Ruben van de Vijver*, Emmanuel Uwambayinema and Yu-Ying Chuang

Comprehension and production of Kinyarwanda verbs in the Discriminative Lexicon

https://doi.org/10.1515/ling-2021-0164
Received September 15, 2021; accepted March 31, 2023; published online November 10, 2023

Abstract: The Discriminative Lexicon is a theory of the mental lexicon that brings together insights from various other theories: words are the relevant cognitive units in morphology, the meaning of a word is represented by its distribution in utterances, word forms and their meaning are learned by minimizing prediction errors, and fully connected networks successfully capture language learning. In this article we model comprehension and production of Kinyarwanda verb forms in the Discriminative Lexicon model. Kinyarwanda is a highly inflectional language, and therefore particularly interesting, because its paradigms are almost unlimited in size. Can knowledge of its enormous paradigms be modeled only on the basis of words? To answer this question we modeled a data set of 11,528 verb forms, hand-annotated for meaning and their grammatical functions, in the Linear Discriminative Learning (LDL), a two-layered, fully connected computational implementation of the Discriminative Lexicon model. We also extracted 573 verbs from our data set for which meanings are available that are based on empirical word embeddings obtained from large text corpora, and modeled them in LDL. Both comprehension and production is learned accurately: Kinyarwanda verb forms can be comprehended and produced relying on words as cognitive units, in a two-layered network, in which prediction errors are minimized.

Keywords: comprehension and production; computational modeling; Discriminative Lexicon; Kinyarwanda (Bantu); word and paradigm

*Corresponding author: Ruben van de Vijver, Institut für Linguistik, Heinrich-Heine-Universität, Universitätsstraße 1, 40225, Düsseldorf, Germany, E-mail: ruben.vijver@hhu.de. https://orcid.org/0000-0003-0761-4649
Emmanuel Uwambayinema, Institut für Linguistik, Heinrich-Heine-Universität, Düsseldorf, Germany, E-mail: emuwa100@hhu.de
Yu-Ying Chuang, Department of Taiwan Culture Languages and Literature, National Taiwan Normal University, Taipei City, Taiwan, E-mail: yuying.chuang@ntnu.edu.tw

Open Access. © 2023 the author(s), published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.
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 (Booij 2010, 2016; Bruening 2018; 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; 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, functions 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 (Blevins 2003, 2006, 2013, 2016; Hockett 1954; Katamba 1978)? In that case speakers could not work out the necessary rules, functions 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. Katamba (1978) observes that Bantu languages are often cited in the company of Turkish as typically agglutinating languages, in which each part of the word has a clearly identifiable, unique morphological function. Yet he provides a number of examples from several Bantu languages which show that this is an idealization.

As morpheme-based analyses overwhelmingly dominate the theoretical literature, and not only the literature on Bantu, we feel we need to back up our skepticism
in this matter with arguments (but see Ainsworth 2019; Engelmann et al. 2019: for analyses of highly inflectional languages using a word-based perspective). After having done so in the next paragraphs for Bantu languages, and also for Kinyarwanda, an agglutinative Bantu language and the empirical focus of this article, we will present an alternative approach to comprehension and production of complex words within the framework of the Discriminative Lexicon (DL, Baayen et al. 2018, 2019b; van de Vijver and Uwambayinema 2022), a word-based theory of the mental lexicon. DL proposes that the mental lexicon contains only word forms and their meanings which form a fully connected network; DL models comprehension as a mapping from word forms onto meaning, and production as a mapping from meaning onto word forms.

1.1 Difficulties of morphemic analyses in Bantu

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 and Inkelas 2017; van der Wal 2015). In his analysis of Kinyarwanda, Banerjee (2019) follows the literature on Bantu (Hyman 2003; Hyman and Inkelas 2017; van der 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 and 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 and Dube 2019). The locative is expressed by a prefix and suffix: e-gwahw-eni Loc-book-Loc “in the book”. Multiple exponence is also reported for verbs in Lusonga, a Bantu language spoken in Uganda (Hyman and Inkelas 2017). In Lusonga there are two causatives which are both expressed in verb forms (Hyman and Inkelas 2017). Multiple exponence is problematic for the morphemic view in which each morpheme is associated with one meaning (Chuang et al. 2020a).
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 and Mpiranya 2010; Nurse and 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” *ndasoma*, is *síinsoma* (Goldsmith and 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 and Mpiranya 2010). Verbs either carry a high tone or no tone (Goldsmith and 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 and 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.
This brings us to word-based theories. Word-based theories of morphology avoid the construct of the morpheme and its associated problems. The Discriminative Lexicon is a word-based theory, as we will discuss below, and is therefore similar in spirit to other word-based theories, especially with those that do not have an explicit layer in which morphology is represented, such as Word and Paradigm (Blevins 2003, 2006, 2013, 2016), and Emergentist Morphology (Rácz et al. 2015). The Discriminative Lexicon differs from word-based theories that assume morphological representations of complex words (Blevins 2016: 120). This is the case, for example, Paradigm Function Morphology (Stump 2001, 2016), and Construction Morphology (Booij 2010, 2016).

