Morphology (2018) 28:297–323
https://doi.org/10.1007/s11525-018-9328-0

The form and productivity of the Maltese morphological diminutive

Shiloh Drake[1]

Received: 21 January 2017 / Accepted: 5 July 2018 / Published online: 19 July 2018

Abstract This paper examines the productivity and form of the morphological diminutive in Maltese. Maltese has lexical items and grammatical properties stemming from both Semitic and Indo-European roots; previous research has shown that there are different levels of productivity for Semitic and Indo-European morphology, which varies even among speakers. In addition, both the Semitic and Indo-European morphological diminutive may take several different forms in Maltese. The goals of this research are to determine whether native speakers of Maltese can use a morphological diminutive (like *wuggie*) rather than a lexical diminutive (like *little wug*); if they can, whether a default form exists for the morphological diminutive, and if so, whether the default form is Indo-European or Semitic in nature. A novel word elicitation task was used to test how speakers use the diminutive, and the results may be explained using a variety of different theoretical frameworks allowing for a hierarchical selection of a diminutive allomorph.

Keywords Root-and-pattern morphology · Maltese · Novel word elicitation task · Diminutive

1 Introduction

One of the central questions in the field of morphology is the issue of productivity, and how this relates to the structure of a lexical entry in the mental lexicon. Speakers of a language can be probed for their unconscious rules about the structure of words via a variety of tasks, such as lexical decision tasks (e.g., Rastle et al. 2004; Frost et al. 1997, 2000; Deutsch et al. 1998) and elicitation tasks, or wug tasks (e.g., Berko 1958).

✉ S. Drake
sndrake@email.arizona.edu

[1] Department of Linguistics, University of Arizona, P.O. Box 210025, Tucson, AZ 85721, USA

These unconscious rules can be used to predict features such as the decomposability of a given word or the status of morphemes in the lexicon.

Compared to other western European languages, Maltese is understudied and under-documented. The language is spoken by approximately 400,000 people on the island archipelago that makes up the Republic of Malta. Due to its history as a trade port located between Africa and Italy, Maltese has a long history of exposure to both Arabic and Sicilian Italian, and more recently, English. Traces of this influence remain today in the form of a dual lexicon: Words that originated in the Semitic lexicon, for instance, generally take morphological patterns more akin to Arabic and other Semitic languages, while words that were borrowed from Sicilian Italian and English tend to take morphological patterns more akin to Indo-European languages.

Diminutives are word forms used with the connotations of smallness, cuteness, affection, familiarity, or contempt, and are frequently used when social distance between two interlocutors is minimal (Schneider 2003:143, 160) or to soften requests (Jurafsky 1996). Diminutives have also been found in some languages (English, Dutch, and Russian) to facilitate word segmentation in infants (Kempe et al. 2007), which occurs due to the final-syllable invariant cases of diminutive marking in those languages (e.g., *birdie*, *horsie*, *doggie*). In the case of minimal social distance between two interlocutors, this may result in the use of nicknames, such as *Nicky* or *Suziekins* for *Nicholas* and *Suzanne* (Jurafsky 1996), or diminutives may be used in order to claim higher status in a social hierarchy ("Let's get started, kiddos" may be appropriate when speaking to children, but is less appropriate in a board meeting unless the speaker is significantly higher in status than their colleagues, or is contemptuous of them) (Schneider 2003:143). Generally speaking, diminutives are used in highly informal situations and primarily used in spoken language, and much less likely to be used in formal situations or in writing (Schneider 2003:138). Cross-linguistically, the diminutive seems to have developed from morphemes meaning "child" and developed other diminutive senses diachronically (Jurafsky 1996:562ff).

This research seeks to answer two questions. First, what form does the Maltese morphological diminutive take and what form occurs most frequently? In current grammars of Maltese (Borg and Azzopardi-Alexander 1997:279–280), there are multiple attested morphological diminutives, some of which are borrowed from Sicilian, and some of which are Semitic in nature. Consultants suggest that people are more likely to use a lexical diminutive because the morphological diminutives are antiquated or lexically frozen and not decomposable (Luke Galea, p.c.). This work seeks to clear the somewhat muddy waters of whether a morphological diminutive is in fact still being used in modern Maltese, and if so, whether the morphological diminutive is Semitic-like or Indo-European-like.

