

Modelling Maltese noun plural classes without morphemes

Jessica Nieder, Fabian Tomaschek, Enum Cohrs & Ruben van de Vijver

To cite this article: Jessica Nieder, Fabian Tomaschek, Enum Cohrs & Ruben van de Vijver (2021): Modelling Maltese noun plural classes without morphemes, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2021.1977835](https://doi.org/10.1080/23273798.2021.1977835)

To link to this article: <https://doi.org/10.1080/23273798.2021.1977835>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 20 Sep 2021.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

Modelling Maltese noun plural classes without morphemes

Jessica Nieder ^a, Fabian Tomaschek ^b, Enum Cohrs^a and Ruben van de Vijver^a

^aDepartment of General Linguistics, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany; ^bDepartment of General Linguistics, Eberhard-Karls-Universität Tübingen, Tübingen, Germany

ABSTRACT

Word-based models of morphology propose that complex words are stored without reference to morphemes. One of the questions that arises is whether information about word forms alone is enough to determine a noun's number from its form. We take up this question by modelling the classification and production of the Maltese noun plural system, using models that do not assume morphemic representations. We use the Tilburg Memory-Based Learner, a computational implementation of exemplar theory and the Naive Discriminative Learner, an implementation of Word and Paradigm, for classification. Both models classify Maltese nouns well. In their current implementations, TiMBL and NDL cannot concatenate sequences of phones that result in word forms. We used two neural networks architectures (LSTM and GRU) to model the production of plurals. We conclude that the Maltese noun plural system can be modelled on the basis of whole words without morphemes, supporting word-based models of morphology.

ARTICLE HISTORY

Received 22 November 2020
Accepted 31 August 2021

KEYWORDS

Computational modelling;
Maltese plurals; word and
paradigm; CV template;
classification; production

1. Introduction

Is a complex word more like a construction made of Lego bricks – built from individual bricks which can be dissembled and reassembled again and again. Or is it more like a cake – baked from individual ingredients that give the cake its delicious taste, but which cannot be dissembled again into its ingredients?

The Lego-view of complex words is represented by morpheme-based models of morphology, such as *Item and Arrangement* or *Item and Process* models (Hockett, 1954), which assume that complex words consist of morphemes. A morpheme is usually defined as the smallest unit that combines sound and meaning (Bauer, 2016; Bloomfield, 1933; Haspelmath, 2020).¹ According to this view, complex words are built by arranging or combining morphemes into word forms (Blevins, 2006; O'Neill, 2014). This can be illustrated with the English past tense form *baked*, which is assumed to be the result of a combination of a root morpheme *bake* with a suffix *-ed*. Speakers can create new complex words by combining the morphemes stored in their memory by means of rules or processes.

Yet, this view breaks down in complex words in which individual morphemes can only be identified with great difficulty, if at all. This can be illustrated with the English past tense form *built* of the verb *build*. Which part of the

sounds in *built* represents the meaning PAST? Is it the voiceless final *t*, or just the phonological feature [-voice]? Neither of these solutions is general enough to cover other past tense forms, such as *constructed* or *assembled*, and would boil down to a complicated way of saying that the past tense of *build* is *built*. In short, morpheme-based theories have an attractive simplicity, which disappears upon scrutiny (Baayen et al., 2019).

The cake-based view of complex words is represented by word-based models of morphology (Blevins, 2006, 2016; Booij, 2010; O'Neill, 2014). Such models propose that complex words are stored without decomposition in the mental lexicon, and that words are the cognitive units in morphology. The difference among word-based models concerns the way in which word forms are related to one another. In realisational models this is achieved by rules, which formalise the relation among forms (Booij, 2010, 2016; Stump, 2001, 2018). In *Word and Paradigm* (WP) models this is achieved by proportional analogy (Blevins, 2006, 2016; O'Neill, 2014). The assumption of *Word and Paradigm* models that whole words are the cognitive units, related to one another through analogy, is shared with usage-based theories (Bybee, 2003; Bybee & McClelland, 2005; Bybee & Beckner, 2010). In usage-based theories, language use impels learning. Words are associated

CONTACT Jessica Nieder  nieder@phil.hhu.de

 Supplemental data for this article can be accessed <https://doi.org/10.1080/23273798.2021.1977835>

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

simultaneously with meaning and grammatical functions, and any word that is new for the user is understood or produced on the basis of words that the user knows. In other words, generalisations over forms in usage-based models emerge directly from stored word forms in the lexicon. These generalisations can take the form of analogy (Bybee, 2003, p. 7) or realisational rules (Booij, 2010, 2016; Stump, 2018)

Another possibility is that the phonological properties of words discriminate between grammatical functions. This possibility is formulated in *Naive Discriminative Learning* (Baayen et al., 2011; Milin et al., 2016). We stress that these hypotheses are not mutually exclusive (Kapatsinski, 2018, p. 16).

Maltese, a Semitic language spoken in Malta, has many complex words in which the morpheme boundaries can only be identified with great difficulty, if at all. The plural of *fardal* “apron” is *fradal*. In other Semitic languages such plurals, which are called broken plurals, have been analysed by means of abstract CV templates. The plural morpheme is conceived of as a string of Cs and Vs, e.g. CCVVCVC, onto which the singular is mapped (Dawdy-Hesterberg & Pierrehumbert, 2014; McCarthy, 1979; Schembri, 2012).

In this paper, we address the question whether it is possible to model classification and production of the Maltese noun plural system without making reference to morphemes. We use Maltese as empirical basis for two reasons.

The first one is that Maltese has a noun plural system that is comparable to Arabic, the noun plural system of which has already been modelled (Dawdy-Hesterberg, 2014; Dawdy-Hesterberg & Pierrehumbert, 2014). Yet, Maltese differs in important details from Arabic and other Semitic languages. Maltese has a lexicon that is for the most part Semitic, but it also has a considerable Sicilian and English share (Ussishkin et al., 2015). Just as other Semitic languages Maltese has two types of plurals. There are a large number of concatenative and a large number of non-concatenative plural forms. Concatenative plurals are expressed by additional material in the plural as compared to the singular, as in the pair *omm* ~ *ommijiet* “mother”. There are 12 different concatenative inflectional plural classes (Nieder et al., 2020). This constitutes a richer system than is usually found in other Semitic languages, which only have two concatenative plural classes (Dawdy-Hesterberg & Pierrehumbert, 2014; McCarthy & Prince, 1990). Non-concatenative plurals are expressed by a change in prosody, and in some cases a change in vowel quality in comparison to the singular, as in the pair *fardal* [fardal] ~ *fradal* [fra:da] “apron”. Maltese has 11 different non-concatenative inflectional plural

classes (Nieder et al., 2020). Detailed information about the Maltese noun plural system is provided in Section 2.

The second reason to use Maltese concerns its orthography, since we use written Maltese to train our models. The orthography of other Semitic languages, such as Arabic, abstracts away from short vowels. Accordingly, written Arabic does not account for morphological information affecting vocalic information. By contrast, Maltese orthography is completely vocalised, and therefore presents a better approximation of the spoken reality in Semitic languages.

1.1. Theories of morpheme representation for Semitic nouns

A widely accepted definition of the morpheme, dating back to Baudouin de Courtenay (1972) according to Bauer (2016), is that it is the smallest phonological form associated with a meaning (for a critical discussion of the notion morpheme see Haspelmath, 2020). As far as non-concatenative broken plurals are concerned, it is not obvious how to isolate one part of the plural word form that carries the meaning of the lexeme, and another part that carries the meaning of the grammatical function “plural”. As mentioned above, in Maltese, the singular *fardal* “apron” has the plural *fradal*. It is unclear how to divide the plural *fradal* in a part that means “apron” and a part that means “plural”.

Nevertheless, there are proposals as to how to achieve just this, often based on data from Arabic. One influential proposal put forward by McCarthy (1979) is that broken plurals in Arabic are represented by a template. This template has the general form CVCVCVCVC, in which C represents a consonant and V represents a vowel. In order to express the plural, segmental material of the singular is mapped onto this template. For example, the segmental material of the singular *sultaan* is mapped onto all available C and V positions in this template, *salaatiin*, resulting in the general form CVCVCVCVC. A singular such as *nafs* “soul”, on the other hand, uses only the first 6 positions of this template and is realised as *nufuus*, resulting in the plural CV-template CVCVC.

However, this mapping onto a CV template does not explain why one singular takes one particular broken plural, and another singular takes another broken plural (Hammond, 1988). In order to address this problem, McCarthy and Prince (1990) define the template not as a string of consonants and vowels, but as a prosodic unit. In the case of broken plurals in Arabic this unit is an iambic foot. They propose that the leftmost foot of the singular is mapped onto an iambic foot, and the prosody of any material left over in the

singular is concatenated to the plural form. For example the leftmost foot in the singular *nafs* is *naf*, the consonants of which are mapped onto an iambic foot *nVfVV*; the vowels are then changed by a special rule to form *nufuu*, after which the final /s/ of the singular is attached to the plural to arrive at the plural *nufuus*. This rule functions to both classify a noun as belonging to a class of plurals and as an instruction for producing a broken plural.

McCarthy and Prince's 1990 theory has been extended to Maltese by Schembri (2012). Schembri's 2012 approach describes the Maltese broken plurals well. The first syllable *far* of the Maltese singular *fardal* can be mapped onto a foot *fra*: with subsequent addition of the final syllable *dal* to arrive at the plural *fradal*. Yet, it does not seem possible to classify the whole Maltese noun plural system on the basis of her representations (even setting aside the issue that her account is not intended to classify concatenative sound plurals). For example, the singulars *birra* "beer" and *bir* "well" have the same first syllable, but different broken plurals: *birra* has the plural form *birer*, and *bir* has the plural form *bjar*. Even though the prosodic account of the shape of the plural morpheme cannot be extended to Maltese nouns, this does not mean that the CV template is equally unsuited for modelling the Maltese noun plural system. The phonological form of the singular may contain information that compels the choice of one particular plural CV template for one singular, and another plural CV template for another singular.

In fact, this is what the work of Dawdy-Hesterberg and Pierrehumbert (2014), who modelled the Arabic noun plural system computationally, suggest. They report support for the CV template in their computational modelling of the Arabic noun plural system. They used the Generalized Context Model (GCM, a model described in Nosofsky (1986)) to predict the plural class for a given Arabic singular. In order to do that, they compiled a corpus of 1945 singular plural pairs: 1384 Arabic sound singular-plural pairs (there are two sound plural suffixes [-uun] and [-aat] in Arabic) and 561 broken singular-plural pairs. The pairs are separated into different groups, so-called *gangs*, based on shared CV templates. Each gang contained singulars with the same CV template which have plurals that also have the same plural CV template; for example the largest broken plural gang had singulars of the form [CVCC] ~ [CVCVC] and included forms such as [dars] ~ [duruus] "lesson", [sijn] ~ [sujuwn] "prison" and [qasr] ~ [qusuwr] "castle".

The GCM predicts the plural pattern for a candidate based on the similarity of a newly encountered form and the gangs of given word forms, weighted on the basis of the number of word forms in each gang (see Dawdy-Hesterberg & Pierrehumbert, 2014; Nakisa et al.,

2001, for a detailed description of the model). Finally, the plural pattern of the gang of word forms having the highest similarity rating with the test word is selected by the GCM as the plural pattern for the test form. The results of their best performing model show an overall accuracy of 66%. The best performing model includes information about the CV-template, and is better than a model without CV-templates. Dawdy-Hesterberg and Pierrehumbert (2014) conclude from their results that the CV-template plays an important role in classifying nouns in the Arabic noun plural system.

Despite the evidence reported by Dawdy-Hesterberg and Pierrehumbert (2014) for CV templates as morphemes in Arabic, the controversy surrounding the construct morpheme provokes a sense of doubt (Baayen et al., 2019; Blevins, 2016; Stump, 2001) – a controversy that is not new either (Hockett, 1954, 1987; Lounsbury, 1953; Matthews, 1965). The morpheme is controversial since it is difficult, if not impossible, to delineate exactly what parts of a complex word correspond to what morpheme. In addition, there is a great deal of evidence that language users (Bybee & Beckner, 2010) do not make use of morphemes to comprehend, produce or process new complex words (Baayen et al., 2019; Chuang et al., 2020; Lõo, Järvikivi, & Baayen, 2018; Lõo, Järvikivi, Tomaschek et al., 2018).