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. This is briefly illustrated using an example from Kinyarwanda. Like all Bantu languages, Kinyarwanda has an intricate class system for nouns, which is sometimes analyzed as a gender system (Güldemann and Fiedler 2021). Class markers classify nouns both semantically and grammatically. The noun class system was semantically productive in Proto-Bantu, but synchronically it no longer is. Each noun has a class marker for the singular and a different one for the plural (Demuth 2000; Nurse and Philippson 2006; van der Wal 2015). Kinyarwanda has 16 different class makers (van de Vijver and Uwambayinema 2022). 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 the morphological properties can be easily identified. For example, the Kinyarwanda verbal root RAS “to shoot” can be associated with the morphosyntactic property set {infinitive}: <RAS, {infinitive}>. A paradigm function will realize this configuration as kuras. 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 verbal root. This problem is compounded by the fact that not all lexemes of this paradigm share the root ras. In lexemes with a causative meaning the verbal root is realized as [-raf-], and the final vowel a is present in all present and future lexemes, in addition to the infinitive forms. In short, it is unclear how native speakers of Kinyarwanda will arrive at an analysis in which they have identified RAS as a unit.
The same problem arises in construction morphology. Constructions are form-meaning correlations.\textsuperscript{1} In a discussion of English agentive constructions, such as for instance \textit{baker}, Booij (2016) gives the following example.

\begin{equation}
\langle[X]_{vi}\rangle N_j \leftrightarrow \text{[Agent of SEM]}_i\
\end{equation}

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 \textit{kurasa}, can be represented as a construction, as is illustrated in (2).

\begin{equation}
\langle ku[X]a \rangle V_i \leftrightarrow \text{[Infinitive of SEM]}_i\
\end{equation}

In (2) the $X$ represents the phonological content of the base verb. In this case, too, it is unclear how a native speaker would arrive at this representation.

In short it is not clear how speakers (as opposed to a linguist) can isolate the morphemes, roots 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 composition, or how they can comprehend complex words by parsing them into smaller units to derive their meaning (Zwitserlood 2018).

After this dyspeptic discussion of problems, difficulties and analytical ambiguities, 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, and builds upon insights from distributional semantics, learning theory and machine learning. We will first introduce its central theoretical tenets, and in Section 2.1 we will introduce the mathematical concepts behind the Discriminative Lexicon, here we will introduce the central theoretical tenets.

\subsection*{1.2 The Discriminative Lexicon}

The Discriminative Lexicon (Baayen et al. 2018, 2019b; Chuang et al. 2020a, 2020b, 2021; Denistia and Baayen 2023; van de Vijver and Uwambayinema 2022) is a comprehensive theory of the mental lexicon 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

\footnote{Kapatsinski (2021) proposes constructions that are mappings from form to meaning and from meaning to form mappings, instead of a bidirectional mapping between form and meaning, which avoid the problems for Booij’s bidirectional instantiation of construction morphology.}
(Blevins 2003, 2006, 2013, 2016); with usage-based theories it shares that structure in language emerges through language use and learning (Bybee 1985; Kapatsinski 2018; Rácz et al. 2015); with distributional semantics it shares the hypothesis that words get their meaning in utterances (Firth 1957; Landauer and Dumais 1997; Sahlgren 2008; Weaver 1955); from error-driven learning it implements the hypothesis that learning is the result of minimizing prediction errors (Rescorla and Wagner 1972; Widrow and Hoff 1960). 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).

The phonology of a word is represented as a subset of the sounds of the language. For example the phonological representation of the word *ndaca* specifies the sounds that it contains and all sounds that are part of the phonology of Kinyarwanda that it does not contain. Furthermore, the sounds of a word are represented as n-grams. Cognitively, this representation seeks to reflect findings of research on which phonological representations are used to discriminate among words.

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 other words, they use allophonic information to discriminate among word variants used by different speakers. As allophonic information is contextual, the Discriminative Lexicon represents words in n-grams. In languages with complex syllable structures, the most informative n-grams may be bi-, tri-, or quatro-grams. In

---

2 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.

