Comprehension, Production and Processing of Maltese Plurals in the Discriminative Lexicon

Jessica Nieder, Yu-Ying Chuang, Ruben van de Vijver
and Harald Baayen

September 23, 2021

Abstract

This study challenges a computational implementation of Word and Paradigm Morphology with the task of modeling the semi-productive noun system of Maltese, which combines a dozen concatenative plural patterns with eleven non-concatenative plural patterns. We show that our model, trained on 6,511 word forms, generates accurate predictions about what meanings listeners understand and what forms speakers produce. Furthermore, measures derived from the model are predictive for Maltese reaction times. Although mathematically very simple, the linear mappings between form and meaning posited by our model are powerful enough to capture the complexity and productivity of the Maltese noun system.

Keywords Discriminative Lexicon; Maltese Plurals; Word and Paradigm Morphology; Linear Discriminative Learning; Computational Modeling; Productivity; Primed Lexical Decision

1 Introduction

In this study, we challenge a computational implementation of Word and Paradigm Morphology (Blevins, 2016; Matthews & Matthews, 1972), the 'discriminative lexicon' (DL) (Baayen et al., 2019), with the task of modeling the noun system of Maltese, a Semitic language spoken in Europe. The DL model differs from most theories of morphology in that it defines mappings between form and meaning (comprehension) and meaning and form (production) without requiring theoretical constructs such as stems, exponents, and inflectional classes. In general, the task of morphological theory is often conceptualized as providing a formal mechanism specifying what sound sequences are possible words. The DL model divides this task into two subtasks: first, to predict what possible forms are, given their meanings; and second, to predict what possible meanings are, given their forms.

Most formal and computational accounts of word structure unfold almost exclusively in the world of forms. Forms are mapped onto forms. For instance, the prosodic theory of non-concatenative morphology laid out in McCarthy (1981) starts with underlying forms that are the starting point for a set of rules that derive words' surface forms. An account of Hebrew non-concatenative morphology within the framework of Optimality Theory (Prince & Smolensky, 2004) is given by Ussishkin (2005). Instead of deriving words from consonantal roots, as argued by McCarthy (1981), Ussishkin proposes that words are derived from other words, subject to a set of prosodic and morphological constraints.

Many computational models for morphology likewise do not predict words' forms from their meanings, but from other forms of these words. Some of these models set up a list of possible changes that have to be applied to move from one form to another, and then seek to predict which of the possible form changes is appropriate given selected properties of the base word. For instance, Ernestus and Baayen (2003) examined several quantitative models that all were given the task to predict whether or not the stem-final obstruent of a Dutch plural noun or verb form is voiced or voiceless. These models, which ranged from recursive partitioning trees and logistic regression models to Analogical Modeling (Skousen, 1989), Memory-Based learning (Daelemans & Van den Bosch, 2005) and Optimality Theory (Boersma & Hayes, 2001), all performed with roughly the same accuracy, suggesting that any reasonably decent statistical classifier, given access to the relevant features of the base word, can accomplish this classification task.[1] However, all these mod-

[1] Thus, the recursive partitioning algorithm of Belth et al. (2021) is also likely to perform well.

els are incomplete, in the sense that to create an actual plural form, the appropriate voicing feature has to be combined with further concatenation of the appropriate plural suffix.

For Semitic languages such as Arabic and Maltese, predicting the plural of a noun is set up as a classification problem by Dawdy-Hesterberg and Pierrehumbert (2014), focusing on Arabic, and by Nieder, Tomaschek, et al. (2021), focusing on Maltese. The former study used the Generalized Context Model (Nosofsky, 1986), the latter study applied Memory-Based learning (Daelemans et al., 2001), Naive Discriminative Learning (Baayen, 2011), as well as an Encoder-Decoder deep learning architecture (McCoy et al., 2020) to generate plurals from singulars. The deep learning model stands in the tradition of the past-tense model of Rumelhart and McClelland (1986), who derived past-tense forms from their present-tense counterparts.

The only way in which semantics plays a role in these models is through an inflectional contrast, such as singular versus plural, which is used to set up separate classes of forms. However, it seems unlikely that native speakers produce plurals from singulars. For second language learners, in pedagogical contexts, instructions for how to create the forms of a paradigm from its principal parts can be quite helpful. But whether native speakers derive forms via other forms is still an open question (Blevins, 2016; Nieder, Tomaschek, et al., 2021).

In psychology, several computational models have been put forward that construct complex words starting from their meanings. The models by Levelt et al. (1999) and Dell (1986) are similar in design to realizational theories of morphology (see, e.g., Bonami & Stump, 2016; Stump, 2001). To our knowledge, these two psychological computational models have not been applied to languages other than English, and it is therefore unclear whether the mechanisms of spreading activation and interactive activation, that they make use of, can be made to work for complex morphological systems such as the Maltese noun system.

Gaskell and Marslen-Wilson (1997) proposed a three-layer network model that maps speech input straight onto semantic representations. The triangle model of Harm and Seidenberg (2004) likewise addresses the relation between words' forms and their meanings, using a more complex multi-layer network. This model has been tested not only on English, but also on Serbo-Croatian (Mirković et al., 2005). Following their lead, the 'Discriminative Lexicon' model (Baayen et al., 2019) zooms in on the mappings from form to meaning in visual and auditory comprehension, and the mapping from meaning to form in production. As in the above connectionist models, both words' forms and their meanings are represented by high-dimensional numeric vectors. However, the DL model simplifies the connectionist multi-layer networks of

Gaskell and Marslen-Wilson (1997) and Harm and Seidenberg (2004) by removing all hidden layers. The simple input-to-output network that results is mathematically equivalent to multivariate multiple linear regression.

By representing words' meanings numerically, it becomes possible to harness the power of distributional semantics (Landauer & Dumais, 1997; Mikolov et al., 2013; Mitchell & Lapata, 2008) when considering the questions of what possible meanings are given words' forms, and what possible forms are given words' meanings. This is important, because form and meaning can show intricate interactions. For instance, Baayen and Moscoso del Prado Martín (2005) called attention to irregular verbs in English (and as well in German and Dutch) being more similar to each other in their meanings than regular verbs. The greater semantic density of irregular verbs in English may underlie the interaction of semantic deficits and regularity in aphasia reported by Bird et al. (2003), and modeled computationally using distributional semantics by Heitmeier and Baayen (2021). Below, we shall see that the broken plurals and the sound plurals of Maltese may also pattern differently in semantic space.

Several studies suggest that the DL correctly predicts the forms of complex words (see Baayen et al. (2018) for Latin verb inflection, Chuang et al. (2020) for Estonian noun inflection, van de Vijver et al. (2021) for Kinyarwanda verbs, and Chuang, Kang, et al. (2021) for Korean verbs). The first goal of the present study is to clarify whether the theory of the DL also correctly predicts Maltese singular and plural nouns. Of particular interest is how well the simple networks used by the DL are able to model not only concatenative morphology, but also non-concatenative morphology.

The framework of the DL has also been used to predict how words are realized phonetically. Tomaschek et al. (2021) modeled the duration of English word-final [s] for different grammatical functions, Chuang, Vollmer, et al. (2021) predicted word duration for English pseudowords as pronounced by native speakers of English, and Chuang, Kang, et al. (2021) applied the model to word duration in Taiwan Mandarin. The latter study also shows that the priming effects reported for Dutch in Creemers et al. (2020) are correctly predicted by the model (see also Baayen & Smolka, 2020, for German). In the light of these results, the second goal of the present study is to clarify whether measures derived from the model help predict lexical processing costs, as gauged with a cross-modal primed lexical decision task.

The remainder of this paper is structured as follows. We first provide an overview of plural formation in Maltese and report previous experimental and computational studies on Maltese plurals. Section 3 proceeds with an introduction to the 'Discriminative Lexicon'. We then present the computational models that we developed for the Maltese noun system. We report how well they perform as a memory for known words, and also examine the extent to which the memory is productive, in the sense that it can handle unseen words that it has not been trained on. Subsequently, we show how the theory can be used to obtain further insight into the lexical processing of Maltese nouns in comprehension. We conclude this study with a discussion of the new insights that our results bring to morphological theory on the one hand, and its limitations on the other hand.

2 Maltese plurals

The turbulent history of Malta is reflected in the national language of the island. Maltese developed from Maghrebi Arabic, and has absorbed influences from Sicilian, Italian and, more recently, from English. These influences affected its lexicon and its morphology (Hoberman, 2007).

The Maltese noun plural system shows a perplexing amount of possible plural forms. Maltese has a great number of typically Semitic non-concatenative plural forms—called broken plurals in the Semitic linguistic tradition. Broken plurals are characterized by differences in the prosodic structure of a plural as compared to its corresponding singular form. For example, the singular form *kelb* 'dog' [kɛlp] has the plural form *klieb* 'dogs' [klɪːp] in which the coda consonant [l] of the singular is found in the onset of the plural form. In addition, the vowel [ɛ] in the singular form corresponds to [ɪː] in the plural. Schembri (2012) distinguishes 11 different broken plural patterns. In Maltese broken plurals account only for a small proportion of plural forms of the language (Borg & Azzopardi-Alexander, 1997, report a proportion of 10%). In addition to broken plurals, Maltese also has a sizable set of sound plurals. The majority of Maltese plurals belong to this category (Borg & Azzopardi-Alexander, 1997; Nieder, van de Vijver, et al., 2021a).

Sound plurals are characterized by additional segmental material at the right side of the plural in comparison to the singular: The singular form *prezz* 'price' has the plural form *prezzijiet* in which the plural differs from the singular in the suffix -*ijiet*. In their work, Nieder, van de Vijver, et al. (2021a, 2021b) distinguish 12 different sound plural patterns (they count the dual forms as a sound plural pattern) with different frequency distributions and productivity. Table 1 below gives an overview of the Maltese sound and broken plural patterns and the two possible dual forms.