In word-based theories, such as the one outlined in Blevins (2016), the controversy over morphemes is solved by declaring words, and not morphemes, the central units in cognition. In order for language users to comprehend and produce new words they use their knowledge of whole word forms (Bybee & Beckner, 2010), and do not parse a complex word into its constituent morphemes.

1.2. The present study

Do words contain enough information for the classification of their morphology? The hypothesis of memory-based models, such as the *Tilburg Memory-Based Learner* (Daelemans, 2005), is that words are grouped together because of their phonological similarity and the form of a new complex word forms is based on its similarity to stored words and the size of the similar group of stored words. Another possibility is that the phonological properties of words discriminate between grammatical functions. This possibility is formulated in *Naive Discriminative Learning* (Baayen et al., 2011; Milin et al., 2016). We stress that these hypotheses are not mutually exclusive (Kapatsinski, 2018, p. 16).

In this study, we want to investigate whether information about abstractive units like morphemes is

necessary for the classification of plural types of Maltese nouns. Our hypothesis is that this is not the case. Instead, we propose that classification of Maltese nouns is learned on the basis of unstructured sublexical information. We will use two computational models that test this hypothesis: Tilburg Memory-Based Learning (TiMBL) (Daelemans, 2005) and Naive Discriminative Learning (NDL) (Arppe et al., 2018; Baayen, 2011).

Since the analytical theories and the modelling of Dawdy-Hesterberg (2014) discussed above also implicitly account for the production of noun plurals, we also want to investigate whether it is possible to model the production of nouns without information about morphemes. To this aim we used an Encoder-Decoder neural network (McCoy et al., 2020).

Specifically, we will address the following questions:

- (1) Can we obtain generalisations about plural classes in Maltese on the basis of similarity?
- (2) Do morphemes play a role in learning the abstractions in Maltese, as Dawdy-Hesterberg and Pierrehumbert (2014) have argued for Arabic?
- (3) Can we produce Maltese sound and broken plural words without morphemes?

1.3. The models: TiMBL, NDL and encoder-decoder network

Data sets are computationally modelled in order to assess predictions about generalisations (Dawdy-Hesterberg & Pierrehumbert, 2014), and these models are often couched in language learning theories. In our case, we base our modelling approach on two important theories of language learning: Memory-based learning (Daelemans, 1995; Keuleers & Daelemans, 2007) and discriminative learning (Ellis, 2006; Ramscar & Yarlett, 2007; Ramscar et al., 2010). Both types of learning explain different aspects of language learning (Kapatsinski, 2018; Milin et al., 2016), and both do so without relying on the construct of the morpheme. By using computational models that can be regarded to be computational implementations of these theories, we want to assess whether it is possible to learn to classify or produce new complex words without recourse to morphemes, or whether the presence of morphemes is necessary for learning.

Memory-based learning assumes that language learning is driven through co-occurrence between cues and outcomes and as a result of which learners establish the probabilities of mappings between forms and function (Daelemans, 1995; Keuleers & Daelemans, 2007). In this view, learning is only based on positive evidence. However, there is also evidence that language learning

is discriminative and based on predictions and prediction error (Ramscar & Yarlett, 2007; Ramscar et al., 2010). The strength of a prediction is based on the informativity of cues that takes into account both the co-occurrence and the non-occurrence of cues and outcomes as a result of which cues compete for informativity about an outcome. A detailed overview of the similarities and differences between these approaches is presented by Kapatsinski (2018).

To model memory-based learning, we used TiMBL (Daelemans, 2005). It is an analogical model that is a variant of the *knn*-nearest neighbour model. The model assumes that exemplars are stored in memory, and that the similarity between new and stored forms decides how the newly encountered form is best classified. More specifically, learning in TiMBL is a more 'static' associative analogical process matching new input to the already stored exemplars in the lexicon in order to assess the similarity to the stored forms (Milin et al., 2020). The number of similar exemplars and the class of these exemplars decides the class of the new form.

TiMBL has been successfully used in natural language processing before. For example, Keuleers and Daelemans (2007) present a memory-based analysis of Dutch plurals. The plural in Dutch nouns are overwhelmingly expressed by a final *-ə* (written *-en*) or *s* (written *-s*). The choice between these plurals depends to a large extent on the phonology of the singular (Haas & Trommelen, 1993). For 19,351 nouns retrieved from CELEX (Baayen et al., 1996), the model performed three tasks. In a first task the model had to predict the plural of 5% of the nouns on the basis of being trained with the rest, in the two remaining tasks the model is expected to match the plural of a nonce word as produced by participants in two different experiments. The accuracy was measured by assessing whether the model assigned a probability ≤ 0.5 to the lexically attested form in task one, and, in the other tasks, to the majority of the answers given by the participants. The results of the modelling for words of 2 or more syllables were 83.9% for the first task, and 77.5% and 73.9% for the other two tasks. In short, it is possible to accurately assess the probability of a plural form given information about the phonology of the singular. A similar result was found by Milin et al. (2016) in modelling Serbian nouns. The phonological properties of the singular are to a large degree predictive of its morphological properties. A similar finding was reported by Vandekerckhove et al. (2008).

To model discriminative learning (Ellis, 2006; Ramscar & Yarlett, 2007; Ramscar et al., 2010), we used Naive Discriminative Learning (NDL) (Arppe et al., 2018; Baayen

et al., 2011), a two-layer input-output neural network. Discriminative learning, as implemented in NDL, has been successfully applied in natural language processing (Arnold et al., 2017; Ramscar & Yarlett, 2007; Ramscar et al., 2010) and learning of phonetic categories (Nixon, 2020; Nixon & Tomaschek, 2021; Olejarczuk et al., 2018). Moreover, discriminative learning as implemented in NDL has been very successful in modelling various aspects of cognitive processes related to morphologically complex words, such as acquisition (Hoppe et al., 2020; Ramscar et al., 2013), production (Baayen & Smolka, 2020; Ramscar & Yarlett, 2007; Tomaschek et al., 2019) or reading (Baayen et al., 2011; Milin et al., 2017).

To give an example, Olejarczuk et al. (2018) investigated whether exemplars of a phonetic category that occur infrequently affect the learning of this category. This would be the case, if negative evidence also plays a role in learning. In order to answer this question, Olejarczuk et al. (2018) created two differently skewed continua of a phonetic token, the syllable /ka/ with different pitch excursions. In one continuum there were more tokens with a small pitch excursion than with a large one, and in another group there were more tokens with a large pitch excursion than with a small one. The participants were exposed to tokens and were told that these are different pronunciation of the same word from a fictitious tone language. Two groups of participants, one trained with one skewed distribution, and another one trained with the other skewed distribution, were asked to rate how well new tokens fitted with the category heard in the training phase. The results showed that the ratings of the participants were skewed in the opposite direction of the skew in the distribution of the training tokens. This shows that infrequent tokens exert influence on learning and that infrequent tokens have a proportional greater influence on learning than frequent tokens. This finding is best explained by prediction and prediction error, in which learning is proportional to the amount of uncertainty of a cue-outcome association.

In their current implementations, TiMBL and NDL cannot be used to concatenate phone sequences that result in word forms. Therefore, we used an Encoder-Decoder neural network with a Long Short-Term Memory (LSTM) architecture and the Gated Recurrent Unit (GRU) architecture (McCoy et al., 2020) to model word production. We decided against a widely used, explicit computational model of morphophonological production, the Minimal Generalisation Learner (MGL) proposed by Albright and Hayes (2003). We did so because the MGL cannot model non-concatenative morphology for principled reasons. MGL compares forms of

different parts of a paradigm, for example singulars and plurals, and models the difference between these forms on the basis of a linear Sound-Pattern-of-English style rule (Chomsky & Halle, 1968). Such linear rules are known to not be able to capture prosodic changes (Goldsmith, 1979; Hayes, 1980; Leben, 1973).

LSTM and GRU architectures have been used in linguistic research before. McCoy et al. (2020) investigated which hypotheses about English question formation are entertained by a learner. The difference between a declarative sentence ("The zebra chuckles.") and a question ("Does the zebra chuckle?") is the position of the verb. McCoy et al. (2020) cite Chomsky (1980) to justify that there are two possible hypotheses compatible with the formation of questions: Move the main verb to the first position in the sentence, or move the first verb of the declarative sentence to the first position.² In order to test which of the hypotheses are likely to be entertained by language learners, McCoy et al. (2020) trained neural networks with LSTM and GRU architectures with declarative sentences and question sentences, each accompanied by a label stating whether the sentence is declarative or a question. The models were trained to create questions from declarative sentences. The results show that LSTM and GRU models are both able to create questions from declaratives with great accuracy. The models did so by learning that the word order in a declarative sentence in English differs from the word order in a question. As we will see below a class of Maltese plurals is characterised, among other things, by a different order of sounds in comparison to their singulars. The finding that LSTMs and GRUs can learn to produce outputs in which the information is ordered in a different way than in the input, suggests that these models could learn to produce plurals from given singulars.

The remainder of this paper is organised as follows. After having introduced the Maltese noun plural system, and our data set, we will turn our attention to our computational modelling. In order to address the question whether morphemes are necessary to classify the plural class of a noun correctly, we will model our data set in three computational models.

We first present the models using TiMBL, which is based on the assumption that learning is associative. This model is in principle agnostic as to the presence or absence of morphemes. We can therefore directly compare models with and without morphemes based on several inputs. In one model we used singulars as input to predict the different Maltese plural classes; in another model we used plurals as input to predict the plural classes; in a third model we used singulars and plurals as input to predict the plural classes and in the

last model we used singulars and plurals and their CV templates to predict the plural classes.

We then present the results of the modelling in NDL, a model based on the assumption that learning is error-driven. This model assesses how expected a plural class is given a certain input structure. We used the same types of input as we did for the TiMBL model, that is singular, plurals or a combination singular and plural forms. All input structures were coded as 2-phones or 3-phones, to investigate how much information of the input forms is needed to correctly predict the Maltese plural classes.

After this, we turn to modelling the production of plurals for any given singular. We do so in an Encoder–Decoder network.

In this model the singular and its plural were used as input to the decoder, and as output the model produced a plural form for a given singular form.

2. Maltese plurals

Maltese is a Semitic language spoken by approximately 500,000 people in Malta (Borg & Azzopardi-Alexander, 1997). The language developed from a spoken Maghrebi Arabic variety but due to extensive language contact the Maltese lexicon shows influences from Sicilian, Italian and English. The opposition of Semitic versus Non-Semitic morphological patterns is visible in the plural formation of the language.

The so-called *sound plurals* make up the majority of plural forms in Maltese and are expressed concatenatively by one of several different suffixes (Borg & Azzopardi-Alexander, 1997; Nieder et al., 2020; Schembri, 2012). For example, the singular *keik* “cake” has the plural form *keikijiet* in which the sound plural suffix *-ijiet* is added to the singular form.

Broken plurals, on the other hand, are expressed non-concatenatively by a different prosodic structure of the plural as compared to the singular. While consonants and their order are maintained, vowels may be changed during the process (Borg & Azzopardi-Alexander, 1997; Nieder et al., 2020; Schembri, 2012). For example the singular form *blakka* “brick” has the plural form *blakək*, in which the coda consonant [k] of the first syllable of the singular form is in the onset of the second syllable in the plural. In addition, another [ɔ] is inserted between the two consonants [k] while the word final vowel [a] from the singular is dropped. According to Schembri (2012), there are eleven different broken plural patterns (*broken A* – *broken K*).

Table 1 displays all Maltese sound plural suffixes and broken plural patterns that are identified by Nieder et al. (2020) and Schembri (2012):

Some Maltese nouns show several plural forms. For example, *kaxxa* “box” can be pluralised as *kaxxi* or as *kaxex*. In this example the types of plural classes differ. In the first case the concatenative sound plural form *-i* is used, in the second case the non-concatenative broken plural pattern *broken B* is shown. However, in other cases the possible multiple plural forms of a singular display a change in the vowels only: The singular *sunnara* “fishing hook”, for example, has the two possible broken plural patterns *snanar* or *sniener* (both *broken A*).

Faced with a great amount of variation within the Maltese plural system the question arises what information native speakers use to generalise to new word forms. Previous studies have shown that Maltese native speakers are aware of the split morphology in their native language, and use sound as well as broken plural pattern as a productive pluralisation strategy for nonce words they never heard before (Drake, 2018; Nieder et al., 2020), although their inflections seem to correlate with the frequency of the patterns and suffixes in their lexicon.

