammatical functions) that are mapped onto each other in LDL to model comprehension and production.

We choose the grammatical functions in such a way that common ones, such as tense, aspect, mood, number, person, voice, causative and applicative, and not so common functions, such as iterative and frequentative, were all represented. And even though we stress again that our dataset is only a poor man’s version of the complexities of verb forms in Kinyarwanda, we believe that it is representative of its complexity. This level of complexity allows us to explore our hypothesis, without claiming that we provide the final and definitive analysis of Kinyarwanda. One important aspect of Kinyarwanda verbs that we have not considered is the complexities surrounding the distribution of tones (Goldsmith & Mpiranya 2010), which we leave for future work.

We used our data set of 11528 verbs to create a second dataset. We did this because the grammatical functions and lexical meaning have been defined by us, but meaning in the Discriminative Lexicon is based on the distribution of words in sentences (Baayen et al. 2019b; Landauer & Dumais 1997). We therefore extracted from a set of word embeddings for Kinyarwanda (Niyongabo et al. 2020) all verbs that were in our dataset. This resulted in a dataset of 573 verbs. Niyongabo et al. report that the word embeddings were calculated on the basis of 21.268 news articles covering a variety of topics. The articles contained about 300.000 unique words. The word embeddings were created by training Word2Vec embeddings (Mikolov et al. 2013), using the skip-gram algorithm with hierarchical softmax. This resulted in two word embeddings, one with a 50 dimensions and one with 100 dimensions. We used the one with 100 dimensions (Niyongabo et al. 2020).

The word embeddings that we used are derived in a different way than the embeddings used by Baayen et al. (2019b). They used the treetagger algorithm (Schmid 1999) to extract from all the words their stems and their Part-Of-Speech, and included embeddings for inflectional functions, since inflectional does not carry. Niyongabo et al. (2020) used only words and no further information, as no such information is available for Kinyarwanda.

The data sets were used in three different models: in the first the verb forms were represented by means of trigrams of letters. Trigrams of letters have also been used in work on Latin (Baayen et al. 2018) and Estonian (Chuang et al. 2020b), and have been argued to reflect allophonic aspects of segments, which are relevant for lexical access and processing (McQueen 2007; Mitterer et al. 2018). In the second model the verb forms were represented by means of bigrams of syllables, and in the last one the verb forms were represented by means of trigrams of syllables. The syllable-based representations were

chosen, since syllables reflect the phonotactics of Kinyarwanda well, and phonotactics has been proposed to play an important role in learning morphophonology (Hayes 2004; Prince & Tesar 2004).

4 Results

We first present the results of modeling all data, and followed by the results of modeling based on training with 90% of the data and testing on the 10% of the data that were held-out during training. These learning simulations address different properties of the mental lexicon of a native speaker. If we model all data, we pretend that a native speaker knows all of the verb forms in the data set. Considering the size of our data set this is not unrealistic. Brysbaert et al. (2016) estimates that an American 20-year-old native speaker knows about 42.000 word types (lemmas plus their inflections) and a 60-year-old knows about 48.200 word types. Even though these data are probably not directly transferrable to Kinyarwanda, because Kinyarwanda has a richer morphology than English, our data set contains only 19 lemmas, and 11528 inflected forms. Since it is likely that an average 20-year-old Kinyarwanda speaker knows more than 19 lemmas and most of the inflected forms of these lemmas, we expect that an average 20-year-old speaker of Kinyarwanda knows vastly more inflected forms than we used in our dataset. It is therefore reasonable to model all data. It is, however, unrealistic to assume that a speaker would know all forms in each paradigm. In reality of course, some verb forms will occur frequently while others do not occur, even in very large corpora (Karlsson 1986; Lõo et al. 2018; Milin et al. 2009). This is why it is also reasonable to model using held-out data. The results of the held-out data will reflect comprehension and production of very rare word forms, and, of course, reflects the learning situation of a child even for more frequent words.

After these results we will present the results of the data set with the verbs of which the meanings are based on word-embeddings.

4.1 Comprehension and production of all data

In order to assess the comprehension we calculated the predicted semantic vectors \hat{S} on the basis of the word form matrix C and the F matrix (Chuang et al. 2020b). We briefly reiterate how the accuracy is assessed: A prediction is considered correct if the word form with its predicted semantic vector is identical to the target word form. The prediction, in turn, is selected as the word form of which the semantic vector has the highest correlation with the predicted semantic vector as the comprehended one.

The comprehension accuracy of the model in which the word forms were represented as trigrams of letters was 96.4%; the comprehension accuracy of the model in which the word forms were represented as bigrams of syllables was 98.6%; the comprehension accuracy of the model in which the word forms were represented as trigrams of syllables was 99.8% (see table 8.)

As for production, we assessed the accuracy of two measures. We briefly reiterate the assessment of the accuracy for the production (build) measure. The production (build) algorithm selects the 15 closest neighbors of target, considers all ngrams in these neighbors, orders them so that they form words with a beginning and an end, checks the semantic vectors of each potential word form, and selects as prediction the predicted word form of which the semantic vector shows the highest correlation with the semantic vector of the target meaning. If the predicted word form and the target word form are identical, the prediction is counted as correct.

The production (build) accuracy of the model in which the word forms were represented as trigrams of letters was 78.8%; of the model in which the word forms were represented as bigrams of syllables the accuracy was 89.3%; of the model in which the word forms were represented as trigrams of syllables the accuracy was 99.9% (see table 8.)

We briefly reiterate the assessment of the accuracy for production (learn) measure. This measure also creates a path from the first ngram of a word form to its last ngram. During path construction, the algorithm assesses how well ngrams fit a given position, and only ngrams with sufficient positional support are considered. The correlation between the semantic vector of the potential predicted word forms and the target word form is established and the predicted word form with the highest correlation is the predicted word form. If the predicted word form and the target word form are identical, the prediction is counted as correct.

The production (learn) accuracy of the model in which the word forms were represented as trigrams of letters was 82.0%; of the model in which the word forms were represented as bigrams of syllables the accuracy was 94.5%; of the model in which the word forms were represented as trigrams of syllables the accuracy was 99.5% (see table 8.)