The complexity of the Maltese noun system is not a consequence of different case-based declension patterns; unlike Akkadian and Modern Standard Arabic, Maltese nominals are not marked for grammatical case. Rather, its complexity is exclusively driven by the sheer variety of suffixes and patterns

Singular	**Plural**	**Gloss**	**Plural Type**
fardal	*fradal*	'aprons'	broken A, CCVVCVC
birra	*birer*	'beers'	broken B, (C)CVCVC
kbir	*kbar*	'big (pl.)'	broken C, CCVVC
ftira	*ftajjar*	'type of bread (pl.)'	broken D, CCVjjVC
bitħa	*btieħi*	'yards'	broken E, CCVVCV
sider	*isdra*	'chests'	broken F, VCCCV
marid	*morda*	'sick persons'	broken G, CVCCV
għodda	*għodod*	'tools'	broken H, (għ)VCVC
elf	*eluf*	'thousands'	broken I, VCVC
għaref	*għorrief*	'wise men'	broken J, CVCCVVC(V)
għama	*għomja*	'blind persons'	broken K, (għ)VCCV
karta	*karti*	'paper'	sound, -*i*
omm	*ommijiet*	'mother'	sound, -*ijiet*
rixa	*rixiet*	'feather'	sound, -*iet*
giddieb	*giddieba*	'liar'	sound, -*a*
meħlus	*meħlusin*	'freed'	sound, -*in*
kuxin	*kuxins*	'cushion'	sound, -*s*
triq	*triqat*	'street'	sound, -*at*
sid	*sidien*	'owner'	sound, -*ien*
baħri	*baħrin*	'sailor'	sound, -*n*
ħati	*ħatjin*	'guilty'	sound, -*jin*
qiegħ	*qigħan*	'bottom'	sound, -*an*
spalla	*spallejn*	'shoulder'	dual, -*ejn/ajn*
sieq	*saqajn*	'foot'	dual, -*ejn/ajn*

Table 1: Maltese broken plurals, sound plurals and duals (examples taken from Nieder, van de Vijver, et al., 2021a; Schembri, 2012). The digraph għ is historically a pharyngeal fricative, which was lost in modern Maltese (Borg & Azzopardi-Alexander, 1997).

available for pluralization.

A further complication is the existence of several plural forms for a singular, without a noticeable semantic difference among the plural variants. For example,the singular *kaxxa* (sg.) 'box' has two plural forms, one is a broken plural, *kaxex*, and one is sound plural, *kaxxi*; another example is the singular *giddieb* (sg.) 'liar', which has two sound plural forms, *giddieba* and *giddibin*.

In addition to sound and broken plurals, Maltese shows other plural types for a small number of nouns, such as the suppletive plural, e.g. *mara* - *nisa* 'women' or a double plural marking that is a blend of a broken plural and a sound plural suffix (called *plural of the plural* by Mayer et al. (2013)), the singular *tarf* has the blended plural *trufijiet* 'edge'. A few words are pluralized with a dual suffix but grammatically behave like plural words, for example *sieq* - *saqajn* 'foot' (Borg & Azzopardi-Alexander, 1997; Mayer et al., 2013).

2.1 Experimental and computational research on Maltese plurals

There exists both experimental and computational research on the Maltese nominal system. In the following, we first discuss the experimental research on Maltese nouns before turning to the computational studies.

Two experimental studies have clarified that native speakers use information about pattern frequency to produce plural forms for singulars they never heard before and to access plurals in their mental lexicons (Nieder, van de Vijver, et al., 2021a, 2021b). While some plural suffixes and patterns occur frequently in the language, for example the sound plural forms ending in *-i* and *-ijiet* or the broken plural patterns characterized by the CV-templates *CCVVCVC* and *CCVVC*, others are found in a relatively small amount of plural forms only (see Nieder, van de Vijver, et al., 2021a, 2021b; Schembri, 2012, for detailed information about pattern frequency in Maltese).

In a production study, Nieder, van de Vijver, et al. (2021a) asked Maltese native speakers to produce plurals for existing singulars and pseudo-singulars. The productions of the participants reflected the frequency of the plural patterns in Maltese. The participants made use of more frequent plural suffixes when they produced sound plurals (a finding that is also reported by Drake (2018)), and of more frequent CV templates when they produced broken plurals.

Further evidence for the importance of the type frequency of exponents (sound plurals) and CV templates (broken plurals) emerged from a reaction time study by Nieder, van de Vijver, et al. (2021b). Frequent broken templates and frequent sound plural exponents elicited significantly shorter reaction times than infrequent ones. This experiment did not provide evidence for an effect of plural type (broken versus sound): on average, response times for both kinds of plurals were highly similar. Below, we return to this study, to show that nevertheless the way in which responses are generated in this task differs for broken plurals and sound plurals.

Computational analyses of the Maltese plural formation have focused on form-to-form modeling. The singular form is taken as starting point for predicting its plural form, without making reference to the semantics of the two word forms. Some models are classifiers for plural classes, others generate full plural forms given the corresponding singulars.

Mayer et al. (2013) present a computational study of Maltese broken plurals that focuses on the application of rules to form plurals from singulars. In light of the fact that up until around 2010 the consensus among Maltese scholars was that there are no rules governing broken plurals (as discussed in Schembri, 2012), their approach breaks with this desperate tradition. Mayer et al. (2013) propose a set of four rules, based on the work of Schembri (2012), which would derive broken plurals from their singulars. These rules, implemented in Python, were shown to be very successful, correctly deriving 75% of all forms in their database of 654 word forms that have a broken plural. This study shows unambiguously that the Maltese broken plurals are to a considerable extent systematic. However, this study does not address the question of how speakers select between broken and sound plurals. Furthermore, as mentioned above, it is not self-evident from a cognitive perspective that speakers would create plurals from singulars — neither the production model by Dell (1986) nor the model of Levelt et al. (1999) make such a claim. And in child language it is also far from evident that there are morphophonological processes that link one form in the paradigm with another one. Zamuner et al. (2011), for example, found that Dutch children have difficulties applying the completely regular final devoicing rule of Dutch in forming singulars from plurals.

Farrugia and Rosner (2008) also focused exclusively on broken plurals, using an artificial neural network with encoder and decoder hidden layers, to categorize and produce Maltese broken plurals. As basis for their work they also use the analysis of Schembri (2012). Farrugia and Rosner (2008) included three different encoding methods: a general grapheme-to-phoneme conversion process, a one-to-one mapping of graphemes to phonemes (called *Features Lite* in their study), and a purely grapheme-based encoding method. The best model that was able to categorize nearly all nouns in their dataset with an accuracy of around 98% was the *Features Lite* model. Although they report good results for forms the model had seen in training, it did not perform well on unseen forms, achieving exact matches between predicted and observed plural forms for only 26.6% of the cases. This computational model again shows that there are indeed systematic relations between the form of the singular and its broken plural form. The model shows that these relations can be derived from the data without requiring handcrafted rules. It remains unclear, however, how the model would have performed if it had been trained on both broken plurals and sound plurals jointly.

Nieder, Tomaschek, et al. (2021) compared three different computational models to investigate whether it is in principle possible to account for the form-based relations in Maltese nominal paradigms without making recourse to the construct of the morpheme: the Tilburg Memory-Based Learner (TiMBL) (Daelemans et al., 2004), the Naive Discriminative Learner (NDL) (Baayen, 2011), and an Encoder-Decoder network. TiMBL and NDL are classifiers, the Encoder-Decoder network is a model generating actual plural forms. Models were trained on a dataset consisting of both sound plurals and broken plurals. The classifiers were given the task to predict which class out of 8 plural classes (4 broken plural classes, and 4 sound plural classes: three for the three most frequent exponents, and one for all other exponents) is appropriate for a given singular. TiMBL's best performance under 10-fold cross-validation was 97%, whereas NDL's best performance under 10-fold cross-validation was 88.7%. The Encoder-Decoder model was at 48.22%. Interestingly, although information about the CV template has been reported to increase classification accuracy for Arabic (Dawdy-Hesterberg & Pierrehumbert, 2014), such information did not improve the accuracy of the TiMBL classifier for Maltese.

What these modeling studies clarify is that there is considerable structure in the Maltese noun system. However, the best-performing models are either trained on only broken plurals, or they are trained to predict form classes, including classes that lump together less frequent form changes. Furthermore, all models focus on production, predicting plurals from singulars without considering words' meanings, and do not address the comprehension of Maltese nouns. In what follows, we address this broader range of questions within the framework of the Discriminative Lexicon. Before doing so, we first introduce the dataset that we used for training and evaluating our models.

2.2 Dataset

The dataset consists of all broken plurals listed by Schembri (2012) and all tagged nouns from the MLRS Korpus Malti version 2.0 and 3.0 (Gatt & Čéplö, 2013). The resulting list of nouns was then enriched with information extracted from a Maltese online dictionary (*Ġabra*, Camilleri, 2013) using the free corpus tool *Coquery* (Kunter, 2017), resulting in a dataset with singulars, their corresponding plurals and their glosses. Subsequently, the dataset was manually extended with information about the type of plural (broken vs. sound), CV structure, number of occurrences (based on the Korpus Malti v. 2.0 and 3.0), origin (Semitic vs. Non-Semitic) and grammatical gender (based on Aquilina (1987)), number, concreteness (abstract vs. concrete), and type of noun (verbal noun, dual noun, suppletive noun, or collective noun).

The resulting dataset contains 6511 word forms in total: 3364 plurals, 3132 singulars and 15 dual forms. Of the 3364 plurals, 892 are broken plural forms while 2458 are sound plural forms (with a total of 11 different sound plural types and 11 different broken plural types), reflecting the proportion of plural types in use in Maltese. The remaining 29 nouns of our dataset labeled as plurals have plurals that are neither of the broken nor of the sound type: 8 of these words have a double plural marking, that is a combination of a broken plural type and a sound plural type, e.g. *sema* (sg.) - *smewwiet* (pl.) 'sky'. 15 words are dual forms, such as *id* (sg.) - *idejn* (dual) 'hands', and 6 words have a suppletive plural, e.g. *mara* (sg.) - *nisa* (pl.) 'women', see Borg and Azzopardi-Alexander (1997) for further details.