In this study, we will model Maltese plurals computationally. To ensure that individual plural classes had

Table 1. Maltese broken and sound plurals. Examples are taken from Schembri (2012) and Nieder et al. (2020). The type distribution in the rightmost column is based on the data set that is used for the present study. The pattern *broken H* identified by Schembri (2012) does not have any instances in our data set.

Singular	Plural	Gloss	Plural Type	Distribution
<i>karta</i>	<i>karti</i>	“paper”	sound, <i>-i</i>	1242 words
<i>omm</i>	<i>ommijiet</i>	“mother”	sound, <i>-ijiet</i>	430 words
<i>rixa</i>	<i>rixiet</i>	“feather”	sound, <i>-iet</i>	409 words
<i>giddieb</i>	<i>giddieba</i>	“liar”	sound, <i>-a</i>	109 words
<i>mehlus</i>	<i>mehlusin</i>	“freed”	sound, <i>-in</i>	91 words
<i>kuxin</i>	<i>kuxins</i>	“cushion”	sound, <i>-s</i>	53 words
<i>triq</i>	<i>triqat</i>	“street”	sound, <i>-at</i>	51 words
<i>sid</i>	<i>sidien</i>	“owner”	sound, <i>-ien</i>	5 words
<i>bahri</i>	<i>bahrin</i>	“sailor”	sound, <i>-n</i>	2 words
<i>hati</i>	<i>hatjin</i>	“guilty”	sound, <i>-jin</i>	1 words
<i>spalla</i>	<i>spallejn</i>	“shoulder”	sound, <i>-ejn/ajn</i>	1 word
<i>sieq</i>	<i>saqajn</i>	“foot”	sound, <i>-ejn/ajn</i>	1 word
<i>qiegh</i>	<i>qighan</i>	“bottom”	sound, <i>-an</i>	1 word
<i>fardal</i>	<i>fradal</i>	“aprons”	broken A, CCVVCVC	258 words
<i>birra</i>	<i>birer</i>	“beers”	broken B, (C)CVCVC	181 words
<i>kbir</i>	<i>kbar</i>	“big (pl.)”	broken C, CCVVC	148 words
<i>ftira</i>	<i>ftajjar</i>	“type of bread (pl.)”	broken D, CCVjjVC	64 words
<i>bitha</i>	<i>btiehi</i>	“yards”	broken E, CCWVCV	71 words
<i>sider</i>	<i>isdra</i>	“chests”	broken F, VCCCV	31 words
<i>marid</i>	<i>morda</i>	“sick persons”	broken G, CVCCV	14 words
<i>ghodda</i>	<i>ghodod</i>	“tools”	broken H, (gh)VVCVC	0 words
<i>elf</i>	<i>eluf</i>	“thousands”	broken I, VVCV	5 words
<i>gharef</i>	<i>ghorrief</i>	“wise men”	broken J, CVCCVC(V)	6 words
<i>ghama</i>	<i>ghomja</i>	“blind persons”	broken K, (gh)VCCV	1 word

Table 2. Example of singulars with multiple plurals from the data set.

	Example	Gloss	Proportion
2 plurals	<i>ċappetta</i> sg. – <i>ċappetti</i> pl. vs. <i>ċpiepet</i> pl.	'hinge of a door'	5.38% (160 of 2974)
3 plurals	<i>sejġ</i> sg. – <i>sjuf</i> pl. vs. <i>swejjef</i> pl. vs. <i>sejġijiet</i> pl.	'sword'	0.57% (17 of 2974)
4 plurals	<i>saff</i> sg. – <i>saffi</i> pl. vs. <i>saffijiet</i> pl. vs. <i>saġef</i> pl. vs. <i>sfuf</i> pl.	'layer'	0.07% (2 of 2974)

enough data points during our classification experiments, we simplified the classification of the Maltese plural system illustrated above in the following way.

On the basis of the frequency information provided by Nieder et al. (2020) and our own data set we focused on the three most frequent plurals of each class: In the case of sound plurals, the three most frequent suffixes in the noun list compiled by Nieder et al. (2020) and available in our data set are *-i*, *-iet* and *-ijiet* (see the first three rows of the upper part of Table 1). While the suffix *-i* has a Romance origin and is used with Romance loanwords only, the suffixes *-ijiet* and *-iet* can be used with both, Semitic and Non-Semitic, nouns.

For the broken plurals, the three most frequent patterns in our data set and in the noun list compiled by Nieder et al. (2020) are broken plural patterns A, B and C (CCVVCVC, (C)CVCVC and CCVVC, see the first three rows of the lower part of Table 1). All of these patterns can be used with both Semitic and Non-Semitic nouns (Borg & Azzopardi-Alexander, 1997).

2.1. Data set

For our models, we used an extended version of a data set of Maltese singular-plural pairs originally compiled by Nieder et al. (2020). Their data set was a combination of a set of broken plurals collected by Schembri (2012) and a set of singular-plural pairs from the MLRS Korpus Malti v. 2.0 and 3.0 (Gatt & Čeplo, 2013) that contains a total of ca. 250 million tokens from different text genres. To process the Korpus Malti v. 2.0 and 3.0, we downloaded a list of all nouns in the corpora. In case the list contained a singular without a corresponding plural, or the other way around, we added the missing forms by automatically matching the nouns with an online dictionary (Camilleri, 2013) data using the free corpus tool Coquery (Kunter, 2017).³ We extended the data set described in Nieder et al. (2020) using this method to create the data set used in this study. This method resulted in the final list of 3311 Maltese singular-plural pairs.

We provided a quasi-phonetic transcription to every singular-plural pairs such that every phone is represented as exactly one letter or symbol. We will henceforth refer to the symbols of this quasi-phonetic transcription as "phones".

In the following, sound plurals are coded by listing the suffixes based on the data presented in Table 1.

As can be seen, broken plurals are coded from pattern A to pattern K, a classification adopted from Schembri (2012).

Plurals that could not be assigned to an established Maltese morphological plural class were not taken into consideration for this study. In addition, we identified duplicated items and deleted them from the list. This removed 137 singular-plural pairs from the data set. The resulting data set contained 3174 Maltese singular-plural pairs (2395 sound nouns, 779 broken nouns).

Of all 2974 unique singulars in our data set, 179 (6% of 2974) show multiple plural forms (see Section 2 for a description). Table 2 displays three examples of singulars with multiple plural forms from the data set used for this study:

Within this set of 179 cases with multiple plural forms, 60 are several different broken plurals (*ċapella* sg. – *ċpapel* pl. – *ċpiepel* pl. "round stone"); 41 are several different sound plural forms (*pejġiep* sg. – *pejġiepin* pl. – *pejġiepa* pl. "one who smokes"); and 78 cross the broken-sound boundary (*torta* sg. – *torot* pl. – *torti* pl. "pie"). These cases, where one singular had multiple plurals, were also used in our modelling. In some cases, e.g. in the case of *ċappetta* sg. – *ċappetti* pl. vs. *ċpiepet* pl. (see the first row of Table 2), this might result in a model favouring and thus predicting the overall more frequent plural form than the less frequent one for a singular. In the given example above this would be *ċappetti* instead of the broken plural *ċpiepet*.

In our data set, some singulars map onto the same plural form. As a result, 90 plurals occur two times, e.g. *ħrafa* sg. – *ħrejġef* pl. "fable" and *ħarifa* sg. – *ħrejġef* pl. "autumn". In this example, the singulars show two different meanings. However, in some cases two singulars have the same plural form because they show an opposition of a masculine vs. a feminine singular with the same meaning, e.g. *tabib* sg. – *tobba* "doctor m." vs. *tabiba* sg. – *tobba* "doctor f.". Again, duplicated plurals like these were retained in the data set and might have an influence on the results of the models presented in the following sections.

On the basis of the type distribution described in Section 2, we divided the sound plurals into four categories: one for each of the three most frequent sound plurals suffixes (*-ijiet*, *-iet* and *-i*) and one category that contains all other, less frequent, sound plural forms (*sound (rest)*). We did the same for the broken plurals.

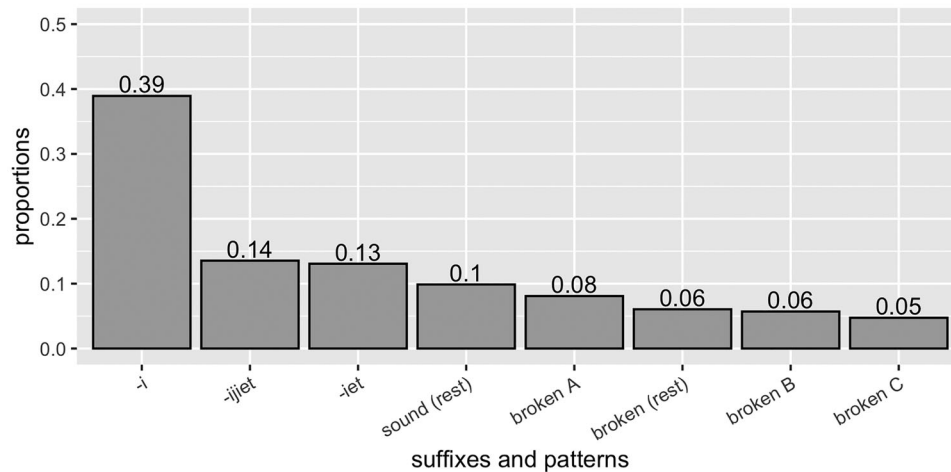


Figure 1. Proportion of sound plural suffixes and broken plural patterns within the data set.

We divided them in four categories: one for each of the most frequent broken plural patterns (*broken pattern A*, *B*, *C*) and one that contained all other broken plural forms (*broken (rest)*)⁴.

The proportions of the different sound plural suffixes and broken plural patterns used for our models is displayed in Figure 1:

The sound plural suffixes *-i*, *-ijiet* and *-iet* are the most frequent plurals in the data set, followed by the category *sound (rest)* that contains 314 words from 9 different sound plural categories. Following Schembri's 2012 classification, the three most frequent broken plural patterns in the data set are CCVVCVC, (C)CVCVC and CCVVC (in the plot and the following referred to as *broken A*, *broken B* and *broken C*). Infrequent broken plural patterns that had only a few instances of words were again combined to the group *broken (rest)*. In total this group contains 191 words from 7 different broken plural categories.

3. Modelling classification

As we have demonstrated above, the Maltese noun system is an instance of a complex pluralisation system. The aim of the present study is to test the hypothesis of Word and Paradigm theory. This theory assumes that words, not morphemes (or stems or exponents) are the relevant cognitive units (Blevins, 2016). We do so by training two computational models to classify the Maltese singular-plural system, and to train one computational model to produce Maltese plurals. In the upcoming sections we will first discuss the two classification models used in the present study – TiMBL and NDL – and their results. Subsequently, we present the results on production. Information on the structure of the models and the learning algorithms can be found

in the Appendix. The code and the data set that we used in our computational experiments can be downloaded from <https://osf.io/pyf7b/>.

3.1. Tilburg Memory-Based Learner

In order to classify a new token, TiMBL relies on stored representations and assesses the similarity of the new token to tokens it has stored in memory. The tokens are stored in memory as a fixed length feature-value vector and each vector is accompanied by information about its class label. All vectors are coerced to the same length, if necessary by padding the vector with extra 0s.

We tested four models: One in which we provided only singulars as input, one in which we provided only plurals as input (Kapatsinski, 2012; Köpcke & Wecker, 2017), one in which we provided singulars and plurals as input, and one in which we provided singulars and plurals with their CV templates as input (Dawdy-Hesterberg & Pierrehumbert, 2014). The output were the eight plural classes described in Section 2.1. These models were chosen for the following reasons: Singulars have been found to contain enough information to predict the morphological class of its paradigm (Keuleers & Daelemans, 2007; Milin et al., 2016). Plurals alone serve as a basis for prediction in product-oriented approaches (Bybee & Beckner, 2010; Kapatsinski, 2012). Singulars and plurals form a paradigm (in Maltese) and Word and Paradigm theory proposes that words and paradigms are cognitively relevant units (Blevins, 2006, 2016). Singulars plus plurals plus their CV templates is the model as proposed by Dawdy-Hesterberg and Pierrehumbert (2014).