3 In phonological theory, error-driven learning theories explain the learning of constraint ranking (Boersma and Hayes 2001; Tesar and 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 2011; Adriaans and Kager 2010), without a role for minimizing errors.
languages that have a limited set of syllable structures, such as Kinyarwanda, it is possible that other sublexical units, for example syllables, are more useful as n-grams (Pham and Baayen 2015).

The semantics of a word is represented as its distribution in utterances in a corpus (Blevins 2016; Firth 1957; Landauer and Dumais 1997; Sahlgren 2008; Weaver 1955). These representations are known as word embeddings or semantic vectors. The vectors capture all information available in the distribution of a word: not only the meaning of its lemma, but also each of its grammatical functions. The vectors are high dimensional and represent meaning in a dynamic way, since each change to the corpus results in a (slight) change in the distribution of the word. This property of the Discriminative Lexicon differentiates it from almost all other theories of morphology, in which meaning is represented as a uniform label, e.g. plural (see Romain et al. 2022: for an interesting study using multiple uniform labels, which are pointers to locations in a high-dimensional semantic space).

The dynamic interpretation of meaning is supported by recent findings on the meaning of grammatical functions. Work by Shafaei-Bajestan et al. (2023) has provided evidence that English plural forms do not have a uniform meaning⁴ (see also Nikolaev et al. 2023: for Finnish). Plural forms of words referring to animals have a different distribution than plural forms of words referring to plants. Nieder et al. (2023) report that sound plurals in Maltese have a different distribution than broken plurals.

The phonological and semantic representations are combined into a fully connected network. This means that each phonological n-gram representation is connected to each cell in the semantic vectors of the language. In representing the mental lexicon as a network, the Discriminative Lexicon resembles the model proposed by Bybee (1985). The differences concern the granularity of the phonological representations, the dynamics of the semantic representations, and, a topic we will turn to now, the way in which the connections between the representations are established.

In the Discriminative Lexicon the weight of each connection in the network is established by minimizing prediction errors. There are two methods to achieve this. One is based on discriminative learning (Rescorla 1988; Rescorla and Wagner 1972), 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 and processing (Baayen 2011; Baayen et al. 2016b, 2018, 2019b; Chuang et al. 2020a, 2021; Denistia and Baayen 2023; Milin et al. 2017b; Nieder et al. 2023; Ramscar 2019; van de Vijver et al. 2023).

---

⁴ A uniform contribution of grammatical functions to semantics is implied, for example, by Romain et al. (2022), in which each label always points to a fixed point in the high-dimensional semantic space and thus contributes a fixed meaning to the whole word.
Ramscar and Gitcho 2007; Ramscar and Yarlett 2007; Ramscar et al. 2013a; van de Vijver and Uwambayinema 2022). In discriminative learning the weight of an association changes as a result of minimizing prediction errors from one representation to another representation (Rescorla and Wagner 1972), by means of the Widrow-Hoff learning rule (Shafaei-Bajestan et al. 2020). In order to incrementally minimize the difference between the predictions and the data, the Widrow-Hoff rule uses gradient descent.\(^5\) This method can be used to model language acquisition. Another method of minimizing prediction errors is by using multivariate, multiple regression. Since the goal of any regression analysis is to minimize the difference of the predictions and the regression line. This method analyzes all data at once, and can therefore be used to model the knowledge of an adult language user. Error-driven learning has been incorporated in Linear Discriminative Learning (LDL) (Baayen 2011; Baayen et al. 2018; Chuang et al. 2020a), a computational implementation of the Discriminative Lexicon.

This brings us to our present study.\(^6\) In order to evaluate the various hypotheses of the Discriminative Lexicon we will analyze comprehension and production of verbs from Kinyarwanda (Banerjee 2019; Nurse and Philippson 2006) in Linear Discriminative Learning. We will do so by modeling comprehension and production of verbs from two data sets and assessing how well LDL predicts their comprehension and production. One data set consists of 11,528 manually annotated verb forms, and the other data set consists of 573 verbs for which empirical word embeddings obtained from large text corpora are available.

Kinyarwanda is well suited for our purposes for a number of reasons. The language is highly inflectional and has a vast number of word forms for each lexeme. In this respect our study adds to the application of DL to another agglutinative language, Estonian (Chuang et al. 2020a). In comparison to the study of Estonian nouns, our study of Kinyarwanda verbs offers an even more daunting number of word forms for each lexeme. Kinyarwanda offers the possibility to investigate both inflection and derivation (see Section 3 for an overview of Kinyarwanda verbs and our dataset.) The extensions of Kinyarwanda are often analyzed as part of derivation (van der Wal 2015). In this respect our study goes further than the analysis of Latin and Estonian (Baayen et al. 2018; Chuang et al. 2020a). In comparison to English (Baayen et al. 2019b) Kinyarwanda has fewer computational resources, and our study will therefore advance our computational and linguistic knowledge of a language about which not nearly enough is known.