Second, are native speakers sensitive to the different word formation rules required by their dual lexicon? When they are presented with a word that has a CV skeleton similar to a Semitic word, will they preferentially apply a Semitic pattern, and vice-versa for a word with an Indo-European CV skeleton? Previous work is somewhat inconclusive. In novel word elicitation tasks, Twist (2006) found that the majority of speakers used an Indo-European affix to form denominal verbs regardless of what the origin of the nonsense word seemed to be. The remaining subjects were sensitive to whether the nonsense words they were presented with had a Semitic

or Indo-European CV skeleton, and tended to apply either a Semitic pattern or an Indo-European affix respectively. However, Spagnol (2011) found that when speakers were asked to form deverbal nouns, virtually without exception Indo-European affixes were used rather than Semitic patterns. Again, this work seeks to shed light on the contradictory results seen in previous work.

2 Background

2.1 Typology of the diminutive

In many language families, the diminutive takes the form of a suffix: for example, English *-ie/y*, *-let*, and *-ling*, German *-lein* and *-chen*, Spanish *-ito/a*, Dutch *-(et/t/p/k)je*, Italian *-etto/a*, *-ello/a*, and *-ino*, Farsi *-cheh* and *-ak*. This type of diminutive formation is also common outside the Indo-European family, and many other languages throughout the world use suffixes for the diminutive: Finnish (*-ke*, *-ka*, and *-nen*) and Estonian (*-ke*) in the Uralic family, Turkish (*-cik* and *-ceğiz*) in the Turkic family, and Chinese (*-zǐ*, though the prefix *xiǎo-* may also be used) in the Sino-Tibetan family.

As with other morphological processes, considerations of allomorphy and blocking come into play in considering productivity and the formation of diminutives. When a language has more than one form that conveys a similar meaning, there may be differences in productivity (Aronoff 1976): In English, for example, it is intuitive to indicate that something is small and cute by attaching the productive *-ie/y* /i/ morpheme, even in nicknames: *Bobby*, *Lizzy*, *Nicky*. An English speaker would probably not use the less productive diminutives *-ling* or *-let* in producing a nickname; forms like "Robling" or "Nicklet" are unlikely to occur spontaneously.

In languages of the Semitic family, which are morphologically nonconcatenative, other types of morphological patterns for the diminutive occur. For instance, Hebrew uses final-syllable partial reduplication to form the diminutive productively (Bolozky 1994), as in (1).

(1)

kélev	*klav -lav*
dog	dog -dim
dag	*dag -ig*
fish	fish -dim
géver	*gvar -var*
man	man -dim

(Bolozky 1994)

Modern Standard Arabic uses separate nominal word patterns to convey the diminutive, one of which is a $C_1uC_2ajC_3$ template.[1]

[1] These templates are referred to generically as *wazan* (pl. *awzaan*) in Arabic grammars, pedagogical literature, and by Arabic linguists, or *binyan* in some Hebrew literature (seen in (2)) (McCarthy and Prince 1990). In this form, the root $\sqrt{f?l}$ is treated as a placeholder for the three root consonants in a template, similar to the CV skeleton above.

(2)

qidħ	*qudajħ*
arrow	arrow.dim
ʔasad	*ʔusajd*
lion	lion.dim
raʒul	*ruʒajl*
man	man.dim

(McCarthy and Prince 1990)

Other dialects of Arabic form the diminutive in other ways, such as via reduplication (*kuskusii* 'puppy', San'ani Arabic (Watson 2006); C_1uC_1u hypocoristics, Egyptian and Lebanese Arabic (Samantha Wray, p.c.)), and the use of other patterns (cf. Watson 2006:196 for further examples).

From these examples, it can be seen that the diminutive form in non-concatenative languages does not have the same morphological properties as it does in concatenative languages. There is a different kind of form consistency, but perhaps not the same type of consistency that would aid infants in word segmentation. This is because infants rely on regularities at the edges of the words in order to segment them from the speech stream (Kempe et al. 2007), and this regular morphological marking does not necessarily occur at word edges. In Hebrew and Arabic, there are few concatenative ways a speaker can form the diminutive (e.g., suffixation of /i/ in San'ani (Watson 2006)). As speakers will discuss things with diminutive qualities, we will assume that if a language only has one diminutive-forming rule, then that rule is productive. This rule may be to use a lexical diminutive, a morphological diminutive, or other form of marking a noun for diminutive characteristics, but can ostensibly be applied to borrowed words or neologisms if they take on diminutive features.