LDL models comprehension and production of Kinyarwanda verbs very well, even if no information about morphemes is provided. The production accuracy is slightly lower, but this is to be expected. There are only a few grammatical functions that have to be mapped onto a great number of possible ngrams for possible words. The resulting uncertainty for a particular form increases the likelihood of an erroneous answer. Furthermore, the lower production accuracy matches data from language acquisition (Pater 2004).

It will nevertheless be instructive to have a look at the errors made by LDL. A closer inspection of the comprehension errors shows that LDL makes very few errors for each grammatical function and lexeme, as is illustrated in table 3. Extensions are the most difficult aspect of Kinyarwanda verbs, but still reach an accuracy between 97.1% and 99.7%.

	Person	Tense	Lexeme	Number	Aspect	Voice	Extension	Mood
Letter trigrams	99.8%	99.0%	99.5%	99.9%	99.5%	99.9%	97.1%	99.1%
σ bigrams	99.8%	99.2%	99.9%	99.9%	99.9%	99.9%	98.7%	99.1%
σ trigrams	99.8%	99.3%	99.9%	99.9%	99.9%	99.9%	99.7%	99.4%

Table 3: Accuracy of the models for each grammatical function. For each grammatical function the number of errors was divided by total number of answers involving that function.

The errors in individual forms for each of the models are illustrated from table 4 to table 6. The tables each list the 5 erroneous predictions that were closest to the intended form (target). The column r is the correlation between the predicted semantic vector based on the word form and the meaning vector of the incorrectly predicted word, and the column r -target is the correlation between the vector of the predicted semantic vector of the word form and the meaning vector of the target word. The Error column shows in which semantic vector the error occurs.

In the first line in table 4, the target verb form *yazimuruzwaga* is the *iterative* form of the third person plural, past, imperfective, passive of the lexeme *kuzimira* “to be lost”, LDL predicted the verb form *yazimuruzwaga*, which is the *reversive* form of the target form. In

the second line the target form is the third person, plural, active, plural, subjunctive of the *neutral* form of the lexeme *guhora* “to calm” for which LDL predicted the *applicative*. The target form in the third line is the third person, plural, present tense, passive, *reversive* of the verb *kuzimira* “to be lost”, and the predicted form is the *iterative* form. The target form in the fourth line is the second person, singular, *present*, passive, comitative, *subjunctive* of the lexeme *kuyobora* “to lead”, while the predicted form differs in being *past* and in being *indicative*. The target form in the fifth line again involve errors between the *reversive* and the *iterative* form of the third person singular, passive, imperfective, indicative of the lexeme *kuzimira* “to be lost”.

Target	Predicted	r	r-target	Error
yazimururuzwaga	yazimuruzwaga	0.95727	0.95725	Extension
bahore	bahore	0.9641	0.9640	Extension
barazimuruzwa	barazimururuzwa	0.9587	0.9586	Extension
wayoboranwe	uyoboranwe	0.9308	0.9305	Voice, Mood
yazimuruzwaga	yazimururuzwaga	0.9570	0.9566	Extension

Table 4: For the model with letter trigrams as cues, the five errors with the smallest different between the correlation with the semantic vector of the erroneously predicted word (r), and the correlation with the semantic vector of the target word (r -target).

There were 11104 correct and 424 false predictions (96.3% correct)

The pattern of errors between *iterative* and *reversive* is also found in the errors in the model that is based on bigrams of syllables (table 5). The target of the first form in table 5, *no.gu.ru.ye*, is the *iterative* of the *reversive* predicted form. There is one cause of substitution, *a.zi.mi.ra.gu.zwe* is repdicted to be *nzi.mi.ra.gu.zwe*

target	predicted	r	r-target	Error
no.gu.ru.ye	no.gu.ye	0.9561332	0.9561068	Extension
two.gu.ra	two.gu.ru.ra	0.9566765	0.9565960	Extension
two.gu.ra.ga	two.gu.ru.ra.ga	0.9575569	0.9570913	Tense, Aspect
bo.gu.ru.ye	bo.gu.ye	0.9533335	0.9528527	Extension
tu.ro.gu.ra	tu.ro.gu.ru.ra	0.9583053	0.9576806	Extension

Table 5: For the model with syllable bigrams as cues, the five errors with the smallest different between the correlation with the semantic vector of the erroneously predicted word (r), and the correlation with the semantic vector of the target word (r -target). There were 11357 correct and 171 false predictions (98.6% correct) .

Table 6 provides an overview of the 5 errors that were closest in meaning to the intended form in the model which used trigrams of syllables. As in the previous tables the distance between the target of the predicted form is usually very small.

As concerns the errors in production we inspected whether the best word form is among the best 15 candidates of the ones predicted by LDL. In case of the production (build) measure, this is often not the case: for the model that used trigrams of letters this was true in 55.6% of the cases, in the model that used bigrams of syllables that was true in 25.4% of the cases and in the model that used trigrams of syllables that was true in a dismal 0.01% of the cases. A production (build) error is often serious and likely an impediment to successful communication.

Target	Form	r	r-target	Error
ni.mu.ba.ga.gu.re	ba.ga.gu.ra	0.9670884	0.9610968	Number
tu.ra.su.ra	tu.ra.su.ru.ra	0.9045014	0.8984885	Extension
na.ku.bi.ta.nwe	nku.bi.ta.nwa	0.9285687	0.9223399	Tense
ba.za.ba.ba.ku.bi.ti.rwa	na.ku.bi.ti.we	0.9851268	0.9381847	Person
ba.za.ba.ba.ku.bi.ti.shwa	na.ku.bi.ti.shi.jwe	0.9843883	0.9371096	Person

Table 6: For the model with syllable trigrams as cues, the five errors with the smallest difference between the correlation with the semantic vector of the erroneously predicted word (r), and the correlation with the semantic vector of the target word (r -target). In total there were 11500 correct and 28 false predictions (accuracy 99.8%).