This article is organized as follows. In Section 2 we introduce one mathematics and the computational implementation of the Discriminative Lexicon, Linear

\(^5\) These representations are called cue and outcome in the technical language of this literature.

\(^6\) The data and scripts for this study are available at: https://osf.io/jdaqy/.
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. 2020a). 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 out modeling, and we conclude the article 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. 2016a, 2016b, 2019b; Chuang et al. 2020a, 2021; Milin et al. 2017a).

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 2.3 we will review the results of modeling comprehension and production in Latin (Baayen et al. 2018) and Estonian (Chuang et al. 2020a).

2.1 Introduction to LDL

LDL models comprehension as the mapping of a word form onto a meaning, and it models production as the mapping of a meaning onto a word form. This is achieved by representing word forms and meanings separately in spaces with a great number of dimensions. Each word form has a position in the word form space, and each meaning has a position in the meaning space. The positions of the word forms and their meaning are defined by their phonological and semantic representation. Each dimension in the word form space is connected to all other dimensions in the meaning space and vice versa, which means that the spaces are fully connected. The connections are weighted differently. The weight of a connection represents how well a sublexical n-gram word form cue predicts different shades of meaning (e.g., mood), and vice versa.

The word form space and the meaning space are represented as matrices. The matrix called $C$ describes the space for word forms and the matrix called $S$ describes the space for meanings. In each matrix each word form is represented as a vector, and its meaning is also represented as a vector. The fully connected network is illustrated in Figure 1.
The mapping from a point in one space onto a point in the other space is achieved mathematically by means of matrix multiplication. In Figure 2 matrix \( C \) represents the phonology of the word forms \( w_1, w_2 \) and \( w_3 \) by means of the phonological properties \( f_1 \) and \( f_2 \) represented as numbers; matrix \( S \) represents the meaning of the word forms \( w_1, w_2 \) and \( w_3 \) by means of the semantic properties \( s_1 \) and \( s_2 \) also represented as numbers. Points from \( C \) can be mapped onto points in \( S \) by means of multiplication by transformation matrices called \( F \) and \( G \) (In other words by solving \( CF = S \) and \( SG = C \)). \( F \) and \( G \) are calculated on the basis of \( C \) and \( S \). The mathematical details of deriving \( F \) and \( G \) are explained in Baayen et al. (2018, 2019b); Nieder et al. (2023). But deriving \( F \) and \( G \) is comparable to solving an equation with one unknown.

Once the transformation matrices have been derived, they can be used to map \( C \) onto \( S \) and vice versa by matrix multiplication. To illustrate this we will use the toy matrices provided in Figure 2. Multiplying the first row of \( C \) by the transformation matrix \( F \) results in the first row in \( S \):

\[
\begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \times 2 + 2 \times 0 \\ 1 \times 0 + 2 \times -2 \end{pmatrix} = \begin{pmatrix} 2 \\ -4 \end{pmatrix}
\]

Multiplying the last row of matrix \( S \) in Figure 2 with matrix \( G \) results in the last row of matrix \( C \):
The examples in Figure 2 need of course to be scaled up, since realistic data sets consist of thousands of words with many different phonological forms and meanings. The word forms of a data set with \( x \) words and \( y \) dimension in which the phonology of these words are represented, will be represented in a matrix \( C \) with \( x \times y \) dimensions, and if there are \( z \) dimensions of meaning of those words, these will be represented in a matrix \( S \) with \( x \times z \) dimensions. The matrix \( F \) with dimensions \( y \times z \) uses \( C \) to predict \( S \), which is how comprehension is modeled; the matrix \( G \) with dimensions \( z \times y \) uses \( S \) to predict \( C \), which is how production is modeled.

The mappings instantiate multivariate, multiple regression. The transformation matrix \( F \) can be understood as containing \( \beta \) coefficients that predict the values in \( S \), and the transformation matrix \( G \) can be understood as containing \( \beta \) coefficients that predict the values in \( C \). Since regression aims at minimizing the square error, this model is error-driven. The predictions of the regression model
can be interpreted as capturing the end-state of learning for the representations in C and S. It is by comparing the predictions of the mappings with the actual forms in C and S that the model’s predictions can be evaluated (see Section 4 for more details).