We decided against a model using singulars and the CV template of their plurals, for the following reason. Since TiMBL uses similarity among words in its input to classify words, a model that has less information in its

Table 3. Example of one line of input for TiMBL for *kelb* “dog” for each of the four models. (a) Singulars only as input, (b) plurals as input, (c) singulars and plurals as input, and (d) singulars, plurals and their CV structure as input. The output is *broken pattern C*, in all four cases. The words are given in our quasi-phonetic transcription (see Appendix D).

input																		output	
(a)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	k	E	I	b	broken pattern C
(b)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	k	I	8	b	broken pattern C
(c)	0	0	0	0	0	0	0	0	0	k	E	I	b	k	I	8	b	broken pattern C	
(d)	k	E	I	b	C	V	C	C	k	I	8	b	C	C	V	V	C	broken pattern C	

input to assess similarity can never be better than a model that has more information. Since a model that has singulars and plurals has more information to base its assessment of similarity on, it would always be better, without giving us insight into the question whether extra abstract information provides a better classification.

We trained all models in the same way. The number of neighbours was set at 5, the similarity among exemplars was computed by means of the (Modified) Value Difference Metric (MDVM) (Daelemans, 2005). This similarity metric takes the relative similarity into account. For example, even though *pack* and *back* are different, *back* is more similar to *pack* than to *sack*. The MDVM would make *pack* and *back* more similar to each other than *sack*.

Examples of four feature-vectors, one for each model we have tested, and their class label are given in Table 3. The mathematics underlying TiMBL are explained in Appendix A.

3.2. Results TiMBL

We present the results of the models after 10 fold cross-validation. We created the files for cross-validation by randomising the order of the pairs and then splitting the file 10 times, in 10 different test and training files. Each training file contains 90% of the data set and the test files contain the remaining 10%. As there are 8 types of plurals to be classified, a classifier that would choose a random plural would achieve an accuracy of 12.5%.

3.2.1. Singulars

We will first discuss the results of modelling singulars alone and how well TiMBL classifies them according to

the eight categories described above. The mean accuracy of the classification of singulars after 10-fold cross validation was 63.4%. The confusion matrix of the best fold in this validation is given in Table 4.

The data show that for the class *broken A* there are 23 true positives (correct) and 5 false negatives (a plural of the class *broken A* was classified 3 times as *sound -iet*, once as *sound -i*, and once as *sound (rest)*), and 11 false positive classifications (4 plurals of class *sound -iet*, 5 plurals of class *sound -i* and 2 plurals of class *sound (rest)* were classified as *broken A*).

The results indicate that the plural class of a singular is difficult to assess on the basis of the information contained in singulars alone.

3.2.2. Plurals

We now turn to the results of classifying plurals as input. The mean accuracy of the classification of plurals after 10-fold cross validation was 95.5%. The confusion matrix is given in Table 5.

The results of this classification are excellent. It is clear that the information about plurals provides TiMBL with enough power to classify new forms with a high accuracy in terms of their plural class. In a next step, we will see what information is added by using plural forms and singular forms together.

3.2.3. Singulars and plurals

We now turn to the results of classifying singulars and plurals as input. The mean accuracy of the classification of singulars and plurals after 10-fold cross validation was 97%. The confusion matrix is given in Table 6.

Table 4. Confusion matrix of the best fold in 10 fold cross-validation of having TiMBL classify the plural class of nouns using **singulars as input**. Rows represent the input category, columns represent their classification. Accuracy of this fold was 68.4%. Its *F*-score was 68. The accuracy of the worst fold was 60.3%.

	broken A	broken B	broken C	broken	sound ijiet	sound iet	sound i	sound
broken A	23 (82%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (11%)	1 (4%)	1 (4%)
broken B	0 (0%)	9 (50%)	2 (11%)	0 (0%)	1 (6%)	4 (22%)	2 (11%)	0 (0%)
broken C	0 (0%)	1 (8%)	5 (39%)	1 (8%)	2 (15%)	2 (15%)	1 (8%)	1 (8%)
broken (rest)	0 (0%)	1 (6%)	1 (6%)	7 (44%)	2 (13%)	0 (0%)	4 (25%)	1 (6%)
sound ijiet	0 (0%)	1 (2%)	3 (6%)	1 (2%)	34 (64%)	3 (6%)	8 (15%)	3 (6%)
sound iet	4 (11%)	1 (3%)	2 (6%)	4 (11%)	2 (6%)	13 (37%)	7 (9%)	2 (6%)
sound i	5 (4%)	1 (1%)	0 (0%)	0 (0%)	5 (4%)	2 (2%)	108 (87%)	3 (2%)
sound (rest)	2 (6%)	0 (0%)	0 (0%)	4 (12%)	3 (9%)	0 (0%)	4 (12%)	20 (61%)

Table 5. Confusion matrix of the best fold in 10 fold cross-validation of having TiMBL classify the plural class of nouns using **plurals as input**. Rows represent the input category, columns represent their classification. Accuracy of this fold was 97.8%. Its *F*-score was.98. The accuracy of the worst fold was 93.4%.

	broken A	broken B	broken C	broken	sound ijiet	sound iet	sound i	sound
broken A	29 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken B	1 (6%)	15 (94%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken C	0 (0%)	0 (0%)	13 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken (rest)	0 (0%)	0 (0%)	0 (0%)	16 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
sound ijiet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	37 (97%)	1 (3%)	0 (0%)	0 (0%)
sound iet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (4%)	45 (94%)	0 (0%)	1 (2%)
sound i	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	124 (99%)	0 (0%)
sound (rest)	1 (3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	34 (97%)

A *t*-test of the mean accuracies of the paradigm and the singular model shows that using singulars and plurals together increases the accuracy ($t(11.0) = -37.3$, $p=5.486e-13$; Bonferroni corrected $p = 2.19426e - 12$).

3.2.4. Singulars and plurals and the CV template of both

We now turn to the results of classifying singulars and plurals and their CV template as input. The mean accuracy of the classification of singulars and plurals and their CV templates after 10-fold cross validation was 96.5%. The confusion matrix is given in Table 7.

TiMBL's results are excellent. A *t*-test of the mean accuracies of the model with singulars and plurals as input and the model using singulars, plurals, and their CV templates as input shows that having singulars and plurals together with their CV templates does not improve the accuracy compared to having only singulars and plurals as input ($t(17.7) = 1$, $p=0.28$; Bonferroni corrected $p=1$). It is clear that the CV templates do not provide any additional information to the classification problem.

3.2.5. Conclusion TiMBL modelling

Our modelling shows that the models with and without an abstract CV template perform equally well. Occam's razor tells us therefore that there is no reason to assume a CV template for nouns in Maltese. These results are in line with the predictions of the Word and Paradigm model of morphology (Baayen et al., 2019; Baayen & Smolka, 2020; Blevins, 2016).

In this respect, our results differ from the results of Dawdy-Hesterberg and Pierrehumbert (2014), who conclude that the CV template is an important factor in learning the noun plural system of Arabic. We propose an explanation in terms of the different writing systems of Arabic and Maltese. The role of the CV template in Arabic could be a result of the pointed Arabic text Dawdy-Hesterberg and Pierrehumbert (2014) used to create their data set. Short vowels are not represented in the pointed Arabic spelling (which is the standard way of spelling Arabic). This spelling is used to create the CV template, for example, the CV template of the word *dars* 'lesson' is CCC, and its segmental representation is *drs*. This increases the number of members in a gang, which then affords the model greater certainty in establishing the correct classification. In contrast, in Maltese short and long vowels are both written.

Furthermore, we show that Maltese is best modelled by taking the entire paradigm of a noun, its singular and its plural, into account (see Tables 7 and 6 and the results for all models in Figure 2). Since Maltese nouns are not marked for grammatical case, singular-plural pairs provided us with the entire paradigms of the nouns in our data set. If Maltese speakers have to find a plural form they do not know for a singular they do know, they are unlikely to rely only on information about similar sounding singulars. Rather they are likely to take into account similar singulars and their plurals.

An important point is that the singulars of Maltese alone provide less information to reliably classify their plurals. This makes the results we obtained for Maltese different

Table 6. Confusion matrix of the best fold in 10 fold cross-validation of having TiMBL classify the plural class of nouns using **singular-plural pairs as input**. Rows represent the input category, columns represent their classification. Accuracy of this fold was 98.1%, its *F*-score was.98. The accuracy of the worst fold was 95%.

	broken A	broken B	broken C	broken	sound ijiet	sound iet	sound i	sound
broken A	29 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken B	1 (6%)	15 (94%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken C	0 (0%)	0 (0%)	14 (93%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)	0 (0%)
broken (rest)	0 (0%)	0 (0%)	0 (0%)	16 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
sound ijiet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	38 (100%)	0 (0%)	0 (0%)	0 (0%)
sound iet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (4%)	45 (94%)	0 (0%)	1 (2%)
sound i	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	124 (99%)	0 (0%)
sound (rest)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (3%)	34 (97%)

Table 7. Confusion matrix of the best fold in 10 fold cross-validation of having TiMBL classify the plural class of nouns using **singular-plural pairs and their CV templates as input**. Rows represent the input category, columns represent their classification. Accuracy of this fold was 98%, its *F*-score was .98. The accuracy of the worst fold was 95%.

	broken A	broken B	broken C	broken	sound ijiet	sound iet	sound i	sound
broken A	17 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken B	0 (0%)	19 (95%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (5%)
broken C	0 (0%)	0 (0%)	14 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken (rest)	0 (0%)	0 (0%)	0 (0%)	25 (96%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)
sound ijiet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	37 (100%)	0 (0%)	0 (0%)	0 (0%)
sound iet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (5%)	42 (96%)	0 (0%)	0 (0%)
sound i	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	113 (99%)	1 (1%)
sound (rest)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	23 (96%)

from the results of Keuleers and Daelemans (2007) for Dutch, and the results of Milin et al. (2016) for Serbian. This is obviously connected to the different types of plurals found in Maltese, and in Semitic in general.

In the next section, we will discuss how an error-driven model classifies the nouns of the Maltese noun plural system without recourse to morphemes.

3.3. Naive discriminative learning (NDL)

Modelling in NDL proceeds as follows. The Maltese data set is divided into cues (the word forms) and outcomes (the plural types). The word forms for the cues are created by dividing the words in *n*-phones, which are considered to be sublexical units, in our case either 2-phones or 3-phones (see Table 8). For each cue the network establishes how strongly they predict an outcome. Take the words *barma* and *ballun* as an example, provided in Table 8 share two 2-phone cues: #b and b6. These cues then are both associated with the outcomes of *barma* and *ballun*, which are the sound plural class *sound -iet* and the broken plural class *broken pattern A* respectively. Thus, these two cues compete for how informative they are about the two sound classes. This also happens to all other cues in the words,

which are associated with other outcomes. The mathematics of the NDL model is explained in Appendix B.

Table 8 shows an example of a training data set for four existing Maltese singular nouns. The column “2-phone cues” contains the singular words *esperiment*, *barma*, *ballun* and *kelb* coded as 2-phone strings, the hash mark illustrates the word boundaries. The column “Outcomes” contains the corresponding plural types of the words specified as their corresponding broken plural pattern or their corresponding sound plural suffix. The singular *esperiment* has the sound plural form *esperiment* (sound -i), the singular noun *barma* has the sound plural *barmiet* (sound -iet). The noun *ballun* shows the broken plural form *blalen* (broken

Table 8. Example of NDL training data for the baseline model using singulars only as cues. The cues are given in our quasi-phonetic transcription (see Appendix B).