3 Predicting Maltese noun inflection

The models for the Maltese plurals reviewed in section 2.1 all seek to predict the appropriate form of a plural from its corresponding singular. However rules for building forms from other forms may be useful for the teaching of a second language, but it is far from clear that native speakers and young L1 learners would follow the same procedure (Blevins, 2016; Dell, 1986; Levelt et al., 1999; Zamuner et al., 2011). The DL model proposed by Baayen et al. (2019) takes as its point of departure that the task of morphology is to explain how listeners understand complex words, and how speakers produce them. In other words, the DL focuses on understanding words' meanings given their forms, and producing words' forms given their meanings. Furthermore, the relation between form and meaning is modeled as immediate, without any further intervening layers of representations.

The mappings that the DL sets up between numeric vectors representing forms and numeric vectors representing meanings are the simplest mappings possible. When conceptualized as an artificial neural network, we have form units (representing dimensions of form) and semantic units (representing dimensions of meaning), with full connectivity between the two sets of units. There are no hidden layers whatsoever.

The mappings of the DL can also be understood as implementing multivariate multiple regression. For a given set of n words and m dimensions in which differences in form are expressed, we bring together their numeric form vectors into an $n \times m$ form matrix $\boldsymbol{C}$. Given k-dimensional vectors representing words' meanings, we set up an $n \times k$ semantic matrix $\boldsymbol{S}$. We can now define a $m \times k$ mapping $\boldsymbol{F}$ that takes the vectors in $\boldsymbol{C}$ and transforms these vectors, as best as it can, into the semantic vectors of $\boldsymbol{S}$. This is accomplished by solving the set of equations $\boldsymbol{CF} = \boldsymbol{S}$. The matrix $\boldsymbol{F}$ consists of the β coefficients of the multivariate multiple regression model. These coefficients can be associated one-to-one with weights on the connections in the network between form and meaning units. For production, the DL model posits a $k \times m$ mapping $\boldsymbol{G}$ from the meaning vectors $\boldsymbol{S}$ to the form vectors in $\boldsymbol{C}$. This matrix is estimated by solving $\boldsymbol{SG} = \boldsymbol{C}$. For all but the smallest toy examples, the predicted form vectors $\hat{\boldsymbol{C}} = \boldsymbol{SG}$ will only approximate the targeted gold-standard form vectors $\boldsymbol{C}$, which is why, following statistical practice, we use the notation $\hat{\boldsymbol{C}}$ rather than $\boldsymbol{C}$. The same holds for the predicted semantic vectors $\hat{\boldsymbol{S}}$. Nevertheless, the estimated weights are optimal, in the sense that they minimize the mean squared error. They represent the 'endstate' of learning that the artificial neural network can achieve by endlessly iterating through the training data with an incremental learning rule such as those of Rescorla-Wagner (Rescorla & Wagner, 1972) and Widrow-Hoff (Widrow & Hoff, 1960). In what follows, we refer to the learning of the mappings using the mathematics of multivariate linear regression as 'Linear Discriminative Learning' (LDL).

3.1 Constructing the form matrix

	Lexeme	Number	Gender
kelb	KELB	singular	M
kelba	KELB	singular	F
klieb	KELB	plural	M

Table 2: Paradigm for the Maltese noun *kelb* 'dog'.

To illustrate the central concepts of LDL, consider the Maltese toy lexicon listed in Table 2. This lexicon consists of a singular word for a male dog, a singular word for a female dog and the plural word for both.[2] The

[2] Note that not all Maltese singular nouns show an opposition of masculine and feminine gender, in many cases only one form is available (see, e.g., Borg & Azzopardi-Alexander, 1997).

first modeling step is to make a decision as to how these word forms can be represented as numeric vectors. One possibility is to decompose word forms into triphones, which target, in a crude way, context-sensitive phone representations. For our example lexicon, there are 11 distinct triphones. We couple each distinct triphone with a form dimension. Words that contain a given triphone receive the value 1 for this dimension, and otherwise the value 0. For our example lexicon, we obtain the following form matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \begin{array}{c} \\ \textit{kelb} \\ \textit{kelba} \\ \textit{klieb} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} \#ke & kel & elb & lb\# & lba & ba\# & \#kl & kli & lie & ieb & eb\# \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array}$$

In this form matrix, the hash mark # represents word boundaries.

Instead of representing words' forms by indicating which triphones are present, we can set up form vectors that decompose a word's form into its constituent syllables. Below, we report results for simulations using these two ways of representing word form information.

3.2 Constructing the semantic matrix

The row vectors of the semantic matrix $\boldsymbol{S}$ represent a word form's meaning numerically. Within the general framework of distributional semantics, many algorithms are now available for deriving semantic vectors (known as embeddings in computational linguistics) from corpora (Baroni et al., 2014; Bojanowski et al., 2017; Joulin, Grave, Bojanowski, Douze, et al., 2016; Joulin, Grave, Bojanowski, & Mikolov, 2016; Mikolov et al., 2013; Pennington et al., 2014; Yang et al., 2017). In the present study, we explore two kinds of semantic vectors: vectors that we constructed ourselves in a linguistically informed way, and ready-made vectors that were generated with `fasttext` (Joulin, Grave, Bojanowski, Douze, et al., 2016; Joulin, Grave, Bojanowski, & Mikolov, 2016).

3.2.1 Simulated vectors

The row vectors of the semantic matrix $\boldsymbol{S}$ represent words' meanings in a high-dimensional space. We can simulate such vectors, using a random number generator. For our example lexicon, we generated 11-dimensional vectors, matching the dimensionality of the form matrix $\boldsymbol{C}$. The result is simply a table with numbers:

$$
\boldsymbol{S} = \begin{array}{c} \\ kelb \\ kelba \\ klieb \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 \end{array} \\
\begin{pmatrix}
0.46 & 4.16 & 8.50 & -4.46 & 8.96 & -4.11 & 8.42 & 9.21 & -25.75 & 15.83 & -14.93 \\
0.61 & -11.93 & 8.09 & 1.00 & 3.44 & -11.98 & 8.72 & -4.75 & -33.29 & 10.39 & -2.12 \\
5.67 & 9.84 & 11.26 & 0.85 & 10.69 & -4.24 & 0.21 & 4.81 & -26.47 & 10.82 & -11.76
\end{pmatrix}
\end{array}
$$

The simplest way in which these vectors could be generated is by creating vectors of 11 random numbers sampled from a standard normal distribution. Unfortunately, this would imply that all forms have meanings that are all completely distinct: random vectors are basically orthogonal.

In order to justice the semantic similarities between words that arise due to shared inflectional features, following Baayen et al. (2019), we construct separate semantic vectors for each inflectional feature-value pair. In our running example we have two numbers and two genders, and therefore we simulate a vector representing singular number, a vector representing plural number, a vector representing masculine gender, and a vector representing feminine gender. The semantic vectors of the words in our example lexicon given above (matrix $\boldsymbol{S}$) were obtained by adding the pertinent inflectional vectors to the vectors of the lexemes:

$$
\begin{array}{llllll}
kelb\text{:} & \overrightarrow{\text{KLB}} & + & \overrightarrow{\text{SINGULAR}} & + & \overrightarrow{\text{MASCULINE}} \\
kelba\text{:} & \overrightarrow{\text{KLB}} & + & \overrightarrow{\text{SINGULAR}} & + & \overrightarrow{\text{FEMININE}} \\
klieb\text{:} & \overrightarrow{\text{KLB}} & + & \overrightarrow{\text{PLURAL}} & + & \overrightarrow{\text{MASCULINE}}
\end{array}
$$

The resulting semantic vectors $\overrightarrow{\text{KELB}}$, $\overrightarrow{\text{KELBA}}$, and $\overrightarrow{\text{KLIEB}}$ now reflect inflectional similarities: singulars will be more similar to other singulars than to plurals, and masculine nouns will be more similar to each other than to feminine nouns. Finally, a small amount of random noise is added to the vectors of individual words to respect word-specific variations.

3.2.2 Corpus-based vectors using `fasttext`

Although such simulated vectors have been found useful for modeling morphological processing in comprehension and production, they make the simplifying assumption that all base word lexemes are semantically unrelated: their simulated semantic vectors are almost completely orthogonal. Instead of working with simulated vectors, Baayen et al. (2019) derived semantic vectors for both content lexemes and inflectional functions such as singular and plural by first morphologically tagging a corpus (the TASA corpus, Ivens & Koslin, 1991), and then using a method from distributional semantics to construct semantic vectors for both content words and for the inflectional (as well as derivational) functions identified by the tagger. Since computational resources for Maltese are limited, for the present study, we complemented modeling using simulated vectors with modeling using ready-made vectors that were created with `fasttext` (Joulin, Grave, Bojanowski, Douze, et al., 2016; Joulin, Grave, Bojanowski, & Mikolov, 2016).

The algorithm underlying `fasttext` constructs semantic vectors for words from semantic vectors of substrings of words. This enables the algorithm to pick up, to some extent, morphological effects. Modeling with `fasttext` vectors therefore has as advantage, compared to simulated vectors, that LDL will now be able to take into account similarities in meaning between content words, but a potential disadvantage is that it might not pick up inflectional semantics as precisely.

We extracted the `fasttext` vectors that are available for 4,056 of the nouns in our dataset[3], of which 2266 are singulars and 1781 are plurals. These `fasttext` embeddings are 300 dimensional vectors.

In order to obtain some insight in how well `fasttext` captures inflectional semantics, we projected the 300-dimensional `fasttext` space onto a 2-dimensional plane using Principal Components Analysis. A scatterplot of nouns in this plane, color-coded for number and plural type, is shown in Figure 1. Interestingly, we find distinguishable clusters of singulars (light blue) and plurals (orange, violet), albeit with considerable overlap. In addition, sound plurals (orange) and broken plurals (violet) seem to dwell in somewhat different semantic subspaces as well. This is confirmed by a Linear Discriminant Analysis (LDA), which showed that a classification of singular, sound plural and broken plural words using the first fifty principal components reaches 85% classification accuracy. Apparently, number and type of plural are to some extent intertwined with word meaning. This interaction of regularity with semantics replicates a similar interaction for English regular and irregular verbs reported by Baayen and Moscoso del Prado Martín (2005). We cannot rule out, however, that the present finding for Maltese is a straightforward consequence of the way in which `fasttext` constructs semantic vectors. Because `fasttext` takes sublexical strings into account, it is conceivable that the semantic differences that we observe between broken plurals and sound plurals are simply reflecting form differences between broken plurals and sound plurals. On the other hand, it might be argued that the relevant sublexical strings taken into account by `fasttext` must have different distributional properties in Maltese, otherwise it would be impossible for the semantic vectors of broken plurals and sound plurals to show different clustering in semantic space. Nevertheless, replication of the present inter-

[3]available at https://fasttext.cc/docs/en/crawl-vectors.html

action of plural type and semantics using, for instance, `word2vec` (Mikolov et al., 2013), would strengthen the present tentative conclusions for Maltese.