In case of production (learn) the errors are likely less serious. For the model that used trigrams of letters this was true in 80.0% of the cases, in the model that used bigrams of syllables that was true in 94.1% of the cases and in the model that used trigrams of syllables that was true in 47.7% of the cases. A production (learn) error is often serious in a model based on trigrams of syllables, but much less so in models build on trigrams of letters and bigrams of syllables.

Accuracy	letter trigrams	σ -bigrams	σ -trigrams
Production (build)	55.6%	25.4%	0.01%
Production (learn)	80.0%	94.1%	47.4%

Table 7: Presence of the correct verb form among the 15 best candidates that LDL predicted.

As can be seen in table 8, comprehension was very accurate for all models. The accuracy of production ranged from good (letter trigrams) to stellar (trigrams of syllables).

Accuracy	letter trigrams	σ -bigrams	σ -trigrams
Comprehension	96.3%	98.6%	99.8%
Production (build)	78.8%	89.3%	99.9%
Production (learn)	82.0%	94.5%	99.5%

Table 8: Overview of the accuracy of the models for all data.

4.2 Comprehension and production of held-out data

The models presented so far used all data from the data set, and even though, as we have argued above, this is not an unreasonable assumption, we also need to know how the model would do if it encounters forms it does not know. In order to assess this, we carefully split the data in 90% for training and assessed the accuracy on the basis of the remaining held-out 10% of the data, but we made sure the held-out data did not contain any phonological ngrams that were not part of the training set. This is because LDL cannot create those ngrams from scratch, and an ngram that is not in the training data would stump the algorithm. In real life such unknown ngrams would likely be reinterpreted in terms of known ngrams, as speakers do when confronted with loan words with unknown sounds or phonotactics (Daland et al. 2019; Rwamo & Ntiranyibagira 2020). The data

reported here is the accuracy on the held-out data, with the accuracy of the trained data in brackets.

For the model in which the word forms were represented as trigrams of letters, the comprehension accuracy of the 10% held-out data was 98.2% (and for the 90% of the seen data the accuracy was 95.8%). For the model in which the word forms were represented as bigrams of syllables, the comprehension accuracy was 98.9% (97.9%). For the model in which the word forms were represented as trigrams of syllables, the comprehension accuracy was 98.9% (98.8%). Across all three models comprehension was excellent, even for held-out data.

The accuracy of production (build) for the model in which the word forms were represented as trigrams of letters, was 53.6% (77.1%). For the model in which the word forms were represented as bigrams of syllables, it was 42.8% (89%). For the model in which the word forms were represented as trigrams of syllables, it was 7.2% (99.1%). The performance of the syllable models on this measure of accuracy is clearly worse than the letter based model.

The accuracy of production (learn) for the model in which the word forms were represented as trigrams of letters was 67% (82.2%). For the model in which the word forms were represented as bigrams of syllables, it was 84.9% (92.6%). For the model in which the word forms were represented as trigrams of syllables, it was 80.4% (98.2%).

It turns out that comprehension is excellent, even for unseen forms. As for production, the build algorithm is worse for all models than the learn model. It is also clear that in the build algorithm trigrams of syllables cannot be put to use to produce new forms, as only 7% will be correct. Even though the other models are better in this respect, but they can still not really be used to produce new forms. The learn algorithm, though, achieves good results, ranging from 67% correct predictions for the model of letter trigrams, 80.4% for trigrams of syllables, and 84.9% for bigrams of syllables. Table 9 gives an overview of all accuracies.

Accuracy	letter trigrams	σ -bigrams	σ -trigrams
Comprehension	98.2%	98.9%	98.9%
Production (build)	53.6%	42.8%	07.2%
Production (learn)	67.0%	84.9%	80.4%

Table 9: Overview of the accuracy of the models for comprehension and production of held-out data.

The modeling of the held-out data allows us to further differentiate among the models. The models of all data showed that trigrams of syllables performed best, but the models of the held-out data show that this is a consequence of overfitting the data. The model using bigrams of syllables performs well even for the held-out data, and much better than the model based on trigrams of letters.

4.3 Comprehension and production of verbs with meanings derived from word-embeddings

In the data set of 573 verbs with representations of meanings that are derived from word-embeddings, the comprehension accuracy of the model in which the word forms were represented as trigrams of letters is 82%; the comprehension accuracy of the model in which the word forms were represented as bigrams of syllables is 92.7%; the compre-

Accuracy	letter trigrams	σ -bigrams	σ -trigrams
Comprehension	82%	92.7%	98.6%
Production (build)	75.6%	86.2%	91.8%
Production (learn)	89.7%	97.4%	99.5%

Table 10: Overview of the accuracy of the models for comprehension and production of verbs with meaning derived from word-embeddings.

Accuracy	letter trigrams	σ -bigrams	σ -trigrams
Comprehension	98.3%	96.5%	100%
Production (build)	42.1%	15.8%	22.8%
Production (learn)	56.1%	43.9%	80.7%

Table 11: Overview of the accuracy of the models for comprehension and production of verbs with meaning derived from word-embeddings for held-out data.

hension accuracy of the model in which the word forms were represented as trigrams of syllables is 98.6% (see table 10).

The production (build) accuracy of the model in which the word forms are represented as trigrams of letters is 75.6%; of the model in which word forms are represented as bigrams of syllables is 86.2%; of the model in which word forms are represented as trigrams of syllables is 91.8% (see table 10).

The production (learn) accuracy of the model in which the word forms are represented as trigrams of letters is 89.7%; of the model in which word forms are represented as bigrams of syllables is 97.4%; of the model in which word forms are represented as trigrams of syllables is 99.5% (see table 10).

The accuracy achieved in this small data set in all models is excellent as well, even though slightly worse than the accuracy of the models that use the entire data set of 11528 verbs and rely on hand-annotated meanings. This is most likely a consequence of the size of the smaller data set of the verbs with meanings based on embeddings.