Singulars	2-phone cues	Outcomes
<i>esperiment</i> “experiment”	#E, Es, sp, pE, Er, rI, Im, mE, En, nt, t#	sound -i
<i>barma</i> “a twist”	#b, b6, 6r, rm, m6, 6#	sound -iet
<i>ballun</i> “ball”	#b, b6, 6l, ll, lu, un, n#	broken pattern A
<i>kelb</i> “dog”	#k, kE, El, lb, b#	broken pattern C

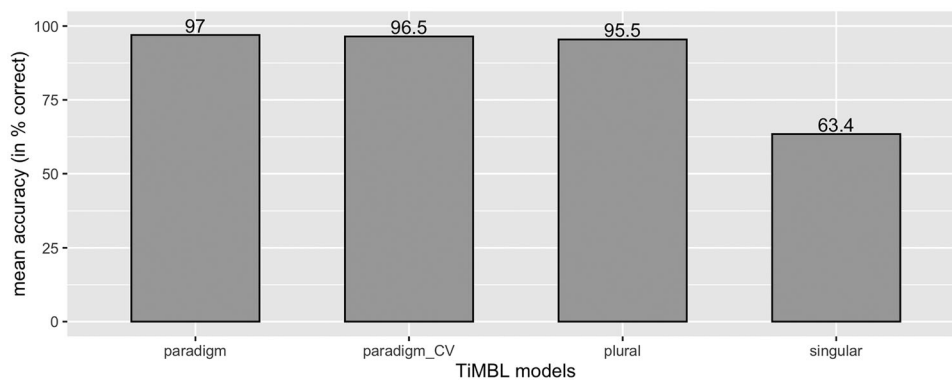


Figure 2. Mean accuracies of all TiMBL models. *Paradigm* is the model with singulars and plurals as input, *paradigm_cv* is the model with singulars and plurals and their CV template as input, *plural* is the model with plurals as input, and *singular* is the model with singulars as input.

Table 9. Confusion matrix of the best fold in 10 fold cross-validation of having the NDL model predict the plural class of a noun with **2-phones of singulars as cues**. Rows represent the input category, columns represent their classification. Accuracy of the best fold was 62.46%, accuracy of the worst fold was 56.47%.

	broken A	broken B	broken C	broken	sound iet	sound ijiet	sound i	sound
broken A	15 (52%)	1 (4%)	0 (0%)	1 (4%)	3 (10%)	1 (4%)	5 (17%)	3 (10%)
broken B	0 (0%)	4 (24%)	0 (0%)	0 (0%)	5 (29%)	2 (12%)	6 (35%)	0 (0%)
broken C	1 (7%)	1 (7%)	6 (40%)	2 (13%)	0 (0%)	1 (7%)	3 (20%)	1 (7%)
broken (rest)	0 (0%)	2 (9%)	2 (9%)	2 (9%)	5 (22%)	2 (9%)	4 (17%)	6 (26%)
sound iet	3 (8%)	2 (5%)	2 (5%)	2 (5%)	16 (42%)	2 (5%)	8 (21%)	3 (8%)
sound ijiet	0 (0%)	1 (2%)	1 (2%)	0 (0%)	3 (6%)	33 (67%)	9 (18%)	2 (4%)
sound i	3 (3%)	2 (2%)	0 (0%)	0 (0%)	2 (2%)	4 (4%)	103 (89%)	2 (2%)
sound (rest)	1 (3%)	0 (0%)	0 (0%)	1 (3%)	2 (7%)	2 (7%)	5 (17%)	19 (63%)

pattern A), and the singular *kelb* is pluralised as *klieb* (broken pattern C).

We created three different versions of the NDL model. In a first version, we used n-phone cues on the basis of singular words to predict plural classes as outcomes. In the second model version we used n-phone cues from plurals to predict the grammatical classes as outcomes (Kapatsinski, 2012; Köpcke & Wecker, 2017). In a third model we used n-phone cues from both singulars and their corresponding plurals to predict the outcomes.

To investigate how informative cues are depending on the size of n-phones, we created two kinds of cue sets for each version: One in which 2-phones were used as cues and another one in which 3-phones were used as cues. We used the same data set as we used for the TiMBL models (see Section 2.1).

Association weights between cues and outcomes were computed using the Danks Equilibrium Equations (Danks, 2003) implemented in the NDL package (Arppe et al., 2018) in R (R Core Team, 2020).

3.3.1. Results NDL

For the predictions with NDL we randomly divided the data set into a training and a test data set. The training data set contained 90% of the corpus data while the test data set contained the remaining 10%. The results of the models are given after performing a 10-fold cross-validation. In the following sections we present the results of the best fold for each version of the NDL model and

for each cue set (accuracies of the best and the worst fold are given in the caption of each table). The mean accuracy of each model is given in the text introducing the model. As there are 8 types of plurals to be classified, a classifier that would choose a random plural would achieve an accuracy of 12.5%.

3.3.2. Singulars

The results of the first version of the NDL model in which 2-phones of singulars predict the plural classes are illustrated in Table 9. The mean accuracy of the model was 60.1%.

Overall, the predictions are above chance level. We find worse predictions for broken plurals than for sound plurals, as the model gives sound plural predictions for broken plural words. This can be seen in the top right quadrant of Table 9.

For sound plurals the model provides better predictions, with 89% for the suffix *sound i* being the most accurate one (103 of 116 cases), followed by *sound ijiet* with an accuracy of 67% (33 of 49 cases) and *sound (rest)* with an accuracy of 63% (19 of 30 cases) (see lower right quadrant of Table 9).

To test the possibility that 2-phones are not informative enough about plural outcomes, we ran a second version using singulars coded as 3-phones as cues. This model showed a mean accuracy of 42.1%. The results are presented in Table 10.

As it turns out, using 3-phones as cues produces a worse overall mean accuracy with 42.1% correct than 2-

Table 10. Confusion matrix of the best fold in 10 fold cross-validation of having the NDL model predicting plural class with **3-phones of singulars as cues**. Rows represent the input category, columns represent their classification. Accuracy of the best fold was 44.79%, accuracy of the worst fold was 38.49%.

	broken A	broken B	broken C	broken	sound iet	sound ijiet	sound i	sound
broken A	14 (48%)	2 (7%)	1 (4%)	1 (4%)	3 (10%)	1 (4%)	4 (14%)	3 (10%)
broken B	1 (5%)	5 (25%)	2 (10%)	2 (10%)	3 (15%)	1 (5%)	6 (30%)	0 (0%)
broken C	0 (0%)	0 (0%)	5 (33%)	1 (7%)	5 (33%)	1 (7%)	3 (20%)	0 (0%)
broken (rest)	2 (13%)	4 (25%)	2 (13%)	3 (19%)	1 (6%)	0 (0%)	1 (6%)	3 (19%)
sound iet	2 (6%)	2 (6%)	2 (6%)	3 (9%)	11 (33%)	0 (0%)	11 (33%)	2 (6%)
sound ijiet	1 (2%)	2 (4%)	4 (8%)	2 (4%)	2 (4%)	21 (42%)	16 (32%)	2 (4%)
sound i	8 (7%)	3 (3%)	5 (4%)	4 (3%)	8 (7%)	12 (10%)	74 (61%)	7 (6%)
sound (rest)	3 (9%)	3 (9%)	0 (0%)	7 (21%)	3 (9%)	4 (12%)	4 (12%)	9 (27%)

Table 11. Confusion matrix of the best fold in 10 fold cross-validation of having the NDL model predicting plural class with **2-phones of plurals as cues**. Rows represent the input category, columns represent their classification. Accuracy of the best fold was 90.85%, accuracy of the worst fold was 86.44%.

	broken A	broken B	broken C	broken	sound iet	sound ijiet	sound i	sound
broken A	15 (68%)	1 (5%)	4 (18%)	0 (0%)	1 (5%)	0 (0%)	0 (0%)	1 (5%)
broken B	2 (10%)	18 (86%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (5%)
broken C	1 (13%)	1 (13%)	4 (50%)	0 (0%)	1 (13%)	0 (0%)	0 (0%)	1 (13%)
broken (rest)	0 (0%)	0 (0%)	3 (17%)	11 (61%)	0 (0%)	0 (0%)	2 (11%)	2 (11%)
sound iet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	40 (89%)	5 (11%)	0 (0%)	0 (0%)
sound ijiet	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	43 (100%)	0 (0%)	0 (0%)
sound i	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	130 (100%)	0 (0%)
sound (rest)	0 (0%)	0 (0%)	0 (0%)	2 (7%)	1 (3%)	0 (0%)	0 (0%)	27 (90%)

phones as cues. In short, the plural class is difficult to discriminate on the basis of singulars alone, even though the classification of each plural class is above chance. NDL is not different from TiMBL in this respect.

In the next section, we present the results of a version in which cues from the plural forms are used to predict the 8 different classes before we finally present a set of models in which both the singular and plural forms were used to predict the plural classes as outcomes.

3.3.3. Plurals

In a second set of NDL models we used the plural forms in our data set coded as 2-phones or 3-phones cues to predict the 8 different plural classes. The mean accuracy of the model using 2-phones of plurals as cues was 88.7%.

Table 11 shows the result of the model using plurals coded as 2-phone cues. With an accuracy of 90.85% for the best fold after 10-fold cross validation, the model shows excellent predictions for the different plural classes. The best predictions are given for the sound plural types *sound ijiet* and *sound i* with an accuracy of 100% for each class (43 and 130 correct predictions respectively). These plural types are the two most frequent plural types in our data set (as well as the two most frequent sound plurals in the work of Nieder et al. (2020)).

In a second step, we tested plurals which were coded as 3-phone cues. The mean accuracy of this model was 79.9%. The results are presented in Table 12 below:

Again, the overall mean accuracy of the model with an accuracy of 79.9% correct turns out lower for 3-phone cues than for 2-phone cues. Within the best fold after 10 fold cross-validation, the best predictions are still given for the two sound plural types *sound ijiet* and *sound i* with an accuracy of 89% (33 of 37 cases) and 94% (117 of 124 cases) respectively. Using plurals only as 3-phone cues thus results in less accurate (but still excellent) classifications for the sound plurals compared to a model that uses 2-phones as cues.

For the broken plural patterns, prediction accuracy is again higher than for the models based on singulars.

3.3.4. Singulars and plurals

In this section, we describe a third NDL model. In this model we used a combination of singular nouns and their corresponding plural forms coded as either 2-phones or as 3-phones to predict the different plural patterns. Again, we started with a version of the NDL model with 2-phones as cues to predict the different Maltese plural forms. Table 13 displays the results of this version of the NDL model that had a mean accuracy of 80.7%.

With a mean accuracy of 80.7% correct, the model shows robust predictions for the different plural classes. In the best fold after 10 fold cross-validation the model is able to provide the best classifications for the broken plural pattern *broken B* (= (C)VVCVC) with an accuracy of 71% and for the sound plurals *sound i* and *sound ijiet* with an accuracy of 99% and 92% (132 of 133 cases and 34 of 37 cases).

Table 12. Confusion matrix of the best fold in 10 fold cross-validation of having the NDL model predicting plural class with **3-phones of plurals as cues**. Rows represent the input category, columns represent their classification. Accuracy of the best fold was 84.54%, accuracy of the worst fold was 71.92%.

	broken A	broken B	broken C	broken	sound iet	sound ijiet	sound i	sound
broken A	24 (73%)	6 (18%)	3 (9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken B	0 (0%)	12 (80%)	2 (13%)	0 (0%)	0 (0%)	0 (0%)	1 (7%)	0 (0%)
broken C	1 (8%)	0 (0%)	9 (69%)	3 (23%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
broken (rest)	0 (0%)	1 (6%)	1 (6%)	10 (63%)	0 (0%)	0 (0%)	0 (0%)	4 (25%)
sound iet	0 (0%)	0 (0%)	2 (4%)	2 (4%)	37 (80%)	5 (11%)	0 (0%)	0 (0%)
sound ijiet	0 (0%)	0 (0%)	0 (0%)	2 (5%)	2 (5%)	33 (89%)	0 (0%)	0 (0%)
sound i	0 (0%)	0 (0%)	1 (1%)	4 (3%)	0 (0%)	0 (0%)	117 (94%)	2 (2%)
sound (rest)	0 (0%)	0 (0%)	1 (3%)	3 (9%)	0 (0%)	0 (0%)	3 (9%)	26 (79%)

Table 13. Confusion matrix of the best fold in 10 fold cross-validation of having the NDL model predicting plural class with **2-phones of the singular-plural paradigm as cues**. Rows represent the input category, columns represent their classification. Accuracy of the best fold was 83.91%, accuracy of the worst fold was 76.34%.

	broken A	broken B	broken C	broken	sound iet	sound ijiet	sound i	sound
broken A	16 (62%)	1 (4%)	2 (8%)	1 (4%)	2 (8%)	0 (0%)	1 (4%)	3 (12%)
broken B	2 (12%)	12 (71%)	0 (0%)	1 (6%)	0 (0%)	0 (0%)	0 (0%)	2 (12%)
broken C	2 (11%)	2 (11%)	12 (63%)	1 (5%)	1 (5%)	0 (0%)	0 (0%)	1 (5%)
broken (rest)	0 (0%)	2 (11%)	2 (11%)	11 (58%)	1 (5%)	0 (0%)	1 (5%)	2 (11%)
sound iet	0 (0%)	0 (0%)	0 (0%)	1 (3%)	34 (92%)	2 (5%)	0 (0%)	0 (0%)
sound ijiet	0 (0%)	0 (0%)	0 (0%)	1 (3%)	2 (5%)	34 (92%)	0 (0%)	0 (0%)
sound i	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	132 (99%)	0 (0%)
sound (rest)	5 (17%)	0 (0%)	1 (4%)	5 (17%)	1 (4%)	0 (0%)	2 (7%)	15 (52%)

We also provided NDL with 3-phones from singulars and their plurals as cues for a second version of the paradigm-model. The mean accuracy of this model was 38.55%. The results are displayed in Table 14.