Maltese is a language that is of particular morphological interest due to its dual lexicon: When a word is a part of the Semitic lexicon, it tends to undergo morphological processes typical of Semitic languages; that is, root-and-pattern morphology is typically used. However, when a word is borrowed from Sicilian or English, it typically undergoes concatenative morphological processes more typical of Romance languages. In (3) below, *xatt* 'shore' and *toqba* 'hole' are words from the Semitic vocabulary, while *kompjuter* 'computer' and *teżor* 'treasure' are words from the Indo-European vocabulary.[2] Exemplified in (3), Arabic-origin words tend to form their plurals via a templatic process, while words incorporated from Indo-European languages tend to form their plurals via a suffixation process, though this is not a hard-and-fast rule and some researchers argue that templatic morphology is not productive in Maltese (e.g., Mifsud 1995; Spagnol 2011).

(3)

xatt	*xtut*	*kompjuter*	*kompjuter -s*
shore	shore.pl	computer	computer -pl
toqba	*toqob*	*teżor*	*teżor -i*
hole	hole.pl	treasure	treasure -pl

(Borg and Azzopardi-Alexander 1997)

[2] All Maltese examples are given in standard orthography. <q> indicates a glottal stop, <ċ> indicates a voiceless palato-alveolar affricate, <ġ> indicates a voiced palato-alveolar affricate, <ż> indicates a voiced alveolar fricative, <z> indicates a voiceless alveolar affricate, and <x> indicates a palatal fricative.

This "dual-lexicon" effect means that there could be multiple ways of marking the diminutive in the language, one of which may look more like a Romance diminutive, and another that looks more like a Semitic diminutive. The productivity of the various diminutive markers may also be affected, depending on which words are more readily accepted into the Maltese lexicon and how the grammar of Maltese treats the accepted words.

In Maltese, there are multiple attested allomorphs of the Semitic diminutive, even within the templatic/affixal distinction. They take the form of an infixed *-ajje-*, *-ajja-*, *-ejje-*, *-ejja-*, *-ajC₃a*, or *-ejC₃a* (Borg and Azzopardi-Alexander 1997:279; see also Aquilina and Cassar-Pullicino 1957), with the C_3 representing the third root consonant. Rather than being infixes, these look like a set of templatic allomorphs, as each of these forms also starts with the C_1C_2 root consonants along with having the vowel and glide infixed. That is, it looks much like an Arabic diminutive template. Example words are shown in (4) with the diminutive pattern in bold. Each word represented is of Semitic origin; other examples of this diminutive template in the grammar are also from the Semitic vocabulary.

(4)

ġobon	*ġb**ejn**a*	
cheese	cheeselet	
tifel	*tf**ajj**el*	
boy	boy.dim	
xatt	*xt**ajt**a*	
shore	shore.dim	(Borg and Azzopardi-Alexander 1997)
toqba	*tq**ajb**a*[3]	
hole	hole.dim	
fqir	*fq**ajj**ar*	
poor	poor.dim	(Aquilina and Cassar-Pullicino 1957)

What these data first demonstrate is that the Indo-European vocabulary is discussed in less detail than the Semitic vocabulary; the grammars provide just a few examples of forming the diminutive with affixes with Sicilian loanwords (seen in (5)). These are not indicated as being borrowed words, but instead are "originally diminutive" (Borg and Azzopardi-Alexander 1997:280). That is, Borg and Azzopardi-Alexander (1997) do not differentiate between the Semitic and borrowed Indo-European vocabularies in this particular instance, but do differentiate between classes of words that are not originally diminutive (that is, words that either originate in the Semitic lexicon, or are borrowed in a non-diminutive form, such as *xatt* 'shore') and lexicalized diminutives that can take an additional diminutive suffix once borrowed into Maltese (such as *berritta* 'cap', which was borrowed with the Sicilian Italian diminutive suffix *-itta* and, as a lexicalized form, it may take a Maltese suffix, shown below).