These results show that meanings of Kinyarwanda words can be learned from the distribution of words in utterances. In a model lexical processing of English words using LDL the semantic vectors were given extra inflectional lexemes (Baayen et al. 2019b). This was not done for Kinyarwanda (Niyongabo et al. 2020), nevertheless LDL was successful in modeling comprehension and production on the basis of verb meanings derived from word embeddings. A linguistic factor that probably has been helpful to achieve this is the great amount of overt agreement markers in Kinyarwanda, that can help determine the inflectional properties of a word (Wal 2015).

We also tested the accuracy of LDL for verbs with meanings based on word-embeddings with held-out data. Of course, since the data set is small, the held-out data set is tiny and consisted of only 60 verb forms. The results are given in Table 11. Again, comprehension is excellent, and across all models the production (learn) algorithm produced more accurate results than the production (build) algorithm.

4.4 Discussion

We modeled two data sets of Kinyarwanda verbs. One consists of 11528 Kinyarwanda verb forms. The verb forms and their grammatical functions were manually annotated. The other in data consisted of 573 verbs of the full data set for which word embeddings

were available. In short, our modelings show that Kinyarwanda verbs can be comprehended and produced accurately. This is true for the data set in which we simulated meaning vectors, as well as in the data set in which the meaning was derived from word embeddings. All models have excellent comprehension. Production is less accurate than comprehension. The production (learn) algorithm has better accuracy than the production (build) algorithm.

Overall the syllable-based models have better accuracy than the models based on letter trigrams, except for the production (build) accuracy on held-out data, in which the model based on trigrams of syllables performs dismally. A comparison with the accuracy of this model for this algorithm on all data suggest that for production (build) the algorithm overfits.

In section 5 we will put our results in the perspective of the Discriminative Lexicon theory and the implications of our findings for our understanding of the language system.

5 Conclusion

In this paper we set out to investigate the hypotheses of the Discriminative Lexicon for comprehension and production of verbs in Kinyarwanda, a highly agglutinative language. To do this, we modeled comprehension and production of verbs in LDL, a computational implementation of the Discriminative Lexicon. It turns out that both comprehension and production can be modeled with great accuracy in LDL. Our results provide support for various hypotheses of the Discriminative Lexicon.

Our results lend support to word-based perspective on morphology of the Discriminative Lexicon (Blevins 2016b). It is supported, because our modeling is based on words and comprehension and production is excellent. Even in agglutinative languages such as Kinyarwanda, which are often thought to be prime examples for the compositional nature of complex words (Katamba 1978), comprehension and production does not rely on morphemes, stems or exponents.

Our results also support the hypothesis, present in both the word and paradigm theory and distributional semantics, that words get their meaning in utterances (Firth 1957; Landauer & Dumais 1997; Sahlgren 2008; Weaver 1955). The Discriminative Lexicon theory goes beyond the assumption of word and paradigm theory, since the relevant distributions are not limited to the paradigm, but are assessed over the entire lexicon (Baayen et al. 2019b). The grammatical function first person, for example, represents all words that occur in utterances where they are intended to convey the meaning of first person, and this meaning can be gleaned from this distribution (Landauer & Dumais 1997). The distributional hypothesis is backed up by the results in both data sets: In the the smaller data set the meanings are based on word embeddings, which are derived from the way in which these words are distributed in utterances. In our larger data set the grammatical functions that we annotated by hand provide a distributional structure to the data set.

Error-driven learning, as implemented in LDL by the mathematics of multivariate multiple regression, is very effective at learning comprehension and production of Kinyarwanda verbs. The findings concerning the production (learn) underline this importance. The algorithm in the JudiLing implementation of LDL that generates predictions for this measure assesses which phonological units are best supported (in terms of learning) in which position of a word to express an intended meaning. In all our simulations this measure was very accurate.

Our modeling also confirms that a fully connected network of two matrices is able to learn comprehension and production of complex words very well. This is in line with a great deal of research in which shallow (Baayen 2011; Baayen et al. 2016; 2018; 2019b; Chuang et al. 2020a; c; b; Tomaschek et al. 2019) or deep neural networks (Hahn et al. 2020; Futrell et al. 2020; Linzen 2019; Magnuson et al. 2020; Malouf 2017; Pater 2019; Prickett et al. 2018) evaluate a range of theoretical questions. However, the shallow networks are preferable, since their workings can be assessed, and rely on linear algebra, while the hidden layers of a deep network are virtually inaccessible and partly rely on complex non-linear algebra (Arras et al. 2016; Baayen et al. 2019b).

The production accuracy is lower in all models than the comprehension accuracy, but this is to be expected. There are only a few grammatical functions that have to be mapped onto a great number of possible phonological units of possible words. This makes the discriminative association between phonological units and meaning less strong. The lower production accuracy matches the difficulty reported in language acquisition for the production of word forms in comparison to their comprehension (Pater 2004).

We evaluated what phonological representation best discriminates among meaning in three different LDL models. In one model the phonology of a word was represented by means of trigrams of letters. As the Kinyarwanda orthography is almost phonemic, the orthography is a reasonable approximation of its phoneme system. In another models the phonology was represented by bigrams of syllables, and in a final model the phonology was represented as trigrams of syllables. All models did well for comprehension. For production in the model which tested all data, the syllable based models were better than the letter based models, and the trigram syllable model was better than the bigram syllable model. This is also true for the data set in which the meaning of verbs was represented by word embeddings. This suggests that syllables would be more useful discriminative units than letter trigrams, which is what can be expected on the basis of the phonotactics of Kinyarwanda. For the held-out data the model based on bigrams of syllables was better than the model of trigrams of letters and of trigrams of syllables. The size of these discriminative units is reminiscent of binary or ternary feet (Elenbaas & Kager 1999; Hayes 1995; Martínez-Paricio & Kager 2015), which have also been reported for Kinyarwanda (Goldsmith & Mpiranya 2010). Phonological stretches of two or three partly overlapping syllables is well-suited to partly discriminate the meaning of the whole word. For stretches of three letters there are more identical stretches which discriminate several meanings and which therefore are slightly less discriminative for the meaning of the whole word. These phonological cues to meaning are learned during acquisition by a process of cue competition (Nixon 2020; Nixon & Tomaschek 2021; Ramscar et al. 2013a) and are therefore language specific.