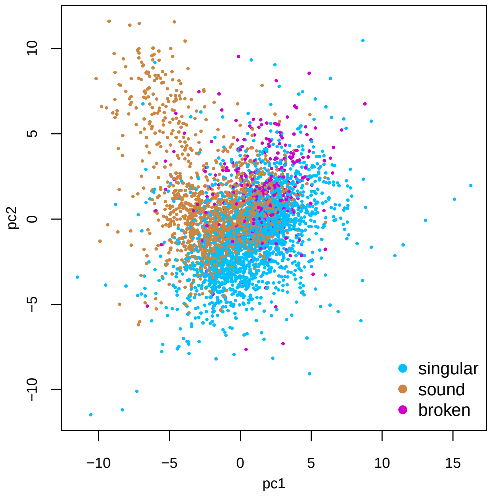

Figure 1: Projection of `fasttext` semantic vectors onto a two-dimensional plane. Number and plural types (sound and broken) are color-coded. Singulars and broken plurals cluster more to the right on PC1, whereas sound plurals and broken plurals cluster more to the top on PC2.

Figure 2 addresses how well `fasttext` captures differences in gender. Despite substantial overlaps, LDA, again using the first fifty principal components, achieved a classification accuracy of 79% and 70% for singular and plural words respectively. For the other semantic features labeled in our dataset (concreteness, verbal noun, collective noun), however, due to the fact that usually one level has overwhelmingly more tokens than the other, no clustering in the semantic space could be observed.

3.2.3 Evaluating model performance

Before reporting how well the DL model approximates the Maltese noun system, we need to explain how we evaluate model performance.

To evaluate comprehension, we calculated the correlations between a given word's predicted semantic vector ($\hat{s}_i$) and all the gold standard semantic vectors in the lexicon (the row vectors of $\boldsymbol{S}$). If $\hat{s}_i$ has the highest correlation with the semantic vector of the targeted word (s_i), comprehension is considered successful. On the other hand, unsuccessful comprehension occurs when the highest correlation is with another word than target word. It should be noted that for homophones, we consider comprehension correct as

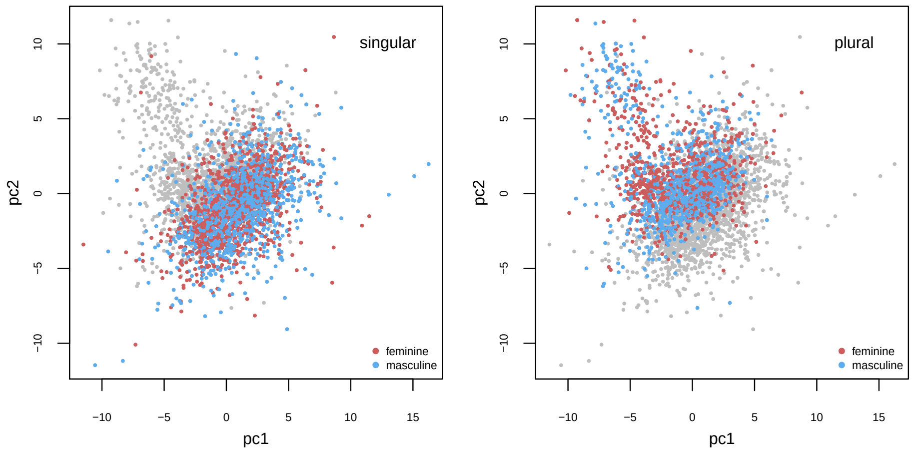

Figure 2: Projection of `fasttext` semantic vectors onto a two-dimensional plane, spanned by the first two principal components. The left panel plots singular feminine (red) and masculine (blue) words, and the right panel plots plural words. PC2 captures to some extent plurality, whereas PC1 captures aspects of gender, resulting in somewhat differentiated clustering within number for feminine vs masculine nouns.

long as $\hat{s}_i$ has the best correlation with one of the homophone meanings, e.g. Maltese *xark* 'shark' [ʃaɐk] and *xark* [ʃaɐk] 'a person who conducts business shrewdly or acts for their own material benefit' (note that there also is a Semitic word to express 'shark' available in Maltese: *kelb il-baħar*). This is because here we are modeling the processing of words in isolation. Given that it is not possible to recognize a specific homophone meaning out of context, we therefore adopted this lenient evaluation metric for comprehension.

With respect to production, as a first step, we generated for each word i the predicted form vector $\hat{c}_i$ from its semantic vector $\hat{s}_i$. This predicted form vector, however, only provides information about the amount of semantic support for the sublexical cues (such as triphones or disyllables); it does not inform us about the order in which well-supported cues have to be placed for articulation. For ordering, the model makes use of the order information that is implicit in the sublexical cues. Take triphones, for example. The triphone `kel` can be followed by `elb` (to form the word *kelb*), given the identity of the final diphone `el` in `kel` and the initial diphone `el` in `elb`. In the absence of such overlap (e.g., for `kel` and `lie`), no sequential ordering is possible. As the lexicon becomes larger, the number of possible triphone combinations also grows, resulting in multiple candidate forms for a given form vector $\hat{c}_i$. The candidate selecting for articulation is chosen such that it best realizes the meaning the speaker has in mind. Technically, this is accomplished by first generating, for each candidate form ω_j its predicted semantic vector $\hat{s}_j$, and then selecting from these semantic vectors the one that is most similar to the targeted semantic vector $\boldsymbol{s}_i$ that is to be expressed. In short, we generate the predicted semantic vectors for all candidate forms and select as model prediction the form vector associated with the predicted semantic vector that has the closest meaning to the targeted meaning (a process called 'synthesis-by-analysis' by Baayen et al., 2018).

For the simulations presented in this study, we used the **JudiLing** package, an implementation of LDL in the `Julia` language (Luo et al., 2021). For production, we used the `learn_paths` function for ordering sublexical features into words. This algorithm takes predicted form vectors, and learns to predict at what position(s) in a word a sublexical cue occurs. In this way, each of a word's sublexical cues is associated with a number reflecting how well it is supported for its position in the word. We refer to this number as a cue's positional support. Only cues with sufficient positional support are taken into account when assembling the set of word candidates. What counts as sufficient positional support is determined by a threshold value θ: Only words with a positional support exceeding θ are taken into consideration. More detail about the `learn_paths` algorithm can be found in Luo et al. (2021). In Section 5.3, we will show that the total amount of positional support for a word's cues is predictive for reaction times to Maltese plurals.

4 Modeling results

4.1 Evaluation on training data

With two cue representations (one using triphones and one using bisyllables) and two semantic representations (one using simulated semantic vectors and the other using `fasttext` vectors), we have in total four models. Comprehension and production accuracies for the full dataset are presented in Table 3. For comprehension, bi-syllables as cues yielded higher accuracies than triphone as cues, regardless of the kind of semantic representation. With respect to production, we also see an advantage of using bisyllables as cues; now model performance is slightly better with simulated vectors than with `fasttext` vectors. However, as the threshold (cf. Section 3.2.3) for the pathfinding algorithm was also set differently for the simulated and `fasttext` vectors, the scores are not readily comparable. By lowering the threshold even further, at considerable computational costs, the performance of models using `fasttext` vectors can be improved further. What this shows is that the simulated vectors are more distinct, and easier to learn, than the `fasttext` vectors.

Given the high accuracy of all four models, we can conclude that the model generally has a good memory for understanding and producing Maltese plurals. However, we do not know how the model performs with respect to inflected forms that it has not encountered during training. In other words, we do not yet know to what extent the model is productive.

	comprehension		production	
	simulated	fasttext	simulated	fasttext
triphone	93.8%	95.6%	92.7%	91.2%
bi-syllable	99.95%	99.93%	100%	96.3%

Table 3: Model performance for comprehension (left) and production (right) for four combinations of cue and semantic representations. For production, the threshold was set to 0.1 for the models with simulated semantic vectors and to 0.005 for the `fasttext` models.

4.2 Evaluation on held-out data

The question of whether our model is productive for Maltese is of considerable theoretical interest because the noun system of Maltese is quite irregular. Although some rules can be formulated, indicating that the system is not random, the many patterns for broken plurals and the wide variety of plural exponents characterize a system for which full productivity cannot be expected. It would actually be strange and worrisome if the models were to be able to predict unseen forms with close to 100% accuracy. Since regularity is generally seen as a prerequisite for productivity, the Maltese noun system is at best semi-productive. This is confirmed by the observation that native speakers of Maltese are often unsure about what the proper plural of an unknown or infrequent word might be, as indicated by the production study in Nieder, van de Vijver, et al. (2021a). In the light of these considerations, a substantial drop in prediction accuracy is expected for held-out data, compared to the accuracy for the training data.

We also expect to find that for held-out data, production accuracy will be somewhat lower than comprehension accuracy. The familiar asymmetry between production and comprehension (Boersma, 1998; Pater, 2004; Smolensky, 1996) was already visible in the results for the training data for the simulations using `fasttext` vectors (see Table 3), and we anticipate it will be present, and perhaps more pronounced, for the held-out data.

To examine whether and to what extent the model is productive for Maltese nouns, we held out 10% of the words in our dataset as testing data. The held-out words were selected based on the criterion that all the sublexical cues and inflectional features of the words have already been available to the model during training. Furthermore, the held-out words were constrained to have lexemes that occurred in the training data.

The testing data contained 298 singular forms and 353 plural forms. Of the plural forms, 332 were sound plurals, 20 were broken plurals, and one had a double plural marking (recall that this is a combination of a broken plural template with a sound plural suffix). Since we do not have `fasttext` vectors for all the words in our dataset, we opted for using simulated vectors rather than `fasttext` vectors. For the representation of words' forms, we made use of bi-syllables.