[3] As an anonymous reviewer points out, tqajba is no longer widely used, and may not be used in everyday speech by the younger speakers who took part in Experiment 2.

(5)

berritta	*berritt-in*
cap	cap -dim
biskott	*biskutt-in*
biscuit	biscuit-dim
ċikkulata	*ċikkulat -ina*
chocolate	chocolate-dim
tromba	*trumb-etta*
cylindrical shape	trumpet

(Borg and Azzopardi-Alexander 1997)

Aquilina and Cassar-Pullicino (1957) suggest that other ways of forming the diminutive exist (including with Sicilian loanwords), but they also note that the words used are considered archaic. This grammar also fills out more templatic diminutives, listed below, which also suffers from the same underrepresentation of the Indo-European sublexicon. These diminutive patterns are as follows: the presence of /*w*/ as a weak root consonant (which typically occurs as the second radical of the word (Aquilina and Cassar-Pullicino 1957:4)) (6a) or adding the feminine suffix *-a* to the end of a masculine noun (6b), and in the case of infant- or child-directed speech, repeating the final consonant of the word and adding a "long *u*", or in other examples given, an *a* (6c). The examples in (6c) seem to be formed similarly to the Hebrew diminutives in (1), with partial final-syllable reduplication.

(6)

a.	*dar*	*dwejra*
	house	house.dim
	ruħ	*rwejħra*
	soul	soul.dim
b.	*bieb*	*bieb -a*
	door	door -dim
	senduq	*senduq -a*
	chest	chest -dim
c.		*bażużu*
		darling (child-directed speech only)
	żaqq	*żaquqa*
	belly	belly.dim
	sinna	*sinnuna*
	tooth	tooth.dim

(Aquilina and Cassar-Pullicino 1957)

With little information about the Indo-European sublexicon, it is possible that there is not a productive way to form the diminutive with the Indo-European loanwords, and the diminutives encountered are simply frozen, lexicalized forms. It is worth noting that when explicitly asked, native speakers are skeptical of any morphological productivity for the diminutive in Maltese, instead asserting that forms such as *tfajjel* 'boy.dim', *tfajla* 'girl.dim', and *ġbejna* 'cheeselet' are lexicalized, and that speakers would more likely use an adjective to denote the diminutive (Luke Galea, p.c.). This would take the form of something like *qanfud żgħir* 'little hedgehog'.

Because this infix is inserted between (or around, in the case of feminine nouns) the two final root consonants, this is similar to the *fu?ajl* pattern in Modern Standard Arabic, which speaks to the shared Semitic roots of the two languages. However, Modern Standard Arabic's vowels in the pattern (e.g., _u_aj_) tend to be invariant. One natural hypothesis concerning the vocalic variation in these allomorphs of the diminutive might be that the choice in vowel is phonologically governed. However, such a hypothesis ascribing phonologically governed allomorphic qualities to these vocalic variations would seem to conflict with the templatic nature of this diminutive morphology, given that the root-and-pattern, non-concatenative qualities of Semitic languages seem to hold here. Thus, the vowels in a given pattern would be expected to behave as morphologically, rather than phonologically, conditioned elements. That is, vocalic alternations in templatic systems are typically governed by factors other than, e.g., coarticulation with consonants or vowel harmony. For example, the history of Maltese suggests that vowel alternations in the patterns have come from borrowed Indo-European words, such as the *i-a* and *o-a* sequences, while other research suggests that some vowel sequences occur based on the phonological qualities of the surrounding root consonants (further discussed in Spagnol 2011). Researchers are not in agreement about this kind of vowel change in the patterns, but future research might benefit from investigating the explanatory potential of such a hypothesis.

2.2 Modeling morphological competition

Some research suggests that type frequency rather than token frequency affects affix productivity (Bybee 1995; Pierrehumbert 2001), while others suggest that neither token *nor* type frequency contributes to productivity, but rather, productivity depends on the frequency of the decomposed forms (Hay and Baayen 2002). This line of research suggests that when a multimorphemic word is compositional (i.e., it can be broken into constituent meaningful pieces with each contributing their interpretation to the interpretation of the whole), this facilitates retrieval from the lexicon based on the frequency of the meaningful parts individually. Hay and Baayen provide *listless* and *tasteless* as examples: *tasteless* is easily decomposed into the discrete morphemes 'taste' and '-less', while the same operation cannot be performed to the same effect with *listless*. This is because *list-* in this context is not a meaningful morpheme unless it appears in the same context as *-less*; the meaning of *listless* is noncompositional.