Our results show that successful comprehension and production of complex word forms does not depend on knowledge of morphemes, stems or exponents. Yet, morphemes and stems play a crucial role in recent work on the order of information in the Kinyarwanda verbs. Banerjee (2019) proposes that syntactic selection of morphemes can account for the order of the morphemes in the template of Kinyarwanda. The syntactic selection operates on morphemes and puts them in order on the syntactic graph. If we are correct in concluding that Kinyarwanda speakers rely on words rather than on morphemes, this raises the question as to where the order of the information in Kinyarwanda verbs comes from.

Our answer to this question is that Kinyarwanda children learn which information discriminates best for which meaning by learning to understand and learning to use the words in utterances that they hear. In usage-based theories using a language spurs learn-

ing a language (Ambridge 2020; Bybee 2001; Ellis & Ogden 2017; Kapatsinski 2018; Tomasello 2003) by inducing a grammar from observable data (Abney 2021). Language users do so by relying on general learning mechanisms, by assuming that language forms are intended to convey meaning, and by assuming lexicon and grammar are inseparable (Ellis & Ogden 2017). These hypotheses fit naturally in the Discriminative Lexicon. Children hear the words uttered with a certain order of information, infer their meaning from their distribution in utterances, and use this knowledge for their own comprehension and production. The learning mechanism would in that case be discriminative learning.

This, in turn, begs the question as to where the order has come from in the first place. An answer is that it is the result of the history of the language, and reflects the pressures of effective communication. (Futrell et al. 2020; Hahn et al. 2020). Hahn et al. (2020) observe that human languages share many properties and offer an explanation in terms of communicative pressures, and support their hypothesis by computational evidence. These communicative pressures require speakers to be succinct, so that a meaning can be expressed easily, and unambiguous, so that a listener can easily understand the intended meaning. Such pressures can be resolved in different ways in different situations, which leads to variation in the distributions of words in utterances. These words, then acquire different shades of meaning. This, in turn, would then lead to differences in comprehension and production—with potential consequences for the order of morphemes in a language if there is enough variation. Variation is the bedrock of grammatical change (Blevins & Garrett 2004). The fact that the order in templates in Bantu languages varies slightly from language to language (Banerjee 2019) is in line with this explanation.

In short, our findings support a Discriminative Lexicon view of the mental lexicon, in which comprehension and production of complex words is learned discriminatively on the basis of whole words. The meaning of word forms is based on their distribution in utterances. Our findings can easily be linked up to proposals that the order of information in a word is a result of the workings of communicative pressures (Kapatsinski 2018).

Obviously there remains a great deal of work ahead of us. Because of the richness of the Kinyarwanda verb, we have only modeled a fraction of its complexity, and there is still work to be done. To give two concrete examples. We ignored vowel length, and did not include tonal information (Goldsmith & Mpiranya 2010). Adding vowel length would not have changed our analysis, or our results, because a long vowel is always long in the same syllable in a word, irrespective of its meaning. As to tones, our data set does not include information that would introduce tones, that are distributed differently in different members of the paradigm (Goldsmith & Mpiranya 2010). Including those affixes would make an even more ecologically valid data set, and is something we will pursue in the future.

One topic that requires further work, but is beyond the scope of our paper, concerns the acquisition of comprehension and production. Our model describes comprehension and production of adults or young children, with access to meanings through the distributions of words in utterances. It leaves open the question as to how infants acquire this knowledge. Infants are exposed to utterances and they obviously learn the morphology of any language. Even though much research remains to be done there is some evidence that this, too, can be learned discriminatively. Nixon & Tomaschek (2021) found that an error-driven computational model of the acquisition of phonetic categories by infants correlates very well with the results of behavioral experiments. Ramscar et al. (2013b) shows that error-driven learning predicts the typical overregularizations in children's speech well.

A methodological issue that needs more research concerns the architecture of the computational network (see Pirrelli 2018: for an overview of computational models for morphology). LDL uses a network of only two dense layers to model comprehension and production. An unresolved question is whether this is enough. Speech recognition has been modeled with more than two layers (Beguš, Gašper 2021) with only modest success, whereas models with only two dense layers are more successful (Arnold et al. 2017; Shafaei-Bajestan et al. 2020). Multilingual acquisition has also successfully been modeled with only two layers (Chuang et al. 2021), as well as several other lexical processing phenomena (Baayen et al. 2019b; Baayen & Smolka 2020; Chuang et al. 2020c; Tomaschek et al. 2019). However, Boersma et al. (2020) argue that it is necessary to assume more layers to model phonetic and phonological knowledge. This is because each level of representation is modeled as one layer, and Boersma et al. (2020) distinguish 4 different layers: the sensorimotor layer, the cue layer, the faithfulness layer and the lexical layer. Many of the assumptions that justify these layers rely on the assumption that phonology ultimately seeks to discover discrete units in the continuous phonetic signal. Even though more research needs to be carried out to adjudicate this question, the fact that word recognition of isolated words from spontaneous speech can be modeled well with only two layers, suggests that phonological knowledge can be modeled with a dense two-layer network.

There is also work in morphology that makes use of a network with more than two layers. In work on the paradigm cell filling problem Malouf (2017) shows that a recurrent neural net is able to fill each cell in a paradigm with a form on the basis of a few other forms in the paradigm for a great number of languages. Yet, it remains unclear whether speakers of a language solve the paradigm cell filling problem in the same way as the neural net does. Our work shows that a model with two layers suffices in morphology, and in order to be able to compare grammars, we need to model the history of a language. A two-layer network is more easily to be interpreted cognitively and would be preferable.

Even though we are aware that much work lies ahead of us, on the basis of these results we feel justified to conclude that comprehension and production of Kinyarwanda verb forms can be modeled successfully without recourse to morphemes, which lends support to the various hypotheses of the theory of the Discriminative Lexicon (Baayen 2011; Baayen et al. 2016; 2018; 2019b; Chuang et al. 2020b; 2021; 2020c). It also provides a straightforward account of comprehension and production of complex words: we can learn to comprehend and produce complex words by figuring out what people mean when they use words.