A model that uses the whole paradigm coded as 3-phones as cues is able to correctly predict Maltese plural nouns (remember that chance level is at 12.5 %), but does so with a lower overall mean accuracy of 38.55% than a model that uses 2-phones as cues.⁵

3.4. Conclusion NDL modelling

Predicting the three most frequent Maltese sound plural suffixes and broken plural patterns (and the *rest* classes) with NDL revealed different results for the different cue structures we tested: (a) singulars only, (b) plurals only and (c) the whole paradigm. All cues were coded as 2-phones or as 3-phones to reveal how much phonological information of the input cues is needed for the correct prediction of the 8 different plural classes.

Figure 3 shows the mean accuracy of each NDL model presented in the sections above.

Plural_2phon is the model that used 2-phones of plurals as cues, *plural_3phon* is the model that used 3-phones of plurals as cues, *paradigm_2phon* is the model that used 2-phones of singulars and plurals as cues, *paradigm_3phon* is the model that used 3-phones of singulars and plurals as cues, *singular_2phon* is the model that used 2-phones of singulars as cues and *singular_3phon* is the model that used 3-phones of singulars as cues.

While both versions using singulars as cues (see Tables 9 and 10) showed a lower overall prediction accuracy for the broken plural classes compared to the plural- or paradigm-models (see Tables 11–14), the prediction accuracy for some of the broken plural classes got better when the singular-model was provided with 3-phone cues. We found the same effect in the model that used plurals coded as 3-phone cues, indicating that the two plural classes belonging to two different morphological systems (concatenative and non-concatenative) are discriminated on the basis of different information. Simply put, longer phone chunks are more informative about broken plurals than about sound plurals.

Overall, the models using plurals as cues showed a high mean accuracy with 88.7% and 79.9% respectively (see Figure 3). Both of these models are able to provide excellent classifications of the different Maltese plural classes. This indicates that information about the plural form is necessary for correct predictions of the different plural types in Maltese. This provides support for product-oriented models, in which a complex word is not constructed from stored morphemes, but rather on the basis of other complex words, with similar properties (Bybee & Beckner, 2010; Kapatsinski, 2013, 2018). If presented with singular only, the error-driven learning mechanism has a harder time to provide more accurate predictions, especially for Maltese broken plurals.

In a final step, we tested a setup using the whole paradigm as cues (see Tables 13 and 14). The mean accuracy was 80.7% for the paradigm-model that used 2-

Table 14. Confusion matrix of the best fold in 10 fold cross-validation of having the NDL model predicting plural class with **3-phones of the singular-plural paradigm as cues**. Rows represent the input category, columns represent their classification. Accuracy of the best fold was 47%, accuracy of the worst fold was 32.18%.

	broken A	broken B	broken C	broken	sound iet	sound ijiet	sound i	sound
broken A	3 (18%)	4 (24%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (24%)	6 (35%)
broken B	0 (0%)	13 (59%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	7 (32%)	2 (9%)
broken C	0 (0%)	0 (0%)	8 (40%)	2 (10%)	0 (0%)	0 (0%)	3 (15%)	7 (35%)
broken (rest)	0 (0%)	2 (17%)	2 (17%)	1 (8%)	0 (0%)	0 (0%)	5 (42%)	2 (17%)
sound iet	1 (3%)	4 (12%)	0 (0%)	1 (3%)	15 (46%)	0 (0%)	6 (18%)	6 (18%)
sound ijiet	1 (2%)	9 (19%)	2 (4%)	0 (0%)	0 (0%)	10 (21%)	18 (38%)	7 (15%)
sound i	5 (4%)	7 (6%)	2 (2%)	2 (2%)	1 (1%)	0 (0%)	84 (66%)	27 (21%)
sound (rest)	4 (11%)	3 (8%)	0 (0%)	1 (3%)	0 (0%)	0 (0%)	15 (40%)	15 (40%)

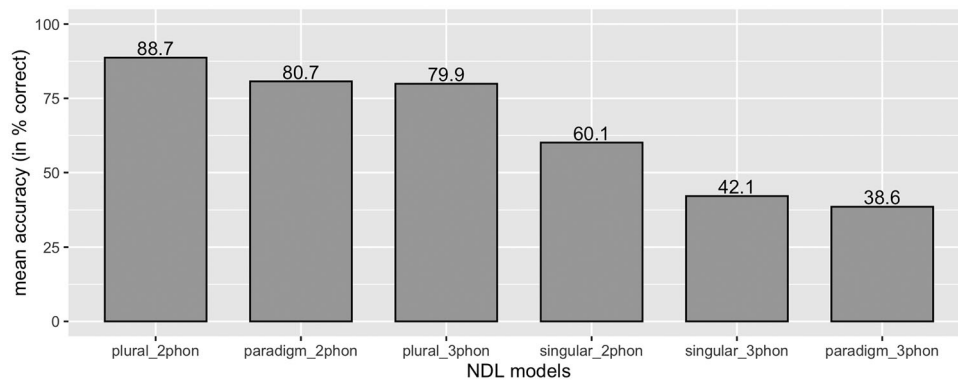


Figure 3. Mean accuracies of all NDL models.

phone cues, making it the second best model of this study. The model provided excellent predictions for all plural classes with errors mainly caused by confusing the different broken plural types or sound plural suffixes with each other in their respective plural group.

When provided with 3-phone cues, the predictions of the paradigm-model with a mean accuracy of 38.6% are less accurate for all plural classes, leading to confusion across categories and making it the worst model of all tested NDL models.

The last question that needs to be answered is why 2-phone cues result in overall better classification accuracy than 3-phone cues. This is surprising as the proportion of 3-phone cues that occur only once in a corpus is typically higher than the proportion of 2-phone cues occurring only once. As can be seen in Table 15, this is also the case in the present Maltese corpus for all noun classes. One would expect that cues that occur only once are actually very informative about their outcome. However, this is only the case if the outcome is known to the network. Since we cross-validated the models, cues that occur only once in the corpus are detrimental to classification. When the network is tested with a stimulus that contains a unique cue, it is not at all informative about its outcome. The remaining cues of the stimulus, i.e. those that occur more than once, cannot salvage the loss of information.

Another possibility for the difference in accuracy between 2-phones and 3-phones is that the cues differ in the amount of uncertainty about the outcomes. To assess this possibility we calculated the entropy in the

input data, as a measure of uncertainty of the cues in relation to the plural types. The entropy measures presented in Table 16 support this explanation. In all noun classes, entropy is higher when 3-phones are used as cues than when 2-phones are used as cues. From this follows that uncertainty is higher for 3-phones than for 2-phones, especially in the case when 3-phones with a low entropy are combined in the paradigm model. In addition, cues for broken nouns have a lower entropy than cues for sound nouns. This is most likely because there is a larger number of sound than of broken nouns.

In summary, the presented NDL models showed the best results when the plural forms of the words were taken into consideration in the cue structure. The most accurate predictions were given when plurals were coded as 2-phone cues, followed by the paradigm coded as 2-phone cues and plurals coded as 3-phone cues (88.7%, 80.7% and 79.9%, see Figure 3).

4. Modelling production of plurals with an Encoder–Decoder network

The TiMBL and NDL models are classifier models that allow us to predict which plural class a singular-plural pair belongs to. The models need to be provided with information about singulars and plurals to predict the plural classes of the nouns. However, the models cannot be used to concatenate phone sequences that result in word forms. To achieve this we implemented an encoder–decoder neural network.

Table 15. Percentage of cue types that appear only once in the corpus. Rows: cue structure. Columns: plural type.

	broken A	broken B	broken C	broken	sound ijiet	sound iet	sound i	sound
singular 2-phones	5.40	8.39	14.62	12.33	3.04	6.35	0.79	7.01
singular 3-phones	26.56	32.70	46.20	45.29	21.44	34.57	9.80	31.71
plural 2-phones	3.92	5.05	11.17	11.50	2.11	4.99	0.77	5.39
plural 3-phones	18.90	29.39	38.32	38.20	14.68	25.53	8.96	27.44
sg+pl 2-phones	3.92	5.56	12.48	7.61	2.16	3.95	0.68	4.71
sg+pl 3-phones	20.60	26.83	40.08	37.01	15.03	23.92	7.86	24.78

Table 16. Entropy in the cue structure (rows) for each plural type (columns).

	broken A	broken B	broken C	broken	sound ijiet	sound iet	sound i	sound
singular 2-phones	0.61	0.38	0.28	0.39	1.09	0.80	2.62	0.69
singular 3-phones	0.70	0.42	0.28	0.43	1.29	0.93	3.42	0.76
plural 2-phones	0.53	0.34	0.24	0.39	1.19	0.80	2.39	0.71
plural 3-phones	0.58	0.37	0.24	0.41	1.40	0.91	3.12	0.81
sgn+plr 2-phones	0.72	0.40	0.33	0.51	1.09	0.81	2.28	0.67
sgn+plr 3-phones	0.90	0.46	0.38	0.59	1.25	0.95	2.91	0.76

We implemented a model in which singulars are mapped onto plurals, and the network has to learn how this mapping best produces plurals. We used the same data set as we used for the TiMBL and NDL modelling (see Section 2.1), with the difference that the encoder–decoder networks were provided only with singulars and their corresponding plurals. Each singular, and each plural was represented as a vector. Each vector coded the presence and absence of the sounds of Maltese. If a sound is present in a word it is coded as 1 and if it is absent it is coded as 0.

We tested two different architectures that have been used for linguistic phenomena: Long short-term memory (LSTM) and gated recurrent units (GRU) (McCoy et al., 2020). The LSTM architecture consists of two networks, the encoder and the decoder, each with a hidden state. The encoder is fed the input, represented as a vector, and it updates its hidden state, also a vector, after each part of the input vector has been processed. The hidden state contains a representation of the information that has been processed. After the entire input has been processed the final state of the hidden state is fed into the decoder, another vector representation, which generates an output, again one slice of its vector at the time. The output is based on the hidden state of the decoder. A special property of LSTMs is their ability to ignore, or forget, some information it has deemed irrelevant. By doing so, an LSTM adds a computational step and this slows LSTMs down. GRU models lack this ability, but are otherwise identical to LSTMs. Each architecture was further tested with attention and without attention. Attention allows the decoder to access all hidden states of the encoder, in models without attention the decoder can only access the last hidden state of the encoder (McCoy et al., 2020). A description of these architectures can be found in Appendix C.

4.1. Results Encoder–Decoder modelling

We fed the model with Maltese singulars and their plural forms. For example *farfett* “butterfly sg.” was mapped onto its broken plural *friefet*. In this way the model has to learn to produce plurals for any given singular as output. Two examples of inputs, corresponding outputs and predicted outputs is given in Table 17.

In the first example in Table 17, there is an exact match between the corresponding output plural and the predicted plural. In the second example, there are two possible plurals, and there is no exact match between the corresponding output plurals and the predicted plural. In order to evaluate the results of the neural networks, we used several measures. In addition to the number of exact matches (in which a plural form exactly matches the predicted plural), we also calculated the overlap between the plural of the output (i.e. the amount of identical segments in the compared items) and the plural predicted by the model. For example, the overlap between *aabcd* and *aaacaa* is 3: *aa* and *c*. Overlap can then be used to calculate the precision ($\frac{\text{overlap}}{\text{length}(\text{output})}$) and recall ($\frac{\text{overlap}}{\text{length}(\text{referenceplural})}$) of the models. Both measures range between 0 and 1, with 1 representing optimal recall/precision. These are used to calculate the f1, which can be understood as a measure for the goodness of a model. It is calculated as follows: $\frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})}$. Furthermore we calculated the Levenshtein distance between the output plural and the predicted plural. The mappings from singular to plural were learned as 1-phone to 1-phone, 2-phone to 2-phone or 3-phone to 3-phone mappings.