If Maltese speakers are able to decompose words into their constituent pieces when asked to form the diminutive, and the diminutive is not a single, listed lexical item, then we can consider the diminutive to be productive. Further, if Maltese speakers decompose words differently depending on whether the words have a triconsonantal root available for decomposition or not, then we might expect to see a difference in whether the diminutive is created via a templatic or suffixal morphological operation.

If speakers have multiple diminutive allomorphs, perhaps one templatic and one affixal, then this must be accounted for by something other than hypothesizing a single morphological rule that is applied regardless of the form that the base takes. This might happen in the case where speakers use a templatic pattern when a nonce word or loanword sounds Semitic in origin, and when speakers use an affix in the case where a word sounds Indo-European in origin. Accounting for this can take place

under many existing frameworks, such as a dual-route system (Pinker and Prince 1988; Prasada and Pinker 1993), a stochastic rule system (Pierrehumbert 2001), a fine-grained list of rules (Albright and Hayes 2003), co-phonologies (after Inkelas and Zoll 2007), or a model based on analogy (Krott et al. 2001). These models and their implications for the results of this work are discussed further in Sect. 5.

In order to ascertain which model or models best accounts for the grammatical split in Maltese speakers' lexicons, and to find out whether it is true that native speakers do not use the Semitic diminutive morpheme productively, I use a corpus analysis (Sect. 3) and a novel word elicitation (Sect. 4). A corpus analysis allows us to see whether each of the diminutive morphemes is actually used in written material, while the novel word elicitation allows us to ask speakers to apply an uncommon derivational morpheme to nonce words to see how speakers prefer to derive words that they are not familiar with.

3 Maltese diminutives in the Maltese *Corpus Malti*

A corpus analysis was carried out to find more information about the distribution and use of morphological diminutives in Maltese. This study had three primary aims: (1) to find out whether morphological diminutives formed productively; if so, (2) whether both affixes and templates were productive; and (3) whether any of the multiple attested allomorphs from Borg and Azzopardi-Alexander (1997) and Aquilina and Cassar-Pullicino (1957) was used more frequently than others.

The Maltese Language Resource Server Corpus (Borg et al. 2011) was used to search for the Semitic templatic diminutives and the Indo-European diminutive affixes described by both Borg and Azzopardi-Alexander (1997) and by Aquilina and Cassar-Pullicino (1957)—that is, the "infixed" forms *-ajje-*, *-ajja-*, *-ejje-*, *-ejja-*, *-ajC₃a*, and *-ejC₃a*, and the suffixed forms *-in*, *-ina*, and *-etta*. The most frequent forms by type that appear in the corpus were taken to be the more productive forms. Thus, although productivity is not conditioned on frequency alone, that is the measure I have opted to use for the purposes of this corpus study.

3.1 Methods

Data were collected from the Korpus Malti v.3.0 on the Maltese Language Resource Server (MLRS) (Borg et al. 2011). This corpus consists of ∼250 million tokens and is made up of legal and academic texts, literary texts, parliamentary debate texts, newspaper texts, religious texts, speeches, general web text, text from the Maltese Wikipedia, and a section called "miscellaneous". In spite of this array of text, we can anticipate that we may not encounter many diminutive forms since the diminutive tends to be more prevalent in informal registers, as noted above (Schneider 2003). However, the general web text section includes text from blogs written in Maltese, so it is possible to encounter some diminutive forms from those sources.

The search string was for the word pattern or affix; namely, the second root consonant preceding *ejje*, *ejja*, *ajje*, *ajja*, *aj*, or *ej*, and then the third root consonant following *ejje*, *ejja*, *ajje*, or *ajja* to end the word, or the third root consonant following

aj or *ej*, followed by *a* to end the word. This was restricted to nouns, as diminutive marking selects for nouns (Borg and Azzopardi-Alexander 1997). A sample search string used to search the corpus is shown in (7).