Funding information (optional)

Emmanuel Uwambayinema is supported by the DAAD, which we gratefully acknowledge. This research was funded by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 ‘Spoken Morphology’, Projects VI 223/3-1, VI 223/3-2), which we gratefully acknowledge.

Acknowledgements (optional)

Thank you: Harald Baayen, Xuefeng Luo, Jessica Nieder, Fabian Tomaschek and the audience of the colloquium of Research Unit 2373 *Spoken Morphology* funded by the Deutsche Forschungsgemeinschaft.

Competing interests (mandatory)

The author(s) has/have no competing interests to declare.

Authors' contributions (optional)

RvdV: conception, writing, analysis

EU: conception, writing, analysis, data

YYC: analysis, coding, writing

References

- Abney, Steven. 2021. Inductive general grammar. *Glossa: a journal of general linguistics* 6(1). <https://doi.org/10.5334/gjgl.1332>.
- Adriaans, F. 2011. The induction of phonotactics for speech segmentation: Converging evidence from computational and human learners. *LOT* 267.
- Adriaans, F. & R. Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62(3). 311–331.
- Alderete, John & Monica Davies. 2019. Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings. *Language and speech* 62(2). 281–317.
- Ambridge, Ben. 2020. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40(5-6). 509–559.
- Arnold, Denis, Fabian Tomaschek, Konstantin Sering, Florence Lopez & R. Harald Baayen. 2017. Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLoS ONE* 12(4). e0174623. <https://doi.org/https://doi.org/10.1371/journal.pone.0174623>.
- Arras, Leila, Franziska Horn, Grégoire Montavon, Klaus-Robert Müller & Wojciech Samek. 2016. Explaining predictions of non-linear classifiers in nlp. *arXiv preprint arXiv:1606.07298*.
- Baayen, R. H., Yu-Ying Chuang & Maria Heitmeier. 2019a. *Wpmwithldl: Implementation of word and paradigm morphology with linear discriminative learning*. R package version 1.3.17.1.
- Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics* 11. 295–328.
- Baayen, R Harald, Yu-Ying Chuang & James P Blevins. 2018. Inflectional morphology with linear mappings. *The mental lexicon* 13(2). 230–268.
- Baayen, R Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P Blevins. 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* 2019.
- Baayen, R Harald, Cyrus Shaoul, Jon Willits & Michael Ramscar. 2016. Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, cognition and neuroscience* 31(1). 106–128.

- Baayen, R. Harald & Eva Smolka. 2020. Modeling Morphological Priming in German With Naive Discriminative Learning. *Frontiers in Communication* 5. 17. <https://doi.org/10.3389/fcomm.2020.00017>.
- Banerjee, Neil. 2019. Templatic morphology through syntactic selection: Valency-changing extensions in Kinyarwanda. *Glossa* 4(1). 112. <https://doi.org/http://doi.org/10.5334/gjgl.920>.
- Bauer, Laurie. 2016. Classical morphemics: Assumptions, extensions and alternatives. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 331–355. Cambridge University Press. <https://doi.org/10.1017/9781139814720.013>.
- Beguš, Gašper. 2021. CiwGAN and fiwGAN: Encoding information in acoustic data to model lexical learning with Generative Adversarial Networks. *Neural networks : the official journal of the International Neural Network Society* 139. 305–325. <https://doi.org/10.1016/j.neunet.2021.03.017>.
- Blevins, J. & A. Garrett. 2004. The evolution of metathesis. In B. Hayes, Robert M. Kirchner & D. Steriade (eds.), *Phonetically Based Phonology*, 117–156. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486401.005>.
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42(3). 531–573.
- Blevins, James P. 2016a. *The minimal sign* 50–69 (Cambridge Handbooks in Language and Linguistics) Cambridge University Press. <https://doi.org/10.1017/9781139814720.003>.
- Blevins, James P. 2016b. *Word and paradigm morphology*. Oxford University Press.
- Boersma, Paul, Titia Benders & Klaas Seinhorst. 2020. Neural network models for phonology and phonetics. *Journal of Language Modelling* 8(1). 103–177. <https://doi.org/10.15398/jlm.v8i1.224>.
- Boersma, Paul & B. Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1). 45–86.
- Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Booij, Geert. 2012. *The Grammar of Words: An Introduction to Linguistic Morphology* (Oxford Textbook in Linguistics). Oxford, UK and Cambridge, MA: Oxford University Press 3rd edn.
- Booij, Geert. 2016. Construction Morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 424–448. Cambridge University Press. <https://doi.org/10.1017/9781139814720.016>.
- Brysbaert, Marc, Michaël Stevens, Paweł Manderla & Emmanuel Keuleers. 2016. How many words do we know? practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in psychology* 7. 1116.
- Bybee, Joan. 2001. Usage-Based Phonology. In Charles W. Kreidler (ed.), *Phonology: Critical Concepts in Linguistics*, 388–411. London: Routledge.
- Chuang, Yu-Ying, Melanie J Bell, RH Baayen et al. 2020a. Bilingual and multilingual mental lexicon: a modeling study with linear discriminative learning .
- Chuang, Yu-Ying, Melanie J Bell, Isabelle Banke & R Harald Baayen. 2021. Bilingual and multilingual mental lexicon: a modeling study with linear discriminative learning. *Language Learning* 71(S1). 219–292.
- Chuang, Yu-Ying, Kaidi Lõo, James P. Blevins & R. Harald Baayen. 2020b. *Estonian Case Inflection Made Simple: A Case Study in Word and Paradigm Morphology with Linear Discriminative Learning* 119–141. Cambridge University Press. <https://doi.org/10.1017/9781108780643.008>.