Comprehension accuracy for the held-out data was 73%. For 33.5% of the errors, the target meaning was ranked among the top five candidate meanings.

For production, accuracy was substantially lower, at 57%. The correct form appeared among the top 10 candidates in only 66% of the words for which the model's proposed candidate was incorrect. Singular forms were produced slightly more accurately than plural forms (61% vs 54%). Within the set of plural forms, while more than half of the sound plural forms were correctly predicted by the model, broken plurals appeared to be especially difficult, as only one of the twenty broken plural forms in the held-out data was produced correctly. An overview of all prediction error types produced by the LDL model is given in table 4.

Error	Target	Predicted (LDL)	Lexeme	Proportion of errors
incorrect word	qta.tes	ġu.ga.rel.li	cats toys	72% (202 of 279)
	fes.ti.ni	par.ti.ti	party (festival) party (political)	
alternative plural	su.fiet	swa.fi	wool	11% (31 of 279)
wrong affix	tu.nel.la.ta	tu.nel.la	ton	8% (22 of 279)
phonetically close	nuf.fa.ta	nuf.fa.ra	blister scarecrow	5% (13 of 279)
alternative singular	o.pe.ra	o.pra	opera	2% (6 of 279)
plural	dnub	dnu.biet	sin(s)	2% (5 of 279)

Table 4: Prediction errors of the LDL model

The model produced a total of 279 errors. 72% of these errors concern incorrect word forms (see rows 1 and 2 of table 4). While some of these incorrect word forms are actual Maltese words, in a few cases LDL produced new word forms, e.g., the pseudo-word *peri*, or produced a word form that has the same meaning. Both *kursar* and *furban*, for instance, are the masculine singular form of 'pirate', and their plural forms are both sound plurals, *kursara* and *furbani* respectively. As a result of this synonymy, the model got confused about which form to use. However, the confusion is to be expected as the semantic vectors for the synonyms in our model are almost identical.

For 11% of the errors, the model predicted an alternative plural for a word that has multiple plural forms. For example, the word *tarf* 'edge (sg)' has two broken plural forms, *traf* and *truf*. While the target form was *truf*, the model produced the alternative form *traf*.[4]

Other errors involve mixing up between sound and broken plural forms. For instance, *qalba* 'core (sg)' has three plural forms, one sound plural (*qalbiet*) and two broken plurals (*qlub* and *qliebi*). The testing data contained both broken plural forms, and the model predicted the sound form for both of them. Such attraction to sound plurals also occurs for the word *tajra* 'fowl (sg)'. Instead of the targeted broken plural form *tjur*, the model predicts the alternative sound plural form *tajriet*. This pattern of results is in line with the finding of Nieder, van de Vijver, et al. (2021a), in which native speakers tend to use frequent sound plural patterns for novel words.

In 8% of the errors, the model produced a word form with additional phonological material, resulting in an error we labeled *wrong affix*. For example, for the target word form *kalamiti* 'magnet (pl)' the model instead predicted the word form *kalamitati* with an additional diphone -*at*- before the correct sound plural ending -*i*.

We labelled 5% of the errors as *phonetically close*. In these cases, LDL predicted a word form that is phonetically similar to the target word form, e.g. *mera* 'mirror' instead of *mara* 'woman'. In a few cases, 2% of all errors, LDL predicted an alternative singular form for a word that has multiple singular forms in our data set, e.g. *lanza* for *lenza* 'fishing-line (sg)'. In another 2% of the cases the errors are a result of the model producing the plural form while the target word form that should have been predicted is a singular word, e.g. *karbiet* instead of *karba* 'groaning (sg f.)'.

When we discount cases where the model produced a non-targeted but otherwise correct form, and cases where the model produced a synonym, production accuracy for the held-out data was 69%. Compared to comprehension accuracy for the held-out data (73%), we observe a lower accuracy for production, as expected.

[4]Although *traf* is listed in Maltese dictionaries (see *Ġabra* (Camilleri, 2013) and Aquilina (1987)), some native speakers find this form unacceptable.

4.3 Discussion

Computational modeling of Maltese noun inflections with LDL showed that it is possible to model comprehension and production of Maltese nouns by considering mappings between form and meaning within the general perspective of Word and Paradigm morphology. Model performance is excellent with the training data. For the held-out data, the model understands and produces unseen form with satisfactory accuracy, around 70%, an accuracy that actually is surprisingly high for a noun system that is irregular in many ways and only semi-productive. Compared to previous modeling results obtained within the framework of Word and Paradigm morphology (Nieder, Tomaschek, et al., 2021), accuracy is much higher than that of an Encoder-Decoder deep learning model, but lower than the exemplar-based model implemented with TiMBL. This model, however, was given a much simpler task, namely, to predict classes of form changes, including classes bringing together many low-frequency patterns of change. In comparison to data from real speakers, the LDL model results on held-out data reflect the uncertainty of native speakers when it comes to infrequent words (Nieder, van de Vijver, et al., 2021a). Nieder, van de Vijver, et al. (2021a) asked participants to provide plural forms for given singulars. They found that participants often were not able to provide the correct plural for existing infrequent singulars.

We conclude that the theory of the Discriminative Lexicon provides a useful framework for predicting what forms are possible for listeners to understand, and what forms are possible for speakers to produce.

In what follows, we address the question of whether the way in which the discriminative lexicon model formalizes listening and speaking (admittedly at a high level of symbolic abstraction, especially when it comes to the representation of words' forms) can contribute to our understanding of human lexical processing. In the next section, we therefore examine whether measures derived from the model can contribute to enhancing statistical models fitted to response latencies in a primed lexical decision experiment with Maltese nouns.

5 Modeling Maltese priming reaction times

5.1 Dataset

For exploring the usefulness of our computational model for understanding actual lexical processing, we re-analyzed the data from a primed cross-model lexical decision task reported by Nieder, van de Vijver, et al. (2021b). Their dataset contains 7885 observations (after removal of incorrect answers, practice trials and outliers) from fifty-nine participants. The dataset provides reaction times for two frequent sound plural suffixes (*-i* and *-ijiet*), and for two infrequent sound plurals (*-a* and *-at*). Likewise, the data cover two frequent broken plural templates (*CCVVCVC* and *CCVVC*) and two infrequent broken plural templates (*CCVjjVC* and *CCVVCV*). In the following, when using the frequency of suffixes and templates as a variable for the model, we will use the terms "pattern frequency" and "patterns" to refer to both suffixes and templates. Nieder, van de Vijver, et al. (2021b) make use of a cross-model lexical decision task with auditorily presented primes and visual targets. All target words were singulars, primes were either unrelated controls, or plural forms corresponding to the singular targets. For the present study, we only used the reaction times for corresponding singular-plural pairs, omitting the control condition. This left us with 3995 observations. We then removed all words for which we did not have `fasttext` semantic vectors, resulting in a dataset with 2951 observations.

5.2 A baseline model

We fitted a generalized additive mixed model (GAMM) (Wood, 2017) to the inverse-transformed RTs, with `pluralType` (sound/broken), `patternFreq` (frequent/infrequent plural pattern), `targetFreq`, and `targetLen` as fixed-effect predictors and covariates. We allowed the two categorical factors to interact, and allowed the covariates to have nonlinear effects by using thin plate regression spline smooths. We also included by-subject random intercepts. A summary of the resulting model is displayed in Table 5.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-1.786	0.038	-47.487	< 0.0001
pluralTypesound	0.008	0.018	0.464	0.642
patternFreqinfrequent	0.042	0.0175	2.389	0.017
pluralTypesound: patternFreqinfrequent	-0.012	0.0285	-0.417	0.677
B. smooth terms	edf	Ref.df	F-value	p-value
s(targetFreq)	3.048	3.567	14.350	< 0.0001
s(participant)	56.277	58.000	35.060	< 0.0001

Table 5: Summary of a GAMM fitted to inverse-transformed RTs, with plural type, pattern frequency, the frequency and the length of target words as predictors, and by-participant random intercepts.

For this smaller dataset there is no evidence that RTs differ for broken

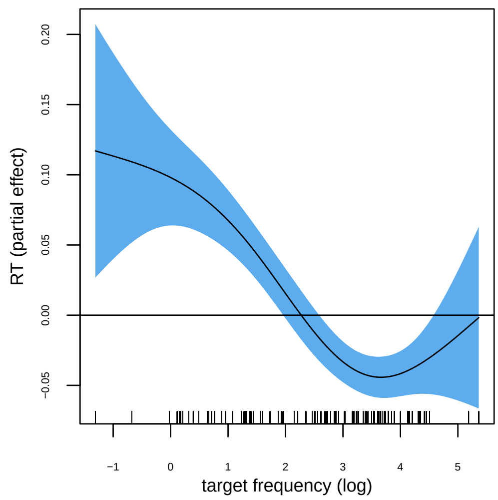

Figure 3: The effects of target frequency on RTs.

and sound plurals, replicating the findings of Nieder, van de Vijver, et al. (2021b) for the original full dataset. The coefficient of pattern frequency indicates that plural primes with infrequent patterns induced longer response times compared to plural with frequent patterns. Figure 3 presents the partial effect of target frequency. The effect is nearly linear, but levels off for the highest-frequency words, a pattern often observed for word frequency in lexical decision tasks (see, e.g., Baayen et al., 2006).[5] Akaike's information criterion for this model was 1920, and the model's Maximum Likelihood score was 1033.

In summary, our baseline model indicates that the pattern frequency of the prime plural and the word frequency of the singular target are predictive for the response latencies, with greater frequencies affording faster responses. This result fits well with usage-based approaches to the lexicon (Bybee, 2010). Primes with infrequent patterns are learned less well, and are therefore less successful in facilitating lexical access to the target. More frequent target words are more entrenched, and hence can be retrieved more quickly from the mental lexicon.

[5]Model fit can be further improved by including by-target random intercepts. However, due to very high concurvity, this model becomes uninterpretable: The covariates do not explain anything that is not already explained by the word-specific random intercepts. In this model, as well as in the model reported below, we therefore did not include by-target random intercepts.