(7) +[aj,ej]?a_NOUN

Notation reads "*one or more characters followed by <aj>* or *<ej>, followed by one character, followed by <a> at the end of the word, restricted to words tagged as nouns*".

Translations for the corpus results were obtained using the Translate feature on the Google Chrome browser. This allowed for a "quick and dirty" translation of the entire webpage to determine whether the words were diminutive or not, as the sequence of letters that make up the diminutive pattern are homophonous with other morphemes, such as the plural. For example, *flejjes* 'monies' or *iskejjel* 'schools' are not diminutive forms of *flus* or *skola* respectively, while a word such as *ġbejna* 'cheeselet' is the diminutive form of the word *ġobon* 'cheese'. The words left untranslated by Google likely to be diminutives (such as those used in a blog that have the Semitic-type pattern, and where the context seemed appropriate for diminutive usage, rather than in the context of a parliamentary debate, where the formal register of the speeches is not typically conducive to the use of a diminutive) were compiled into a list for checking in a dictionary or translation by a native speaker.

As words are not tagged for being diminutive in the corpus and both patterns and suffixes are conflated with homographic morphemes, the corpus was searched both for the sub-lexical search strings provided above and for representative diminutive words. The representative diminutives were either found within the first 10,000 results of the search for a given morpheme, or they were words found in the Maltese grammars (Borg and Azzopardi-Alexander 1997; Aquilina and Cassar-Pullicino 1957). This method of searching means that the numbers represent the frequency of each letter string occurring, and the data had to be modified for it to be meaningful. This being said, it can still provide an idea of the morphemes that are most frequent and are thus faster to access in the mental lexicon (cf. Family Size Effect: de Jong et al. 2000) and hence which allomorph Maltese speakers may access if they want to use a diminutive.

3.2 Results

Table 1 shows each search string used for the corpus search, the number of matches, and the match frequency. This count includes likely false positives for the reasons discussed above.

As it was necessary to clean the data, translations of the first 10,000 results for each letter string were checked by hand. Every instance of a diminutive word was collected and searched for in the corpus. Table 2 contains the diminutive words attested in the corpus and their frequency after processing the data, which is further described in detail below. As outlined below, some of the strings attested in the grammar were not associated with any diminutive words in the grammar, while others (particularly the concatenative *-ina* and *-etta*) were more frequent. Based on the occurrences per million of each word associated with the remaining patterns, the most

Table 1 Match count and frequency of orthographic strings in MLRS Corpus

String	Matches	Match frequency (per million)
+ejja?_NOUN	2870	11.51
+ejje?_NOUN	94,035	377.26
+ajja?_NOUN	13,089	52.51
+ajje?_NOUN	13,296	53.34
+aj?a_NOUN	130,796	524.74
+ej?a_NOUN	114,331	458.69
+in_NOUN	570,190	2,287.56
+ina_NOUN	95,456	382.96
+etta_NOUN	50,795	203.79

Table 2 Diminutive types found, token count for each type, and occurrences of the type per million words as a measure of frequency in the corpus after cleaning the data

Word	# Matches	Occurrences (per million)
tfajjel	849	3.41
tfajla	11,650	46.74
xtajta	394	1.58
dwejra	96	0.39
ġbejna	154	0.62
plattin	90	0.36
biskuttin	19	0.08
sinjorina	580	2.33
kuċċarina	655	2.63
kartolina	443	1.78
ċikkulatina	6	0.02
trumbetta	106	0.43
istatwetta	17	0.07
pipetta	238	0.95
favetta	23	0.09

frequent diminutive is the pattern *-ajC₃a-* (48.32 occurrences per million), followed by the suffix *-ina* (6.76 occurrences per million), the pattern *-ajje-* (3.41 occurrences per million), the suffix *-etta* (1.54 occurrences per million), the pattern *-ejC₃a-* (1.01 occurrence per million), and finally the suffix *-in* (0.44 occurrences per million).

-ejje- and *-ejja-* are common letter combinations in words that are not diminutives, as shown in (8) (word containing the pattern in question is in bold).