- Chuang, Yu-Ying, Marie Lenka Vollmer, Elnaz Shafaei-Bajestan, Susanne Gahl, Peter Hendrix & R. Harald Baayen. 2020c. The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods* 1–32. <https://doi.org/10.3758/s13428-020-01356-w>.
- Creissels, Denis. 2019. Morphology in niger-congo languages. <https://doi.org/10.1093/acrefore/9780199384655.013.535>. <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-535>.
- Daland, Robert, Mira Oh & Lisa Davidson. 2019. On the relation between speech perception and loanword adaptation. *Natural Language & Linguistic Theory* 37(3). 825–868. <https://doi.org/10.1007/s11049-018-9423-2>.
- Demuth, K. 2000. Bantu noun class systems: loanword and acquisition evidence of semantic productivity. *Systems of Nominal Classification* 270–292.
- Elenbaas, Nine & René Kager. 1999. Ternary rhythm and the lapse constraint. *Phonology* 273–329.
- Ellis, Nick C & Dave C Ogden. 2017. Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science* 9(3). 604–620.
- Firth, John R. 1957. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.
- Futrell, Richard, Roger P Levy & Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language* 96(2). 371–412. <https://doi.org/10.1353/lan.2020.0024>.
- Goldsmith, John A., Jackson L. Lee & Aris Xanthos. 2016. Computational Learning of Morphology. *Annual Review of Linguistics* 3(1). 85–106. <https://doi.org/10.1146/annurev-linguistics-011516-034017>.
- Goldsmith, John A & Fidèle Mpiranya. 2010. Rhythm, quantity, and tone in the kinyarwanda verb. *Tones and features: Phonetic and phonological perspectives*.
- Goldwater, Sharon, Thomas L. Griffiths & Mark Johnson. 2009. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112(1). 21–54. <https://doi.org/10.1016/j.cognition.2009.03.008>.
- Hahn, Michael, Dan Jurafsky & Richard Futrell. 2020. Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences* 117(5). 2347–2353. <https://doi.org/10.1073/pnas.1910923117>.
- Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia linguistica* 45(1). 31–80.
- Haspelmath, Martin. 2020. The morph as a minimal linguistic form. *Morphology* 30. 117–134.
- Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Hayes, B. 2004. Phonological acquisition in optimality theory: The early stages. In René Kager, Joe Pater & Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 158–203. Cambridge, UK: Cambridge University Press.
- Heyes, Cecilia. 2012. Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1603). 2695–2703.
- Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10(2-3). 210–234.

- Hyman, Larry M. 2003. Suffix ordering in bantu: A morphocentric approach. In *Yearbook of morphology 2002*, 245–281. Springer.
- Hyman, Larry M & Sharon Inkelas. 2017. Multiple exponence in the lusoga verb stem. In Claire Bower, Laurence Horn & Rafaella Zanuttini (eds.), *On looking into words (and beyond)*, 171–188. Language Science Press Berlin.
- Kapatsinski, Vsevolod. 2018. *Changing minds changing tools: From learning theory to language acquisition to language change*. MIT Press.
- Karlsson, Fred. 1986. Frequency considerations in morphology. *Sprachtypologie und Universalienforschung* 39(1). 19–28.
- Katamba, Francis. 1978. How agglutinating is bantu morphology? *Linguistics* 16(210). 77–84.
- Keating, P.A. 1985. Universal phonetics and the organization of grammars. *Phonetic Linguistics* 115–132.
- Kimenyi, Alexandre. 1979. *Studies in Kinyarwanda and bantu phonology*, vol. 33. Linguistic Research.
- Landauer, Thomas K & Susan T Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review* 104(2). 211.
- Linzen, Tal. 2019. What can linguistics and deep learning contribute to each other? response to pater. *Language*.
- Lõo, Kaidi, Juhani Järvi, Fabian Tomaschek, Benjamin V Tucker & R Harald Baayen. 2018. Production of estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 28(1). 71–97.
- Luo, Xuefeng. 2021. *Judiling: An implementation for discriminative learning in julia*. Germany: Eberhard Karls University of Tübingen MA thesis. https://github.com/MegamindHenry/JudiLing.jl/blob/master/thesis/thesis_JudiLing_An_implementation_for_Discriminative_Learning_in_Julia.pdf.
- Luo, Xuefeng, Yu-Ying Chuang & Harald Baayen. 2021. Judiling: an implementation in julia of linear discriminative learning algorithms for language models. <https://github.com/MegamindHenry/JudiLing.jl/blob/master/docs/src/index.md>. <https://github.com/MegamindHenry/JudiLing.jl/blob/master/docs/src/index.md>.
- Magnuson, James S., Heejo You, Sahil Luthra, Monica Li, Hosung Nam, Monty Escabá, Kevin Brown, Paul D. Allopenna, Rachel M. Theodore, Nicholas Monto & Jay G. Rueckl. 2020. Earshot: A minimal neural network model of incremental human speech recognition. *Cognitive Science* 44(4). e12823. <https://doi.org/10.1111/cogs.12823>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12823>.
- Malouf, Robert. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology* 27(4). 431–458.
- Martínez-Paricio, Violeta & René Kager. 2015. The binary-to-ternary rhythmic continuum in stress typology: Layered feet and non-intervention constraints. *Phonology* 459–504.
- Matthews, Peter Hugoe & PH Matthews. 1972. *Inflectional morphology: A theoretical study based on aspects of latin verb conjugation*, vol. 6. CUP Archive.
- McQueen, James M. 2007. Eight questions about spoken-word recognition. *The Oxford handbook of psycholinguistics* 37–53.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. *arXiv*.
- Milin, Petar, Victor Kuperman, Aleksandar Kostic & R Harald Baayen. 2009. Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In James P Blevins & Juliette Blevins (eds.), *Analogy*