The results of all models after 10-fold cross validation are presented in Table 18.

The GRU architecture with attention (see fourth row of Table 18) does worse than all of the other models. Not only are recall, precision and f1 score worse than the scores of the other models, also, there are fewer exact matches and a larger average Levenshtein distance compared to the GRU without attention and both LSTM models.

The GRU architecture without attention (see third row of Table 18) does better than the GRU architecture with attention, but not better than the models with an LSTM

Table 17. Example of an input, an output and a prediction of the LSTM architecture with attention.

input (singular)	farfett
output (plural)	friefet
predicted plural	friefet
input (singular)	shubija
output (plural)	shubiet
output (plural)	shubijiet
predicted plural	hsiebijiet

Table 18. Average measures, Levenshtein distance and exact matches for the neural network structures.

		Precision	Recall	F1-Score	Average Levenshtein Distance	Exact Matches
LSTM with attention	1-phone	0.89	0.89	0.88	1.43	48.22%
	2-phones	0.73	0.72	0.72		
	3-phones	0.66	0.65	0.65		
LSTM no attention	1-phone	0.89	0.88	0.88	1.7	41.18%
	2-phones	0.68	0.68	0.68		
	3-phones	0.59	0.58	0.58		
GRU no attention	1-phone	0.87	0.86	0.86	1.84	33.55%
	2-phones	0.67	0.65	0.66		
	3-phones	0.56	0.55	0.55		
GRU with attention	1-phone	0.68	0.72	0.67	5.73	16.77%
	2-phones	0.44	0.45	0.43		
	3-phones	0.33	0.33	0.32		

architecture. Even though precision, recall, and the f1 score of the GRU architecture with attention are comparable to both LSTM models, it has fewer exact matches and a larger Levenshtein distance.

The best models are the LSTM models illustrated in the first and second row of Table 18 and of these, the LSTM model with attention is the best one (first row). With 47.11% it has the highest exact matches and with 1.48 the lowest Levenshtein distance of all tested models.

Overall, though, the models are fairly successful in predicting plurals for given singulars, after having learned a number of singular-plural mappings. The LSTM model with attention is most successful, because the architecture employs a forget gate, which allows it to ignore irrelevant information, in addition to attention, which allows it to access all hidden states of the encoder (McCoy et al., 2020). The GRU model with attention is least successful, because the architecture lacks a forget gate. The model considers all information from all hidden states of the encoder, and these, apparently, contain too much noise.

In line with our findings in the TiMBL and NDL simulations that morphological information does not improve classifying plurals, the encoder-decoder nets were not provided with morphological information. Nevertheless the models did well in predicting several types of plurals, and the mistakes they made still look like Maltese plurals. For example the plurals predicted by the model given in Table 19, are possible plural forms, just not for these singulars. The predicted plural *tfalin* is a mix of a broken pattern, *tfal* is the broken plural form of *tifel/tifla* “boy/girl” which are both phonetically close to *tafal*, and a sound suffix, as found in the actual plural *bnedmin* from the singular *bniedem* “man”.

Interestingly, these results resemble incorrect plural forms Maltese native speakers would use in cases of uncertainty about low frequent plurals. The predicted plural *pinniet* would be a perfect example of a sound plural that ends in *-iet*. In their production experiment with nonce singulars and existing singulars, Nieder et al. (2020) found for the existing singulars that

especially when sound plural suffixes were used instead of the correct broken plural form, these suffixes are part of the high frequency forms that are available for sound pluralisation, that is either *-i*, *-ijiet* or *-iet*. In the light of Nieder et al.’s 2020 results, the incorrectly predicted plural form *pinniet* in Table 19 is not surprising but rather a perfect example of an error humans would run into as well.

The results may not seem very good in comparison with other deep learning results, for example McCoy et al.’s 2020. But a number of things have to be kept in mind: The data set is by necessity relatively small, and the number of different sounds in the data set is much smaller than the number of different words in the sentences used by (McCoy et al., 2020).

4.2. Conclusion Encoder-Decoder modelling

The question addressed by our modelling efforts was whether neural networks are able to learn to predict a plural for any given singular without any decomposition into morphemes. The models perform well in this respect, especially given the limited amount of data they were provided with.

Our modelling of production shows that there is much relevant information in the phonological forms that can be used to predict a plural from for an unknown singular form. It is also clear that this is not a plausible model for how children learn morphology. It is unlikely that they compare forms of a paradigm in order to arrive at predictions for novel forms (Ramscar & Yarlett, 2007).

Table 19. Examples of Maltese-like mistakes from the best performing neural network: LSTM with attention.

input (singular)	tafal
output (plural)	itfla
predicted plural	tfalin
input (singular)	pinna
output (plural)	pinen
predicted plural	pinniet

5. Discussion

In the present study, we investigated the following three questions:

- (1) Can we obtain generalisations about plural classes in Maltese on the basis of similarity? The results of our TiMBL modelling indicate that the answer to this question is yes.
- (2) Do morphemes play a role in learning the abstractions in Maltese, as Dawdy-Hesterberg and Pierrehumbert (2014) have argued for Arabic? The results of both the TiMBL modelling and the NDL modelling indicate that the answer is no.
- (3) Can we produce Maltese sound and broken plural words without morphemes? The RNN indicates that the answer is yes.

In the following, we will discuss the details of our study and our results. Our starting point was to investigate the role of morphemes and CV-templates in the classification of broken and sound noun classes of Semitic languages, whether plurals can be predicted on the basis of the phonology of their singulars, and the consequences of assuming or not assuming morphemes and CV-templates for morphological theory. This is relevant in the light of the debate as to whether morphology is morpheme-based (Bauer, 2016; Halle & Marantz, 1993) or word-based (Blevins, 2006, 2016; Booij, 2010).

In order to test to what cues are informative about noun classes (broken classes vs. sound classes) in Maltese, we ran several computational analyses with the Maltese noun plural system. We trained TiMBL (Daelemans, 2005) on different combinations of cues – segmental information, or segmental information in addition to an abstract morpheme in the form of a CV-template. We found that adding a CV-template to segmental information does not improve classification. A highly accurate classification can be accomplished using segmental information only. Our findings differ from the ones reported by Dawdy-Hesterberg and Pierrehumbert (2014) for Arabic noun plurals. They found that adding a CV template improved classification. We attributed this difference to two factors: the Arabic writing system does not represent short vowels, which increases the informativity of consonants, whereas in Maltese all vowel and consonants are written. The second difference is rooted in the Maltese lexicon which contains a great deal of non-Arabic words. These words do not obey the typical Arabic consonantal root pattern.

Using NDL (Arppe et al., 2018; Baayen et al., 2011), we further tested the informativity of differently sized

chunks (2-phone or 3-phone cues) of segmental information for noun classes. This analysis indicated that 2-phones are more informative about sound nouns and 3-phones are more informative about some broken nouns.

Moreover, the results of modelling with TiMBL and NDL show that while noun classes can be classified using singulars alone, providing the model with singulars and plurals offers a much better classification. This indicates that the words in the entire paradigm are informative about noun classes. This result is also reflected in the results of our encoder–decoder production model, in which the model learned to predict the form of the plural of a singular on the basis of its plural. The information in the phonology of singulars helps in producing plurals, but as the results leave room for improvement, it is likely that other information also plays a role – a probable contender is the semantics of the noun. These findings are in agreement with experimental work on Maltese nouns, which has provided support for word-based lexical processing (Nieder et al., 2020, 2021) and further strengthens a word-based formalism of morphological processing such as the Word and Paradigm model of morphology (Blevins, 2006, 2016).

Our findings underline an apparent dichotomy in the Maltese lexicon. For nouns, the present work and our experimental work (Nieder et al., 2020, 2021) has shown that they are best represented as whole words and that there is no need to assume the presence of morphemes. By contrast, experimental studies (Ussishkin et al., 2015) and computational work (Borg, 2015) on Maltese verbs has argued that representation of verbs includes their consonantal roots, which are assumed to be morphemes that encode the lexical meaning of lexemes (Ussishkin et al., 2015). This raises the question as to why there would be such a difference between nouns and verbs. We speculate that this is caused by the fact that a morphological family (Moscoso del Prado Martín et al., 2004) of Maltese verbs contains many word forms that share the consonantal root, whereas for nouns the family is smaller – nouns are not inflected for tense or person.

In verbs, the shared consonantal root results in a strong similarity among verb forms. From an associative learning perspective, the consonants then will become a strong force to attract perceived forms. From a discriminative learning perspective, the root becomes the most informative cue about their lexical meaning. From both perspectives, there is no need to assume that the roots are conceived of as morphemes. We leave this issue open for future research.

The present findings raise an important methodological point. To investigate morphological production,

studies use wug tests in which one form is used in order to make a participant produce another form (Berko-Gleason, 1958; van de Vijver & Baer-Henney, 2014). In addition to being a very meta-linguistic task, this method may leave participants stumped in case of languages such as Maltese. One form may not provide enough information to participants for them to reliably produce another form. This might be why participants are often so tongue-tied in this task or why they often simply repeat the form provided to them rather than inflecting them (Nieder et al., 2020; van de Vijver & Baer-Henney, 2014).

Furthermore, our results have implications for models that formalise morphological production from a source-oriented or realisational perspective (Booij, 2010; Bybee, 2003; Köpcke & Wecker, 2017), or a product-oriented perspective (Bybee & Beckner, 2010; Kapatsinski, 2018; Köpcke & Wecker, 2017). Source-oriented and realisational models rely on rules which formalise how one form in the paradigm can be changed into, or relates to, another form. For example, using the Minimal Generalisation Learner, Albright and Hayes (MGL, 2003) modelled morphological processes from a source-oriented perspective. The MGL compares two forms and the difference between them is stated as a rule. Provided with many pairs, the model comes up with many rules that can be used to derive one form from another one. As some of our best models were based on information about the plural alone, this indicates that cues that reflected the product-oriented perspective performed better than cues that reflected the source-oriented perspective. As the advantage for the product-oriented approaches emerged in our NDL models, it may indicate that our result is the consequence of the different learning theories that underlie NDL and TiMBL. More research is needed to explore these differences and the predictions that follow from them for product-oriented and source-oriented or realisational models.

Finally, it is evident that our models come with a clear shortcoming. Like other modelling approaches (Albright & Hayes, 2003; Dawdy-Hesterberg & Pierrehumbert, 2014), we only used phonological information but did not include semantic information. There is evidence that phonological information alone is not informative enough for the production of morphologically complex word forms (see Baayen et al., 2018, 2019, who modelled production from a discriminative learning perspective). This means that once information about semantics is included to the production model, its accuracy might improve. In other words, a more complete picture can only arise when semantic information is also taken into account. This would be a very different study, though, one which we leave for future research.

In conclusion, we have shown that classification and production of the Maltese noun plural system can be successfully modelled without recourse to morphemes, and that complex word are like tasty cakes, that we can, and should, enjoy as a whole.

Notes

1. for a critical discussion of the notion morpheme see Haspelmath (2020).
2. Another hypothesis would be that questions are not formed by rearranging the order of the words in a declarative sentence, but are stored as exemplars (Bod, 2006). Pursuing this hypothesis is a research programme in itself.
3. available at <http://coquery.org/>
4. We also tested models with all plural classes, but the results of these models were difficult to interpret, because of the sparsity of some plural classes.
5. Following the suggestion of an anonymous reviewer, we have also tested models with a combination of 2-phones and 3-phones as cues. However, these models achieved an accuracy that was the same as models that used 2-phones or 3-phones separately, or less.

Data availability statement

The data that support the findings of this study are openly available at <https://osf.io/pyf7b/>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 'Spoken Morphology', Projects VI 223/3-1, VI 223/3-2, BA 3080/3-1 and BA 3080/3-2), which we gratefully acknowledge. We thank the audience of the Colloquium of the research unit Spoken Morphology and the audience of the International Conference on Error-Driven Learning in Language 2021 for their feedback and comments.