5.3 Predicting reaction times with LDL predictors

For predicting reaction times with measures based on discriminative learning, we opted for using the model with bisyllables as cues, and `fasttext` word embeddings as semantic vectors. We used `fasttext` vectors because, unlike simulated vectors, as we demonstrated above, they are remarkably sensitive to semantic differences between stems, number, plural types (broken vs. sound), and gender.

There are several potential measures that can be derived from an LDL model (see Chuang & Baayen, 2021, for an overview). We found two measures particularly useful for understanding Maltese primed lexical decision latencies, one measure quantifying how well primes' forms can be learned, and the other measure quantifying how closely the meaning of the prime plural already approximates the meaning of the target singular.

First consider the form measure, henceforth labeled `prime support`. The measure is defined as the sum of the positional semantic supports that the bi-syllable cues of a given plural prime word receive. By way of example, the word *trabi*, the plural form of *tarbija* (sg.f.) 'baby', contains three bi-syllable cues: `#.tra`, `tra.bi`, and `bi.#`, at positions 1, 2, and 3, respectively ("." denotes syllable boundaries). As described in Section 3.2.3, the `learn_paths` function in the **JudiLing** package calculates, for each cue position, the amount of support that a bi-syllable cue of the target word receives. That is, given the semantics of `trabi`, the positional support measure quantifies how certain the model is that `#.tra` should occur at position 1, `tra.bi` at position 2, and `bi.#` at position 3. For this example, the positional supports that the three bi-syllable cues receive are 0.25, 0.20, and 0.28, respectively. The `prime support` measure sums these three individual supports (i.e., $0.25 + 0.20 + 0.28 = 0.73$). The larger the `prime support` is, the more predictable a prime word's form is given its semantics, and the better its form can be learned.

The second measure, henceforth labeled `pre-activation distance`, addresses the relation between plural prime words and their corresponding singular target words. It gauges the extent to which listening to a prime plural word semantically pre-activates (or "primes") the meaning of the target singular word. The `pre-activation distance` is defined as the Euclidean distance between the predicted semantic vector of a prime word and the gold standard semantic vector of its target word ($\mathrm{dist}(\hat{s}_{prime}, s_{target})$). A large value of this measure indicates that the predicted meaning of the plural prime word is far away in semantic space from the meaning of the singular target word. Conversely, a small pre-activation distance indicates that the prime word already closely approximates the meaning of the target word.

This measure is inspired by a similar measure proposed in Baayen and Smolka (2020) on the basis of a naive discrimination learning network, `prime-to-target pre-activation`, which calculates the extent to which a target word is already activated by the cues of the prime word. We also calculated the prime-to-target pre-activation measure for the present Maltese data, but, although significant, its use led to a substantially worse fit by nearly 50 AIC units (details are available in the supplementary materials: https://osf.io/rxsbu/). For the usefulness of a distance measure rather than a correlation (or cosine similarity) measure, see the study by Chuang, Vollmer, et al. (2021) on the semantics of English pseudowords.

We fitted a GAMM to the inverse-transformed response latencies with prime support and pre-activation distance as covariates, along with by-participant random intercepts. Incremental model fitting revealed that goodness of fit increased by allowing the two covariates to interact (the AIC of the model decreased by 17 units, amounting to an evidence ratio of 4915). The summary of the resulting model is presented in Table 6. Figure 4 visualizes the partial effect of the interaction.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	-1.7664	0.0364	-48.4714	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
te(primeSupport, preActDist)	12.5678	15.2192	6.5171	< 0.0001
s(participant)	56.3071	58.0000	35.1927	< 0.0001

Table 6: Summary of a GAMM fitted to RTs, with the interaction between prime support and pre-activation distance as fixed-effect predictors, and by-participant intercept as a random-effect predictor.

The effect of prime support emerges as roughly ∩-shaped: with increasing prime support, RTs first increase (indicated by warmer colors) and then decrease (indicated by colder colors). Where the maximum is reached depends on pre-activation distance: for greater values of pre-activation distance, the maximum appears for greater values of prime support. Furthermore, especially the decrease in RTs for larger values of prime support is more noticeable for low values of pre-activation distance. With respect to pre-activation distance, here we see, across all values of prime support, that reaction times decrease as pre-activation distance increases.

At first sight, these results are puzzling. One might have expected that if the prime fails to pre-activate the target, reaction times should be longer, but in reality, they are shorter. Furthermore, why is it that the effect of prime support is ∩-shaped, instead of simply facilitating response latencies

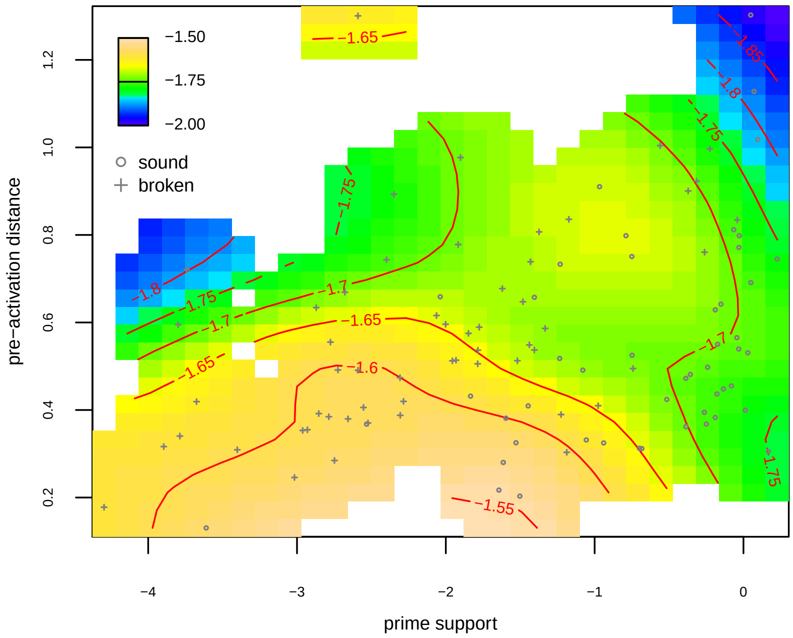

Figure 4: The interaction effect of prime support and pre-activation distance on RTs.

to the target?

To answer these questions, we need to take a step back and have a critical look at the priming paradigm. Priming is often understood as involving facilitation of lexical access to the target. This is how we interpreted the priming effect in our baseline model. However, in general, compared to an identity baseline, primes typically give rise to longer instead of shorter response latencies. Primes are only facilitating when they are compared to an unrelated control baseline. In other words, unrelated primes are more disruptive than related primes, and related primes are more disruptive than identity primes.

The interpretation of primes as disrupting and interfering with normal lexical processing is supported by the experiments reported by Libben et al. (2018). Their study made use of primed visual lexical decision, with two-constituent compounds as target words and one of their constituents as primes. They observed longer reaction times for more frequent primes, in combination with the usual shorter reaction times for more frequent target compounds. In other words, their experiment indicates that the more frequent a prime is, the more it disrupts the processing of the target.

With respect to the present experiment, a greater pre-activation distance likewise bears witness to a similar disruptive effect of the prime. The greater the distance of a prime to its target, the faster participants were able to make a lexicality decision. Since the plural and singular of a word are semantically highly similar in comparison to an unrelated control prime and its target words, it seems unlikely that it is the pre-activation distance by itself that is at issue. Instead, our hypothesis is that when prime and target are already quite similar, they become highly confusable and render deciding on the target's lexicality more difficult and more time-consuming.

For understanding the inverse U-shaped effect of prime support, an unprimed experiment reported in Pham and Baayen (2015) is relevant. In this experiment, Vietnamese two-constituent (or, more precisely, two-syllabeme) compounds were presented. The frequency of the compound itself emerged with the expected facilitatory effect on response latencies. However, the frequency effects of the constituents were fundamentally different in nature. Response times were shortest when both constituents had a frequency close to their mean frequency. Response times increased, the further these frequencies were away from their mean frequencies. In this experiment, it was the probability of a constituent's frequency that determined reaction times, not the frequency value itself. Response times were fastest for highly probably constituent frequencies, and slowest for very atypical constituent frequencies. This suggests that the constituents of Vietnamese compounds are interfering with the processing of the compound itself, and that lexical decision making is optimized for dealing with average interference (gauged through average constituent frequency).

In the present experiment with Maltese nouns, we may be observing a similar optimization strategy with respect to prime support, which gauges how well the pronunciation of a word is learnable given its meaning. Recall that above we observed that the maximum of the effect of prime support on the reaction times is reached for larger values of prime support as pre-activation distance is increased. As can be seen by comparing the left and right panels of Figure 5, these RT maxima are not well predicted when pre-activation distance is regressed on prime support (left panel), but they are predicted reasonably well when regressing prime support on pre-activation distance (right panel). We can therefore interpret the maximal RTs as arising for the maximal probabilities of prime support values, conditional on the prime's pre-activation distance. This conceptualization is highlighted in the right panel of Figure 5 for a pre-activation equal to 0.8. The most probable values are located around the mean, indicated by a blue dot. As one moves away from the mean, either up to higher values of prime support, or down to lower values of prime support, the values of prime support become less probable.