- in Grammar: Form and Acquisition*, 214–252. Oxford University Press Oxford.
- Milin, Petar, Eva Smolka & Laurie Beth Feldman. 2017. Models 11 of lexical access and morphological processing. *The Handbook of Psycholinguistics* 240.
- Mitterer, Holger, Eva Reinisch & James M McQueen. 2018. Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language* 98. 77–92.
- Myers, Scott. 2005. Vowel duration and neutralization of vowel length contrasts in Kinyarwanda. *Journal of Phonetics* 33(4). 427 – 446. <https://doi.org/https://doi.org/10.1016/j.wocn.2005.02.002>. <http://www.sciencedirect.com/science/article/pii/S0095447005000148>.
- Ndlovu, Mbulisi & Progress Dube. 2019. Extended exponence in isindebele morphology. *Stellenbosch Papers in Linguistics Plus (SPiL Plus)* 56. 47–57.
- Nixon, Jessie S. 2020. Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition* 197. 104081.
- Nixon, Jessie S. & Fabian Tomaschek. 2021. Prediction and error in early infant speech learning: A speech acquisition model. *Cognition* 212. 104697. <https://doi.org/10.1016/j.cognition.2021.104697>.
- Niyongabo, Rubungo Andre, Hong Qu, Julia Kreutzer & Li Huang. 2020. KINNEWS and KIRNEWS: Benchmarking Cross-Lingual Text Classification for Kinyarwanda and Kirundi. *arXiv*.
- Nurse, Derek & Gérard Philippson. 2006. *The Bantu Languages*. London and New York: Routledge.
- Olejarczuk, Paul, Vsevolod Kapatsinski & R Harald Baayen. 2018. Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard* 4(s2).
- Pater, J. 2004. Bridging the gap between receptive and productive development with minimally violable constraints. In Rene Kager, Joe Pater & Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 219–244. Cambridge, UK: Cambridge University Press.
- Pater, Joe. 2019. Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95(1). <https://doi.org/10.1353/lan.2019.0009>. <http://par.nsf.gov/biblio/10123481>.
- Pham, Hien & Harald Baayen. 2015. Vietnamese compounds show an anti-frequency effect in visual lexical decision. *Language, Cognition and Neuroscience* 30(9). 1077–1095.
- Pirrelli, Vito. 2018. Morphological theory and computational linguistics. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 572–593. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.32>.
- Prickett, Brandon, Aaron Traylor & Joe Pater. 2018. Seq2seq models with dropout can learn generalizable reduplication. In *Proceedings of the fifteenth workshop on computational research in phonetics, phonology, and morphology*. 93–100.
- Prince, A. & B. Tesar. 2004. Learning phonotactic distributions. In René Kager, Joe Pater & Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 245–291. Cambridge, UK: Cambridge University Press.
- Ramscar, Michael. 2019. Source codes in human communication. <https://doi.org/10.31234/osf.io/e3hps>. [psyarxiv.com/e3hps](https://arxiv.org/abs/1905.08282).
- Ramscar, Michael, Melody Dye & Joseph Klein. 2013a. Children value informativity over logic in word learning. *Psychological science* 24(6). 1017–1023.

- Ramscar, Michael, Melody Dye & Stewart M McCauley. 2013b. Error and expectation in language learning: The curious absence of "mouses" in adult speech. *Language* 760–793.
- Ramscar, Michael & Nicole Gitcho. 2007. Developmental change and the nature of learning in childhood. *Trends in cognitive sciences* 11(7). 274–279.
- Ramscar, Michael & Daniel Yarlett. 2007. Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive science* 31(6). 927–960.
- Rescorla, Robert A. 1988. Pavlovian conditioning: It's not what you think it is. *American psychologist* 43(3). 151.
- Rescorla, Robert A & Allan R Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.B.W. Prokasy (ed.), *Classical conditioning ii: Current research and theory*, vol. 2, 64–99. New York: Appleton Century Crofts.
- Rwamo, Alice & Constantin Ntiranyibagira. 2020. Phonological and perceptual factor symbiosis in loanword adaptation. *Revista Odisseia* 5(1). 22–39.
- Saffran, Jenny R, Richard N Aslin & Elissa L Newport. 1996. Statistical learning by 8-month-old infants .
- Sahlgren, Magnus. 2008. The distributional hypothesis. *Italian Journal of Disability Studies* 20. 33–53.
- Schmid, Helmut. 1999. Improvements in part-of-speech tagging with an application to german. In *Natural language processing using very large corpora*, 13–25. Springer.
- Shafaei-Bajestan, Elnaz, Masoumeh Moradipour-Tari, Peter Uhrig & R H Baayen. 2020. LDL-AURIS: Error-driven Learning in Modeling Spoken Word Recognition <https://doi.org/10.31234/osf.io/v6cu4>.
- Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax-morphology interface*, vol. 149. Cambridge University Press.
- Stump, Gregory. 2018. Paradigm function morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 284–304. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.23>.
- Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure* (Cambridge Studies in Linguistics). Cambridge University Press.
- Tesar, B. & P. Smolensky. 1998. Learnability in optimality theory. *Linguistic Inquiry* 29(2). 229–268.
- Tessier, Anne-Michelle. 2019. U-shaped development in error-driven child phonology. *Wiley Interdisciplinary Reviews: Cognitive Science* 10(6). e1505. <https://doi.org/10.1002/wcs.1505>.
- Tomaschek, Fabian, Ingo Plag, Mirjam Ernestus & Harald Baayen. 2019. Phonetic effects of morphology and context: Modeling the duration of word-final s in English with naïve discriminative learning. *Journal of Linguistics* 1–39. <https://doi.org/10.1017/S0022226719000203>.
- Tomasello, Michael. 2003. *Constructing a language*. Cambridge, MA: Harvard University Press.
- Wal, Jenneke van der. 2015. Bantu Syntax, <https://doi.org/10.1093/oxfordhb/9780199935345.013.50>.
- Walker, Rachel, Dani Byrd & Fidèle Mpiranya. 2008. An articulatory view of kin-yarwanda coronal harmony. *Phonology* 25(3). 499–535. <https://doi.org/10.1017/S0952675708001619>.
- Weaver, Warren. 1955. Translation. *Machine translation of languages* 14(15-23). 10.

- Widrow, Bernard & Marcian E Hoff. 1960. Adaptive switching circuits. Tech. rep. Stanford Univ Ca Stanford Electronics Labs.
- Zorc, R. David & Louise Nibagwire. 2007. *Kinyarwanda and Kirundi Comparative Grammar*. Hyattsville, MD: Dunwoody Press.
- Zwitserslood, Pienie. 2018. Processing and representation of morphological complexity in native language comprehension and production. In *The construction of words*, 583–602. Springer.