### 2.2 Word form and meaning matrices

The phonology of the word forms is represented in a matrix. The rows of the matrix consist of word 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).

\[
C = \begin{pmatrix}
    #nd & nda & dar & ara & ras & asa & sa# & #ur & ura & rar \\
    1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
    0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
    0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
\end{pmatrix}
\]

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. Triphones of letters were chosen, since they contain contextual information, which has been shown to play an important role in lexical processing (Mitterer et al. 2018). 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 and Tesar 2004).

We can now turn to the representation of meaning. This is a matrix as well, 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 and 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 11,528 verbs in our dataset are included in the word embeddings, we decided to use both full data set of 11,528 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. For each word the value of the lexeme and the value of each grammatical function is represented by a vector. 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. 2020a). The meaning of, for example, ndarasa is represented analytically as a vector consisting of the values of its lexeme and of its grammatical functions, as presented in (4).

\[
\begin{align*}
\text{ndarasa} & \rightarrow = \text{Kurasa} + \text{one} + \text{singular} + \text{present} + \text{active} + \text{indicative} + \text{neutral} \\
\text{urarasa} & \rightarrow = \text{Kurasa} + \text{two} + \text{singular} + \text{present} + \text{active} + \text{indicative} + \text{neutral}
\end{align*}
\]

For words of the same lexeme (e.g., ndarasa and urarasa), given that they only differ in a few grammatical or semantic functions (the value of person for these two words), their simulated semantic vectors will therefore be similar as well.

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).

\[
S = \begin{pmatrix}
\text{ndarasa} & 13.99 & 0.63 & 12.13 & -4.22 & 16.17 & 15.13 & 1.26 & 26.50 & -4.07 & 4.18 & 0.77 & 4.53 \\
\text{urarasa} & 27.42 & -3.04 & 1.41 & 8.85 & 17.04 & 14.14 & -0.53 & 22.03 & -6.60 & 13.43 & 10.92 & 3.83 \\
\text{bararasa} & -1.97 & -8.90 & -0.19 & -0.75 & 11.16 & 7.59 & 4.21 & 12.19 & 2.32 & 9.06 & -4.50 & 10.60
\end{pmatrix}
\]

The fact that the S matrix in (5) has column headers S1, S2, S3, ..., S12 is due to the fact that we want to represent word meanings in a high dimensional vector space. We could have used the same encoding as we did for the word form matrix C by using the lexemes and grammatical functions to name each column in S, and use one-hot encoding to specify which lexemes or functions are present or absent in a word’s meaning. This, however, comes with the risk that once the data set becomes bigger, the number of columns in S will increase rapidly, which causes difficulties in computation. To address this problem, we therefore represent the meaning of each lexeme and grammatical function by a vector of real numbers drawn from a normal
distribution. To optimize mapping accuracy, we usually set the number of dimensions (columns) in $S$ the same as the number of dimensions in $C$. The meaning of a word is again the sum of the real-valued vectors of its pertinent lexeme and grammatical functions. 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.\(^8\)

The semantic relation between words in our simulations is determined by the similarity between vectors, commonly assessed by correlation. For example, the correlation between \textit{ndarasa} and \textit{urarasa} is 0.70, which is larger than that between \textit{ndarasa} and \textit{bararasa}, 0.53. This is because \textit{urarasa} is closer to \textit{ndarasa} in meaning than \textit{bararasa} is: while \textit{ndarasa} and \textit{urarasa} only differ in Person (‘I shoot’ vs. ‘you (sg.) shoot’), \textit{bararasa} differs from \textit{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.

Let us now describe how the accuracy of the model 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 $C$ 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 multiple candidates are formed. 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 support ngrams receive in each position. To be more specific, again take the word \textit{ndarasa} for example. The first cue (letter trigram) of this

\(^8\) This is not always the case. See Baayen et al. (2019b) for methods of constructing semantic vectors with interpretable semantic dimensions based on discriminative learning.
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 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 provide 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 support in order to be considered by the path-finding algorithm. The amount of positional 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 and Davies (2019) show that speakers do make few mistakes, this is not an implausible assumption.

LDL is computationally implemented as the package JudiLing (Luo 2021; Luo et al. 2021), for the programming language julia.9 The implementation of LDL in julia offers a better algorithm for production and testing of held-out data, has an extra function 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.

---

9 JudiLing’s manual is available at: https://megamindhenry.github.io/JudiLing.jl/stable/. Upon publication our data sets and scripts will be made available through OSF.
2.3 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. (2020a) modeled the comprehension and production of 232 nouns, each inflected for 14 cases and two numbers, yielding a dataset of 6,496 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. (2020a) 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 2003, 2006, 2013, 2016). 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. (2020a) 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 Kinyarwana 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, comitative, revesive), 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; van der 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

Table 1: A few of the verbs forms of 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.