If our interpretation of the wiggly regression surface for reaction times

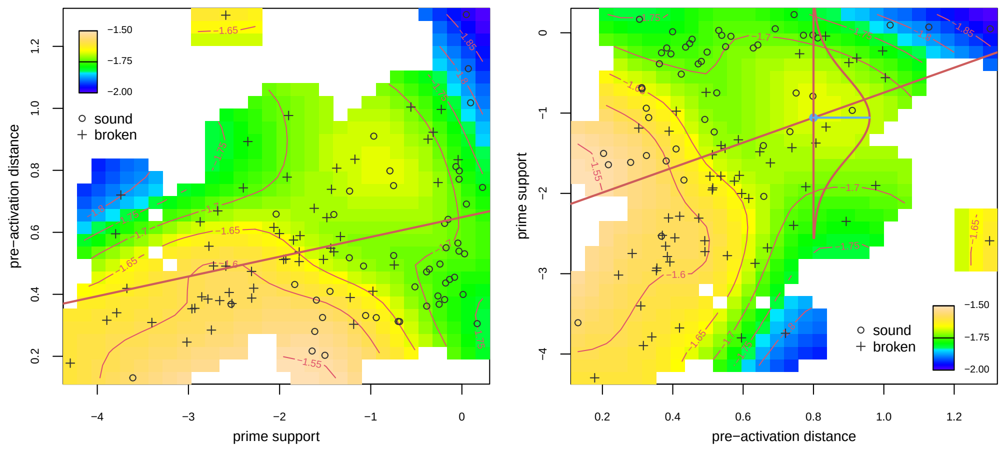

Figure 5: Regressing pre-activation on prime support fails to capture the ridge of maximal RTs (left panel), but regressing prime support on pre-activation distance provides a better approximation. The locations of the maximal reaction times can therefore be interpreted as reflecting the maximal probabilities of prime supports conditional on the values of pre-activation distance. A Gaussian density function illustrates this for the mean at pre-activation distance equal to 0.8.

in Figure 5 is correct, then what the GAMM model is telling us is that, conditional on the pre-activation distance (i.e., conditional on how close the prime's semantics approximate the target's semantics), primes with average (and thus most probable) prime support are the most disruptive: here, participants are most 'wary' to respond. As prime support increases, the prime provides stronger evidence for lexicality: for half of the trials in the experiment, the target word itself is a nonword. Participants therefore have to be cautious not to rely on the evidence for lexicality originating from the prime. Apparently, they set their decision criteria conservatively, such that they are not misled by the most probable prime supports, given the pre-activation distance.

It is important to note that this interpretation of the effect of prime support does not claim that participants are silently repeating the prime words, or that we are observing here some articulatory component residing in the speech perception process, as hypothesized by the motor theory of speech perception (Liberman & Mattingly, 1985). If that were the case, the prime support value itself should have been predictive, instead of the probability of that value. What we do claim is that in the meta-linguistic task of primed lexical decision making, participants can base their decisions in part on how well they know the pronunciation of the primes.

5.4 Discussion

How does the GAMM with LDL predictors compare to the baseline model with template frequency and target frequency as predictors? To address this question, we compared Akaike's Information Criterion (AIC) for the two models. The AIC of the baseline model is 1920, and that of the LDL-based model is 1907. The corresponding evidence ratio is 665, indicating that the LDL-based GAMM is 665 times more likely than the baseline model to minimize the information loss.

We did not include target frequency as a predictor in the GAMM with prime support and pre-activation distance, for two reasons. First, within the framework of the discriminative lexicon, there are no word units with which frequency counts can be associated. Second, for modeling, we have made use of the multivariate multiple regression method for estimating weights, which represents the endstate of learning. At the endstate of learning, for which all token frequencies have increased to infinity, frequency effects are no longer present (see Heitmeier et al., 2021; Shafaei-Bajestan et al., 2021, for a detailed discussion).

Frequency of occurrence does come into play when incremental learning algorithms are used. For the present study, we did not explore incremental learning, for two reasons. First, for representing words' meanings, we need incrementally updated semantic vectors. Unfortunately, incrementally updated `fasttext` vectors are not available for Maltese. Second, although incremental updating of the network is implemented in the **JudiLing** package for comprehension, it is not fully implemented for production. Developing a fully-fledged incremental version of the model is a target for further research. We do note, however, that when target frequency is added as predictor to the GAMM with LDL predictors, both prime support and pre-activation distance remain significant.

6 General Discussion

We conclude this study with a discussion of the new insights that our results bring to morphological theory on the one hand, and its limitations on the other.

The semi-productivity of the Maltese plural poses a challenge for computational modeling. Any system, whether based on rules, analogy, or machine learning, needs to strike a balance between providing a good memory for the forms in use, and doing justice to the extent that the system is productive. We have shown that the Discriminative Lexicon (DL) model finds such a balance: it provides an accurate memory for both the comprehension and production of known words, and it also performs reasonably well when given the task to produce or understand novel, unseen forms. Given the semi-productivity of the Maltese plural, it is actually surprising how well prediction for unseen words works. This finding supports earlier descriptive studies that have called attention to substantial regularities in the Maltese plural system (Mayer et al., 2013; Nieder, van de Vijver, et al., 2021a; Schembri, 2012).

The theory of the DL currently does not include algorithms implementing decision making in experimental tasks such as lexical decision. Nevertheless, some headway can be made by incorporating measures derived from the theory as predictors in statistical models for experimental measures such as reaction times. Two such measures, one gauging how well we know a word's form, and the other assessing how closely the meaning of the prime approximates the meaning of the target, were found to improve the quality of a GAMM model fitted to the reaction times in a primed lexical decision task. The resulting model forced us to reconsider how to understand priming. Instead of understanding primes as facilitating lexical access to the target, primes may actually be disruptive. Among highly semantically relevant singular-plural word pairs, primes that are less similar in meaning to the target give rise to reduced interference. Furthermore, subjects appear to have developed a response strategy in which they are most wary to respond to target words that have the highest probabilities of having well-learned pronunciations. It should be kept in mind, however, that these results are tentative, based as they are on a post-hoc reanalysis, using exploratory data analysis, of the experiment reported earlier by Nieder, van de Vijver, et al. (2021b), and further replication studies will be essential for consolidating the present findings.

From this set of results, we conclude that the algorithm of linear discriminative learning, previously tested on Latin (Baayen et al., 2018), Estonian (Chuang et al., 2020), English (Chuang, Vollmer, et al., 2021), German (Heitmeier et al., 2021), Indonesian (Denistia & Baayen, 2021), Kinyarwanda (van de Vijver et al., 2021), and Korean (Chuang, Kang, et al., 2021), also provides a fruitful window on non-concatenative morphology.

The approach to the Maltese plural system that we have worked out in this study, which is a computational implementation of Word and Paradigm morphology (Blevins, 2016), differs from previous studies using computational modeling in that both production and comprehension are modeled. Instead of defining the task of morphological theory as providing a formal mechanism specifying what sound sequences are possible words, the DL framework explicitly addresses two challenges, first, to predict what possible forms are, given their meanings; and second, to predict what possible meanings are, given their forms. The present study is limited by the fact that the form representations that we made use of are based on abstract sublexical features such as letter or syllable n-grams, and it is currently an open question how the model will perform when, for instance, features derived from the acoustic signal are used (see Shafaei-Bajestan et al., 2021, for an exploration).

Our study also contributes to the theory of morphological productivity. Productivity is usually investigated for specific affixes. We have shown that we can assess the productivity of a whole system by inspecting how well the model's networks generalize to understanding and producing unseen forms. Several researchers have suggested that the productivity of rival affixes (e.g., -al, -ion, -ment) should be assessed jointly (Corbin, 1983; Wurzel, 1970; Zwanenburg, 1983). The present model for Maltese provides one way in which this suggestion can be implemented: many different suffixes for sound plurals, and many different templates for broken plurals, are all considered jointly.

Inspection of the semantics of Maltese singulars and plurals, using distributional semantics, clarified that the broken plurals, sound plurals, and singulars form partly overlapping but distinguishable clusters in semantic space. Furthermore, feminine and masculine nouns show some clustering in semantic space that is slightly different for singulars and plurals. These results show that the semantic vectors of inflected words have considerably more structure than expected in approaches in which plural inflection realizes a fixed morpho-syntactic feature. As the semantic vectors that can be simulated for inflected words with the **JudiLing** package implement fixed shifts for a given morpho-syntactic feature, it is clear that such vectors capture only part of the true complexity and richness of the semantics of inflected words. Simulated vectors construct a useful scaffolding for inflected words' semantics, sufficient to set up effective mappings between form and meaning, but insufficient for modeling the details of how form and meaning interact.

Since all models, including the one we presented in this paper, are idealizations, it is useful and necessary, we think, to reflect upon the differences between our model and native speakers of Maltese. The input to our model is a list of words and their semantics, conceptualized as embeddings. The model assumes that these forms and meanings are correct for any given speaker, but, of course, this is an idealization given that actual usage varies across speakers (Bybee, 2010; Sinclair, 1991). Native speakers reported to us that they frequently hear other speakers use plurals that they had not heard before, but find understandable nevertheless.

Whereas native speakers learn continuously and incrementally, we have modeled the endstate of learning, of a learner with perfect memory and undivided attention to nouns alone. Obviously the existence of such a learner is a myth. It is possible to model incremental learning in LDL, but we do not have a sufficient amount of learning data of Maltese nouns to reliably model their learning. We leave this open for further research.

Our model represents a single (mythical) learner, but in reality there are individual differences between learners. Milin et al. (2017), for example, found evidence from skilled Russian readers that some readers accelerated as they progressed in a new text, whereas others slowed down. They connected this behavior to individual differences in the use of perceptual cues. Such individual differences in the use of cues would also affect acquisition of Maltese nouns. This could be modeled by learner-specific thresholds determining the number of candidate forms a speaker is willing to take into consideration (see also Chuang et al., 2020).

Keeping its limitations in mind, we contend that our model is useful as a quantitative tool for investigating high-level properties of human learning. Our model not only goes beyond predicting possible forms given another form, as is usual in computational models of morphophonology, but also provides model-based measures that predict human processing.

We conclude with noting that the LDL learning engine of the DL model strives for simplicity and interpretability. Formally, this engine carries out multivariate multiple linear regression on form and meaning. The assumption that mappings between form and meaning are linear undoubtedly involves substantial simplifications. Nevertheless, as illustrated in the present study, this simple approach already works surprisingly well, suggesting that the noun system of Maltese itself is also roughly 'linear'. Because the architecture of the network is fixed, the performance of the model is completely determined by the representations selected by the researcher for representing form and meaning, and the data. This, we think, makes the model especially useful as a tool for linguistic analysis.

Data Availability Statement

The data that support the findings of this study are openly available at https://osf.io/rxsbu/.

References

Aquilina, J. (1987). *Maltese-English dictionary*. Midsea Books.

Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, *11*, 295–328.

Baayen, R. H., Chuang, Y.-Y., & Blevins, J. P. (2018). Inflectional morphology with linear mappings. *The Mental Lexicon*, *13*(2), 230–268.

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

Baayen, R. H., Feldman, L., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *53*, 496–512.

Baayen, R. H., & Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, *81*, 666–698.