ORCID

Jessica Nieder  <http://orcid.org/0000-0001-5977-8958>

Fabian Tomaschek  <http://orcid.org/0000-0003-2246-663X>

References

- Albright, A., & Hayes, B. (2003). Rules vs. Analogy in English past tenses: A computational/experimental study. *Language*, 90 (2), 119–161. [https://doi.org/10.1016/S0010-0277\(03\)00146-X](https://doi.org/10.1016/S0010-0277(03)00146-X)
- Arnold, D., Tomaschek, F., Lopez, F., Sering, K., Ramscar, M., & Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy

- by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLOS One*, 12 (4). <https://doi.org/10.1371/journal.pone.0174623>
- Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2018, September). *ndl: Naive discriminative learning*. Retrieved September 20, 2020, from <https://CRAN.R-project.org/package=ndl>
- Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Revista Brasileira de Linguística Aplicada*, 11(2), 295–328. <https://doi.org/10.1590/S1984-63982011000200003>
- Baayen, R. H., Chuang, Y. Y., & Blevins, J. P. (2018). Inflectional morphology with linear mappings. *The Mental Lexicon*, 13(2), 230–268. <https://doi.org/10.1075/ml>
- 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. <https://doi.org/10.1155/2019/4895891>
- Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438–481. <https://doi.org/10.1037/a0023851>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). The CELEX lexical database (cd-rom).
- Baayen, R. H., & Smolka, E. (2020). Modeling morphological priming in German with naive discriminative learning. *Frontiers in Communication*, 5(17). <https://doi.org/10.3389/fcomm.2020.00017>.
- Baudouin de Courtenay, J. (1972). An attempt at a theory of phonetic alternations. In E. Stankiewicz (Ed.), *A Baudouin de Courtenay Anthology: The beginnings of structural linguistics* (pp. 144–212). Indiana University Press.
- Bauer, L. (2016). Classical morphemics: Assumptions, extensions and alternatives. In A. Hippisley and G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 331–355). Cambridge University Press.
- Berko-Gleason, J. (1958). The child's learning of English morphology. *Word*, 14(2–3), 150–177. <https://doi.org/10.1080/00437956.1958.11659661>
- Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, 42(3), 531–573. <https://doi.org/10.1017/S0022226706004191>
- Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press.
- Bloomfield, L. (1933). *Language*. H. Holt and Company.
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, 23(3), 181–320. <https://doi.org/10.1515/TLR.2006.012>
- Booij, G. (2010). *Construction morphology*. Oxford University Press.
- Booij, G. (2016). Construction morphology. In A. Hippisley and G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 424–448). Cambridge University Press.
- Borg, A., & Azzopardi-Alexander, M. (1997). *Maltese*. Routledge.
- Borg, C. (2015). *Morphology in the Maltese language: A computational perspective* [Doctoral thesis]. University of Malta.
- Bybee, J. L. (2003). *Phonology and language use* (Vol. 94). Cambridge University Press.
- Bybee, J. L., & Beckner, C. (2010). Usage-based theory. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 827–855). Oxford University Press.
- Bybee, J. L., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22(2–4), 381–410. <https://doi.org/10.1515/tlir.2005.22.2-4.381>
- Camilleri, J. J. (2013). *A computational grammar and lexicon for Maltese* [Master thesis]. University of Gothenburg.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3(1), 1–15. <https://doi.org/10.1017/S0140525X00001515>
- Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. (2015). Attention-based models for speech recognition. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15)*, Volume 1, MIT Press, Cambridge.
- Chuang, Y. Y., Löö, K., Blevins, J. P., & Baayen, R. (2020). Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning. In L. Körtvélyessy and P. štekauer (Eds.), *Complex words: Advances in morphology* (pp. 119–114). Cambridge University Press.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
- Cost, S., & Salzberg, S. (1993). A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning*, 10(1), 57–78. <https://doi.org/10.1023/A%3A1022664626993>
- Daelemans, W. (1995). Memory-based lexical acquisition and processing. In Steffens P. (Eds.), *Machine Translation and the Lexicon*, WMTL 1993. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), Vol. 898. https://doi.org/10.1007/3-540-59040-4_22
- Daelemans, W. (2005). *Memory-based language processing*. Cambridge University Press.
- Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, 47(2), 109–121. [https://doi.org/10.1016/S0022-2496\(02\)00016-0](https://doi.org/10.1016/S0022-2496(02)00016-0)
- Dawdy-Hesterberg, L. G. (2014). *The structural and statistical basis of morphological generalization in Arabic* [Doctoral thesis]. University of Illinois.
- 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>
- Drake, S. (2018). *L1 biases in learning root-and-pattern morphology* [Doctoral thesis]. University of Arizona.
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24. <https://doi.org/10.1093/applin/ami038>
- Gatt, A., & Čéplö, S. (2013). Digital corpora and other electronic resources for Maltese. In *Proceedings of the international conference on corpus linguistics*.
- Goldsmith, J. (1979). *Autosegmental phonology* [Doctoral thesis]. MIT, Published by Garland Press.

- Haas, W. G., & Trommelen, M. T. G. (1993). *Morfologisch handboek van het nederlands: Een overzicht van de woordvorming*. Sdu.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale and S. J. Keyser (Eds.), *The view from building 20* (pp. 111–176). MIT Press.
- Hammond, M. (1988). Templatic transfer in Arabic broken plurals. *Natural Language & Linguistic Theory*, 6(2), 247–270. <https://doi.org/10.1007/BF00134231>
- Haspelmath, M. (2020). The morph as a minimal linguistic form. *Morphology*, 30(2), 117–134. <https://doi.org/10.1007/s11525-020-09355-5>
- Hayes, B. (1980). *A metrical theory of stress rules* [Doctoral thesis]. Massachusetts Institute of Technology.
- Hockett, C. F. (1954). Two models of grammatical description. *Word*, 10(2–3), 210–234. <https://doi.org/10.1080/00437956.1954.11659524>
- Hockett, C. F. (1987). *Refurbishing our foundations: Elementary linguistics from an advanced point of view*. John Benjamins Publishing Company.
- Hoppe, D. B., van Rij, J., Hendriks, P., & Ramscar, M. (2020). Order matters! Influences of linear order on linguistic category learning. *Cognitive Science*, 44(11), 1–43. <https://doi.org/10.1111/cogs.v44.11>
- Kapatsinski, V. (2012). What statistics do learners track? Rules, constraints and schemas in (artificial) grammar learning. In S.T. Gries and D. Divjak (Eds.), *Frequency effects in language learning and processing* (pp. 53–82). De Gruyter Mouton.
- Kapatsinski, V. (2013). Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language*, 89(1), 110–148. <https://doi.org/10.1353/lan.2013.0003>
- Kapatsinski, V. (2018). *Changing minds changing tools: From learning theory to language acquisition to language change*. MIT Press.
- Keuleers, E., & Daelemans, W. (2007). Memory-based learning models of inflectional morphology: A methodological case-study. *Lingue e Linguaggio*, 6(2), 151–174. <https://doi.org/10.1418/25649>
- Köpcke, K., & Wecker, V. (2017). Source- and product-oriented strategies in L2 acquisition of plural marking in German. *Morphology*, 27, 77–103. <https://doi.org/10.1007/s11525-016-9292-5>
- Kunter, G. (2017). Coquery: A free corpus tool. Version 0.10.0. <https://www.coquery.org/>
- Leben, W. R. (1973). *Suprasegmental phonology* [Doctoral thesis]. Massachusetts Institute of Technology.
- Lõo, K., Järvikivi, J., & Baayen, H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition*, 175, 20–25. <https://doi.org/10.1016/j.cognition.2018.02.002>
- Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B., & Baayen, H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, 28(1), 71–97. <https://doi.org/10.1007/s11525-017-9318-7>
- Lounsbury, F. G. (1953). *Oneida verb morphology* (No. 48). Yale University Press.
- Loy, J. (2019). *Neural network projects with python: The ultimate guide to using python to explore the true power of neural networks through six projects*. Packt Publishing Ltd.
- Matthews, P. H. (1965). The inflectional component of a word-and-paradigm grammar. *Journal of Linguistics*, 1(2), 139–171. <https://doi.org/10.1017/S0022226700001146>
- McCarthy, J. J. (1979). *Formal problems in semitic phonology and morphology* [Doctoral thesis]. Massachusetts Institute of Technology.
- McCarthy, J. J., & Prince, A. S. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8(2), 209–283. <https://doi.org/10.1007/BF00208524>
- 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, 1125–140. https://doi.org/10.1162/tacl_a_00304
- Milin, P., Divjak, D., Dimitrijević, S., & Baayen, R. H. (2016). Towards cognitively plausible data science in language research. *Cognitive Linguistics*, 27(4), 507–526. <https://doi.org/10.1515/cog-2016-0055>
- Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PLOS One*. <https://doi.org/10.1371/journal.pone.0171935>
- Milin, P., Tucker, B. V., & Divjak, D. (2020). A learning perspective on the emergence of abstractions: the curious case of phonemes. arXiv preprint arXiv:2012.07499.
- Moscato del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & R. H. Baayen (2004). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1271–1278. <https://doi.org/10.1037/0278-7393.30.6.1271>
- Nakisa, R., Plunkett, K., & Hahn, U. (2001). Single- and dual-route models of inflectional morphology. In P. Broeder and J. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 201–222). MIT Press.
- Nieder, J., van de Vijver, R., & Mitterer, H. (2020). Knowledge of Maltese singular-plural mappings. *Morphology*, 31, 147–170. <https://doi.org/10.1007/s11525-020-09353-7>
- Nieder, J., van de Vijver, R., & Mitterer, H. (2021). Priming Maltese plurals: Representation of sound and broken plurals in the mental lexicon. *The Mental Lexicon*, 16(1), 69–97.
- Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, 197, 104081. <https://doi.org/10.1016/j.cognition.2019.104081>
- Nixon, J. S., & Tomaschek, F. (2021). Prediction and error in early infant speech learning: A speech acquisition model. *Cognition*, 212, 104697. <https://doi.org/10.1016/j.cognition.2021.104697>
- Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Experimental Psychology: General*, 115(1), 39–57. <https://doi.org/10.1037/0096-3445.115.1.39>
- Olejarczuk, P., Kapatsinski, V., & Baayen, R. H. (2018). Distributional learning is error-driven: The role of surprise in the acquisition of phonetic categories. *Linguistics Vanguard*, 4(s2), 20170020. <https://doi.org/10.1515/lingvan-2017-0020>
- O'Neill, P. (2014). The morpheme in constructive and abstract models of morphology. *Morphology*, 24(1), 25–70. <https://doi.org/10.1007/s11525-014-9232-1>

- Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of “mouses” in adult speech Error and expectation in language learning: The curious absence of “mouses” in adult speech. *Language*, 89(4), 760–793. <https://doi.org/10.1353/lan.2013.0068>
- Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6), 927–960. <https://doi.org/10.1080/03640210701703576>
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909–957. [https://doi.org/10.1111/\(ISSN\)1551-6709](https://doi.org/10.1111/(ISSN)1551-6709)
- R Core Team (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. <https://www.R-project.org/>.
- Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black and W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts..
- Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3), 151. <https://doi.org/10.1037/0003-066X.43.3.151>
- Schembri, T. (2012). *The broken plural in maltese: A description*. Univ.-Verl. Brockmeyer.
- Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure*. Vol. Cambridge University Press.
- Stump, G. T. (2018). Paradigm function morphology. In J. Audring & F. Masini (Eds.), *The oxford handbook of morphological theory* (pp. 284–304). Oxford University Press.
- Sutskever, I., Vinyals, O., & Le, Q. V.. (2014). Sequence to sequence learning with neural networks. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14)*. MIT Press, Cambridge, 3104–3112.
- Tomaschek, F., Plag, I., Baayen, R. H., & Ernestus, M. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naive discriminative learning. *Journal of Linguistics*, 57(1), 123–161. [doi:10.1017/S0022226719000203](https://doi.org/10.1017/S0022226719000203)
- Ussishkin, A., Dawson, C. R., Wedel, A., & Schluter, K. (2015). Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience*, 30(9), 1096–1115. <https://doi.org/10.1080/23273798.2015.1005635>
- Vandekerckhove, B., Keuleers, E., & Dominiek, S. (2008). The role of phonological distance and relative support in the productivity of the Dutch simple past tense. In *Quantitative investigations in theoretical linguistics (QITL3)* (p. 70).
- van de Vijver, R., & Baer-Henney, D. (2014). Developing biases. *Frontiers in Language Sciences*, 5(634). <https://doi.org/10.3389/fpsyg.2014.00634>