<table>
<thead>
<tr>
<th>Verb form</th>
<th>Person</th>
<th>Tense</th>
<th>Lemma</th>
<th>Number</th>
<th>Aspect</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ndaca</td>
<td>1</td>
<td>present</td>
<td>guca</td>
<td>sing</td>
<td>imperfective</td>
<td>to cut</td>
</tr>
<tr>
<td>uraca</td>
<td>2</td>
<td>present</td>
<td>guca</td>
<td>sing</td>
<td>imperfective</td>
<td>to cut</td>
</tr>
<tr>
<td>araca</td>
<td>3</td>
<td>present</td>
<td>guca</td>
<td>sing</td>
<td>imperfective</td>
<td>to cut</td>
</tr>
<tr>
<td>nca</td>
<td>1</td>
<td>present</td>
<td>guca</td>
<td>sing</td>
<td>perfective</td>
<td>to cut</td>
</tr>
<tr>
<td>uca</td>
<td>2</td>
<td>present</td>
<td>guca</td>
<td>sing</td>
<td>perfective</td>
<td>to cut</td>
</tr>
<tr>
<td>aca</td>
<td>3</td>
<td>present</td>
<td>guca</td>
<td>sing</td>
<td>perfective</td>
<td>to cut</td>
</tr>
<tr>
<td>nacaga</td>
<td>1</td>
<td>past</td>
<td>guca</td>
<td>sing</td>
<td>imperfective</td>
<td>to cut</td>
</tr>
<tr>
<td>wacaga</td>
<td>2</td>
<td>past</td>
<td>guca</td>
<td>sing</td>
<td>imperfective</td>
<td>to cut</td>
</tr>
<tr>
<td>yacaga</td>
<td>3</td>
<td>past</td>
<td>guca</td>
<td>sing</td>
<td>imperfective</td>
<td>to cut</td>
</tr>
<tr>
<td>naciye</td>
<td>1</td>
<td>past</td>
<td>guca</td>
<td>sing</td>
<td>perfective</td>
<td>to cut</td>
</tr>
<tr>
<td>waciye</td>
<td>2</td>
<td>past</td>
<td>guca</td>
<td>sing</td>
<td>perfective</td>
<td>to cut</td>
</tr>
<tr>
<td>yaciye</td>
<td>3</td>
<td>past</td>
<td>guca</td>
<td>sing</td>
<td>perfective</td>
<td>to cut</td>
</tr>
<tr>
<td>ndasoma</td>
<td>1</td>
<td>present</td>
<td>gusoma</td>
<td>sing</td>
<td>imperfective</td>
<td>to read</td>
</tr>
<tr>
<td>urasoma</td>
<td>2</td>
<td>present</td>
<td>gusoma</td>
<td>sing</td>
<td>imperfective</td>
<td>to read</td>
</tr>
<tr>
<td>arasoma</td>
<td>3</td>
<td>present</td>
<td>gusoma</td>
<td>sing</td>
<td>imperfective</td>
<td>to read</td>
</tr>
<tr>
<td>nsoma</td>
<td>1</td>
<td>present</td>
<td>gusoma</td>
<td>sing</td>
<td>perfective</td>
<td>to read</td>
</tr>
<tr>
<td>usoma</td>
<td>2</td>
<td>present</td>
<td>gusoma</td>
<td>sing</td>
<td>perfective</td>
<td>to read</td>
</tr>
<tr>
<td>asoma</td>
<td>3</td>
<td>present</td>
<td>gusoma</td>
<td>sing</td>
<td>perfective</td>
<td>to read</td>
</tr>
<tr>
<td>nasomaga</td>
<td>1</td>
<td>past</td>
<td>gusoma</td>
<td>sing</td>
<td>imperfective</td>
<td>to read</td>
</tr>
</tbody>
</table>
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,

---

10 The feature [nasal] could be analyzed as a morpheme with accompanying linearization contraints
(Tebay and Zimmermann 2020; Zimmermann 2017), yet this analysis would make it necessary to also
assume that the obstruent part of the prenasalized stop would be a morpheme with the meaning
present, but only for the first person. This analysis then would run afoul of the idea that a morpheme
has a unique meaning.
they are realized as vowels, but when they are followed by a vowel, they are realized as glides, [j] and [w] respectively (Zorc and 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, ɱ, f, v, s, z, ş, ʐ, h, ɲɖ, ɲʂ, ɲʑ, ɲɟ, ɲ, c, ŋ, k.