Baayen, R. H., & Smolka, E. (2020). Modeling morphological priming in German with naive discriminative learning. *Frontiers in Communication*, *5*, 17.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 238–247.

Belth, C., Payne, S., Beser, D., Kodner, J., & Yang, C. (2021). The greedy and recursive search for morphological productivity. *arXiv*. https://arxiv.org/abs/2105.05790

Bird, H., Lambon Ralph, M., Seidenberg, M., McClelland, J., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language*, *48*, 502–526.

Blevins, J. P. (2016). The minimal sign. In A. Hippisley & G. Stump (Eds.), *The Cambridge Handbook of Morphology* (pp. 50–69). Cambridge University Press. https://doi.org/10.1017/9781139814720.003

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, *32*, 45–86.

Boersma, P. (1998). *Functional Phonology* (Doctoral dissertation). University of Amsterdam.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *arXiv*. https://arxiv.org/abs/1607.04606

Bonami, O., & Stump, G. T. (2016). Paradigm function morphology. In A. Hippisley & G. T. Stump (Eds.), *Cambridge Handbook of Morphology* (pp. 449–481). Cambridge University Press.

Borg, A., & Azzopardi-Alexander, M. (1997). *Maltese*. Routledge.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Camilleri, J. J. (2013). *A Computational Grammar and Lexicon for Maltese* (Master Thesis). University of Gothenburg. Gothenburg, Sweden.

Chuang, Y.-Y., & Baayen, R. H. (2021). Discriminative learning and the lexicon: NDL and LDL. *Oxford Research Encyclopedia of Linguistics* (in press). Oxford University Press.

Chuang, Y.-Y., Kang, M., Luo, X., & Baayen, R. H. (2021). Vector space morphology with linear discriminative learning. *arXiv*. https://arxiv.org/abs/2107.03950

Chuang, Y.-Y., Lõo, K., Blevins, J. P., & Baayen, R. H. (2020). Estonian Case Inflection Made Simple. A Case Study in Word and Paradigm Morphology with Linear Discriminative Learning. In Körtvélyessy, Livia and Štekauer, Pavol (Ed.), *Complex Words: Advances in Morphology* (pp. 119–141). Cambridge University Press.

Chuang, Y.-Y., Vollmer, M.-L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., & Baayen, R. H. (2021). The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using linear discriminative learning. *Behavior Research Methods*, *53*(3), 945–976. https://doi.org/10.3758/s13428-020-01356-w

Corbin, D. (1983). Le monde étrange des dictionnaires (4): La créativité lexicale, le lexicographe et le linguiste. *Lexique*, *2*, 43–68.

Creemers, A., Davies, A. G., Wilder, R. J., Tamminga, M., & Embick, D. (2020). Opacity, transparency, and morphological priming: A study of prefixed verbs in dutch. *Journal of Memory and Language*, *110*.

Daelemans, W., & Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge University Press.

Daelemans, W., Zavrel, J., van der Sloot, K., & van den Bosch, A. (2001). TiMBL: Tilburg Memory-Based Learner-version 4.1-Reference Guide.

Daelemans, W., Zavrel, J., Van Der Sloot, K., & Van den Bosch, A. (2004). *Timbl: Tilburg memory-based learner* (tech. rep.). Tilburg University.

Dawdy-Hesterberg, L. G., & Pierrehumbert, J. B. (2014). Learnability and generalisation of Arabic broken plural nouns. *Language, Cognition and Neuroscience*, *29*(10), 1268–1282. https://doi.org/10.1080/23273798.2014.899377

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Denistia, K., & Baayen, R. H. (2021). The morphology of indonesian: Data and quantitative modeling. In C. Shei (Ed.), *The Routledge Handbook of Asian Linguistics*. Routledge.

Drake, S. (2018). *L1 biases in learning root-and-pattern morphology* (Doctoral dissertation). University of Arizona.

Ernestus, M., & Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*, 5–38.

Farrugia, A., & Rosner, M. (2008). Maltimorph, a computational analysis of the maltese broken plural. *Proceedings of Workshop on ICT 2008*.

Gaskell, M. G., & Marslen-Wilson, W. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*, 613–656. https://doi.org/10.1080/016909697386646

Gatt, A., & Čéplö, S. (2013). Digital corpora and other electronic resources for Maltese. *Proceedings of the International Conference on Corpus Linguistics*.

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

Heitmeier, M., & Baayen, R. H. (2021). Simulating phonological and semantic impairment of English tense inflection with Linear Discriminative Learning. *The Mental Lexicon*, *15*, 385–421.

Heitmeier, M., Chuang, Y.-Y., & Baayen, R. H. (2021). Modeling morphology with linear discriminative learning: Considerations and design choices.

Hoberman, R. D. (2007). Maltese morphology. In A. S. Kaye (Ed.), *Morphologies of asia and africa* (pp. 257–281). Eisenbrauns.

Ivens, S. H., & Koslin, B. L. (1991). *Demands for reading literacy require new accountability methods*. Touchstone Applied Science Associates.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). Fasttext.zip: Compressing text classification models. *arXiv*. https://arxiv.org/abs/1612.03651

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. *arXiv*. https://arxiv.org/abs/1607.01759

Kunter, G. (2017). Coquery: A free corpus tool. version 0.10.0.. https://www.coquery.org

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

Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Libben, G., Goral, M., & Baayen, R. H. (2018). What does constituent priming mean in the investigation of compound processing? *The Mental Lexicon*, *13*(2), 269–284.

Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Luo, X., Chuang, Y.-Y., & Baayen, R. H. (2021). Judiling: An implementation in Julia of Linear Discriminative Learning algorithms for language modeling. https://megamindhenry.github.io/JudiLing.jl/stable/

Matthews, P. H., & Matthews, P. (1972). *Inflectional morphology: A theoretical study based on aspects of Latin verb conjugation* (Vol. 6). CUP Archive.

Mayer, T., Spagnol, M., & Schönhuber, F. (2013). Fixing the broken plural in Maltese. *Perspectives on Maltese Linguistics* (pp. 129–158). Akademie Verlag.

McCarthy, J. J. (1981). A Prosodic Theory of Non-concatenative Morphology. *Linguistic Inquiry*, *12*, 373–418.

McCoy, R. T., Frank, R., & Linzen, T. (2020). Does Syntax Need to Grow on Trees? Sources of Hierarchical Inductive Bias in Sequence-to-Sequence Networks. *Transactions of the Association for Computational Linguistics*, *8*, 125–140. https://doi.org/10.1162/tacl_a_00304

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 3111–3119.

Milin, P., Divjak, D., & Baayen, R. H. (2017). A learning perspective on individual differences in skilled reading: Exploring and exploiting orthographic and semantic discrimination cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1730.

Mirković, J., MacDonald, M. C., & Seidenberg, M. S. (2005). Where does gender come from? evidence from a complex inflectional system. *Language and cognitive processes*, *20*, 139–167.

Mitchell, J., & Lapata, M. (2008). Vector-based models of semantic composition. *ACL*, 236–244.

Nieder, J., Tomaschek, F., Cohrs, E., & van de Vijver, R. (2021). Modelling Maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2021.1977835

Nieder, J., van de Vijver, R., & Mitterer, H. (2021a). Knowledge of Maltese singular–plural mappings. *Morphology*, *31*, 147–170.

Nieder, J., van de Vijver, R., & Mitterer, H. (2021b). Priming Maltese plurals: Representation of sound and broken plurals in the mental lexicon. *The Mental Lexicon*, *16*(1), 69–97.

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Experimental Psychology: General*, *115*(1), 39–57.

Pater, J. (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in Phonological Acquisition* (pp. 219–244). Cambridge University Press.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. http://www.aclweb.org/anthology/D14-1162

Pham, H., & Baayen, R. H. (2015). Vietnamese compounds show an anti-frequency effect in visual lexical decision. *Language, Cognition, and Neuroscience*, *30*(9), 1077–1095. https://doi.org/DOI:10.1080/23273798.2015.1054844

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Prokasy (Ed.), *Classical conditioning ii: Current research and theory* (pp. 64–99). Appleton Century Crofts.

Rumelhart, D. E., & McClelland, J. L. (1986). On Learning the Past Tenses of English Verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models* (pp. 216–271). The MIT Press.

Schembri, T. (2012). *The broken plural in Maltese—A description.* Brockmeyer.

Shafaei-Bajestan, E., Moradipour-Tari, M., Uhrig, P., & Baayen, R. H. (2021). LDL-AURIS: A computational model, grounded in error-driven learning, for the comprehension of single spoken words. *Language, Cognition and Neuroscience.* https://doi.org/https://www.tandfonline.com/doi/full/10.1080/23273798.2021.1954207

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford University Press.

Skousen, R. (1989). *Analogical modeling of language.* Kluwer.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic inquiry*, *27*(4), 720–731.

Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure* (Vol. 93). Cambridge University Press.

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2021). Phonetic effects of morphology and context: Modeling the duration of word-final

S in English with naïve discriminative learning. *Journal of Linguistics*, *57*(1), 123–161. https://doi.org/10.1017/S0022226719000203

Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language & Linguistic Theory*, *23*(1), 169–218. http://www.jstor.org/stable/4048019

van de Vijver, R., Uwambayinema, E., & Chuang, Y.-Y. (2021). Comprehension and production of Kinyarwanda verbs in the Discriminative Lexicon. *PsyArXiv*. https://doi.org/10.31234/osf.io/h63zy

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits.

Wood, S. N. (2017). *Generalized Additive Models*. Chapman & Hall/CRC.

Wurzel, W. U. (1970). *Studien zur deutschen Lautstruktur*. Akademie-Verlag.

Yang, W., Lu, W., & Zheng, V. (2017). A simple regularization-based algorithm for learning cross-domain word embeddings. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. https://doi.org/10.18653/v1/d17-1312

Zamuner, T., Kerkhoff, A., & Fikkert, P. (2011). Phonotactics and morphophonology in early child language: Evidence from Dutch. *Applied Psycholinguistics*, *33*(3), 481–499. https://doi.org/10.1017/S0142716411000440

Zwanenburg, W. (1983). *Productivité Morphologique et Emprunt. Études des dérivés déverbaux savants en français moderne*. Benjamins.