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 in Kinyarwanda. Vowels preceding prenasalized consonants are intermediate in length between short and long vowels. Myers suggests that 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 pre-nasalized 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.

### Table 2: Consonants of Kinyarwanda.

<table>
<thead>
<tr>
<th></th>
<th>labial</th>
<th>labio-dental</th>
<th>alveolar</th>
<th>postalveolar</th>
<th>palatal</th>
<th>velar</th>
<th>glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>plosive</td>
<td>p</td>
<td>t, d</td>
<td>c, j</td>
<td></td>
<td>k, g</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td>j</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fricative</td>
<td>f, v</td>
<td>s, z</td>
<td>ş, ʐ</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>affricate</td>
<td>pf</td>
<td>ts</td>
<td>ʂ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flap</td>
<td></td>
<td></td>
<td>ɭ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>approximant</td>
<td>w</td>
<td></td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 possible Extensions: applicative, causative, comitative, frequentative, iterative and reversive. This resulted in 11,528 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 and Mpiranya 2010), which we leave for future work.

We used our data set of 11,528 verbs to create a second one. We did this because the grammatical functions and lexical meaning in our data set have been defined by us, and even though we aimed for our annotations to simulate the distribution of the verb forms in Kinyarwanda sentences, a more unbiased distribution would be derived from word embeddings of actual sentences. Embeddings are a good way to instantiate the hypothesis of the Discriminative Lexicon that meanings of words are based on the distribution of words in sentences (Baayen et al. 2019b; Landauer and 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 trained with Word2Vec (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 they included embeddings for inflectional functions. Niyongabo et al. (2020) we used only words and no further information, as such information is not 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. 2020a), 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 and 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. If we only focus on 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 11,528 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. 2020a). 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 by means 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 candidate word forms and the target word form is established and the candidate 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%.

The errors in individual forms for each of the models are illustrated from Tables 4-6. 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,

| 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. |
|---------------------------------|----------|----------|---------|---------|---------|---------|----------|----------|
| Person | Tense | Lexeme | Number | Aspect | Voice | Extension | Mood |
| 99.8% | 99.0% | 99.5% | 99.9% | 99.5% | 99.9% | 97.1% | 99.1% |
| 99.8% | 99.2% | 99.9% | 99.9% | 99.9% | 99.9% | 98.7% | 99.1% |
| 99.8% | 99.3% | 99.9% | 99.9% | 99.9% | 99.9% | 99.7% | 99.4% |
Table 4: For the model with letter 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). There were 11,104 correct and 424 false predictions (96.3 % correct).

<table>
<thead>
<tr>
<th>Target Form</th>
<th>Predicted</th>
<th>r</th>
<th>r-target</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>yazimururuzwaga</td>
<td>yazimuruzwaga</td>
<td>0.95727</td>
<td>0.95725</td>
<td>Extension</td>
</tr>
<tr>
<td>bahore</td>
<td>bahore</td>
<td>0.9641</td>
<td>0.9640</td>
<td>Extension</td>
</tr>
<tr>
<td>barazimuruzwa</td>
<td>barazimuruzwa</td>
<td>0.9587</td>
<td>0.9586</td>
<td>Extension</td>
</tr>
<tr>
<td>wayoboranwe</td>
<td>uyoboranwe</td>
<td>0.9308</td>
<td>0.9305</td>
<td>Voice, Mood</td>
</tr>
<tr>
<td>yazimuruzwaga</td>
<td>yazimuruzwaga</td>
<td>0.9570</td>
<td>0.9566</td>
<td>Extension</td>
</tr>
</tbody>
</table>

Table 5: For the model with syllable bigrams 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). There were 11,357 correct and 171 false predictions (98.6 % correct).

<table>
<thead>
<tr>
<th>Target Form</th>
<th>Predicted</th>
<th>r</th>
<th>r-target</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>no.gu.ru.ye</td>
<td>no.gu.ye</td>
<td>0.9561332</td>
<td>0.9561068</td>
<td>Extension</td>
</tr>
<tr>
<td>two.gu.ra</td>
<td>two.gu.ru.ra</td>
<td>0.9566765</td>
<td>0.9565960</td>
<td>Extension</td>
</tr>
<tr>
<td>two.gu.ra.ga</td>
<td>two.gu.ru.ra.ga</td>
<td>0.9575569</td>
<td>0.9570913</td>
<td>Tense, Aspect</td>
</tr>
<tr>
<td>bo.gu.ru.ye</td>
<td>bo.gu.ye</td>
<td>0.9533335</td>
<td>0.9528527</td>
<td>Extension</td>
</tr>
<tr>
<td>tu.ro.gu.ra</td>
<td>tu.ro.gu.ru.ra</td>
<td>0.9583053</td>
<td>0.9576806</td>
<td>Extension</td>
</tr>
</tbody>
</table>

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 11,500 correct and 28 false predictions (accuracy 99.8 %).

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

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 he fifth line again involves errors between the reversive and the iterative form of
the third person singular, passive, imperfective, indicative of the lexeme *kuzimira* “to be lost”.

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*.

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.

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.

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 (Table 7).

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).
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 and 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. 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 comprehension 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).

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Letter trigrams</th>
<th>σ-bigrams</th>
<th>σ-trigrams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension</td>
<td>82 %</td>
<td>92.7 %</td>
<td>98.6 %</td>
</tr>
<tr>
<td>Production (build)</td>
<td>75.6 %</td>
<td>86.2 %</td>
<td>91.8 %</td>
</tr>
<tr>
<td>Production (learn)</td>
<td>89.7 %</td>
<td>97.4 %</td>
<td>99.5 %</td>
</tr>
</tbody>
</table>

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

Table 10: Overview of the accuracy of the models for comprehension and production of verbs with meaning derived from word-embeddings. It should be kept in mind that the data set based on word embeddings is only 20% of the size of the entire data set.
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 11,528 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 Kinyarwanda words can be learned by using meaning representations based on distributional semantics. In a model of lexical processing of English words, which used semantic vectors that were created by LDL, the semantic vectors were given extra inflectional lexomes (Baayen et al. 2019b). This was not done for Kinyarwanda (Niyongabo et al. 2020), since we did not create the embeddings ourselves. 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 (van der Wal 2015).

For completeness sake, 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, but we do not accord much weight to them. Again comprehension is better than production.

### 4.4 Discussion

We modeled two data sets of Kinyarwanda verbs. One consists of 11,528 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 Conclusions

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 2003, 2006, 2013, 2016). The procedures to calculate the accuracy for comprehension and production in LDL rely on correlation between vectors of whole words, and that, as a consequence, LDL is word-based. As there are no morphemes, stems or exponents in the input to LDL, its results can be interpreted in terms of word and paradigm morphology (Blevins 2003, 2006, 2013, 2016). The accuracy of comprehension and production is very good. 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 (Blevins 2003, 2006, 2013, 2016; Matthews 1972) and distributional semantics, that words get their meaning in utterances (Firth 1957; Landauer and 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 and Dumais 1997). The distributional hypothesis is backed up by the results in both data sets: in 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. 2016b, 2018, 2019b; Chuang et al. 2020a, 2020b, 2021; Tomaschek et al. 2021) or deep neural networks (Futrell et al. 2020; Hahn 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 model 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 and Kager 1999; Hayes 1995; Martínez-Paricio and Kager 2015), which have also been reported for Kinyarwanda (Goldsmith and 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 and 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 learning a language (Ambridge 2020; Bybee 2001; Ellis and Ogden 2017; Goldberg 2019; Kapatsinski 2018; Tomasello 2003) by inducing a grammar from observable data (Abney 2021; Goldberg 2019; Mayer 2020). 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 and Ogden 2017). Morphology emerges as a consequence of usage and learning, as is proposed in Emergentist Morphology (Rácz et al. 2015). 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 and Garrett 2004).

The fact that the order in templates in Bantu languages varies slightly from language to language (Banerjee 2019), and that it is words, not morphemes, that drive analogical change (Hill 2020), 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, of which the meaning is represented as their distribution (either a simulated distribution, or their distribution in utterances) (Goldberg 2019). 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 and 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 and Mpiranya 2010). Including those affixes would make an even more ecologically valid data set, but would require modeling (small) phrases which is something we will pursue in the future.

One topic that requires further work, but is beyond the scope of our article, 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 and 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 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 2021) with only modest success, whereas models with only two 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 and Smolka 2020; Baayen et al. 2019b; Chuang et al. 2020b; Tomaschek et al. 2021). 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 (Arnold et al. 2017; Shafaei-Bajestan et al. 2020), suggests that phonological knowledge can be modeled with a dense two-layer network. Our work provides evidence that this is true for morphology as well.

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 theory of the Discriminative Lexicon (Baayen 2011; Baayen et al. 2016b, 2018, 2019b; Chuang et al. 2020a, 2020b, 2021). 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.

References


Brysbaert, Marc, Michaël Stevens, Paweł Mandera & 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.


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