(8) a. *Il- proġett jagħti sapport lill- **iskejjel** biex jgħallmu*
DEF project give.PROG support to.the school.PL to teach.3PL
suġġet-ti relevanti għall globalizzazzjoni…
subject-PL relevant to.the globalization
'The project provides support to **schools** to teach subjects relevant to globalization…'

b. *L- ebda annimal ma għandu jiġi mġiegħel isofri*
DEF no animal with have.PST.3PL to force.PST suffer.3SG
uġigħ, tbatija, jew ***dwejjaq*** *mingħajr bżonn…*
pain suffering or sadness without need
'No animal should be forced to suffer pain, suffering, or **sadness** unnecessarily…'

In the examples in (8), both the meaning and context of the words (*iskejjel* 'schools' and *dwejjaq* 'sadness') make them unlikely to contain a diminutive morpheme, and this is confirmed with a Maltese-English dictionary (Vella 2015), as these are the listed dictionary forms. In general, dictionary forms do not have derivation attached to them; that is, entries in a dictionary are either the bare uninflected form, or may (in this case) include plural inflection, but little else.

The *-ejje-* sequence is much more common than *-ejja-* is; however, very common words happen to match a particular string type. Many instances of *flejjes* 'monies' (1,807 tokens) or *iskejjel* 'schools' (19,318 tokens), caused the count to skew high. *-ejje-* was associated with a much larger set of word types (330 types, of which 172 occurred only once), but no diminutives were found in spite of this larger set. The search for *-ejja-* yielded no diminutive words, and the words that were matched were limited to a small set of types. The matches consisted primarily of forms of *dwejjaq* 'sadness' (1,833 tokens), *tfejjaq* 'cure' (79 tokens), and *ħlejjaq* 'wildlife, beings' (212 tokens)—that is, the type count was very low. *-ajja-* was similar, though it had a higher number of untranslated words in the results than the other searches, which may indicate the presence of diminutive forms that would not occur as the dictionary forms. None of the results for *-ajja-* were diminutives.

When searching for *-ajje-*, it was a different story. The diminutive form of *tifel* 'boy', *tfajjel*, occurred 849 times in the corpus, or approximately 3.41 times per million words. This may be because the usage of this diminutive is similar to Italian *ragazzo/a* 'young person', which is used to describe people in their teens and early twenties. Making this more plausible is the fact that it was used to refer to both males and females, and was defined as "youngster" as opposed to "small boy" or similar. *-ajje-* is also confounded with a plural form—there were many results for *dgħajjes* 'boats', for example, and the singular form is conflated with the *-ajC₃a-* diminutive pattern (sing. *dgħajsa*, 'boat', but not '*little boat'). *Dgħajsa* is also not morphologically related to semantically similar words, such as *vapuri* 'ship', which would suggest the possibility that it is a diminutive form of an unmarked word with similar but non-diminutive semantics.

-ajC₃a- showed a result similar to *-ajje-*. *tfajla*, the diminutive form of *tifla* 'girl', appeared 223 times in the first 10,000 results, making up 2.23% of the matches for the search. *Tfajla* occurs 11,650 times in the corpus, or 46.74 instances per million words, making it much more common than *tfajjel* in the corpus. This may be a form that is similar to the Italian *ragazza*, referring specifically to a female young person, and it may be used more often than *tfajjel* for sociocultural reasons. For example, it is more common in spoken Maltese to refer to a young female as a *tfajla* than it is to refer to a young male as *tfajjel*; in fact, *tfajjel* is used mostly to refer to toddler-aged or young school-aged boys, while *tfajla* could refer to a four-year-old or a 20-year-old (Luke Galea, p.c.). As the results are substantially different from other words

searched (e.g., *tfajjel* or the other words that follow), the corpus was also searched for *xtajta* 'beach', the diminutive form of *xatt* 'shore' (Borg and Azzopardi-Alexander 1997:279). *Xtajta* appears 394 times in the corpus, for a frequency of 1.58 instances per million words.

-ejC₃a- also had a similar result, with *dwejra*, the diminutive form of *dar* 'house', providing a high number of matches in the first 10,000 results. When the corpus is searched for *dwejra*, there were 906 tokens (frequency of 3.63 instances per million words), but only 96 of those were instances of a word that was translated as 'cottage'; that is, a 'little house'. The remaining matches were translated as 'azure', likely because the location of a natural landmark, the Azure Window (or the Dwejra Window), located in the town of Dwejra on the island Gozo. The instances where *dwejra* was translated as 'azure' were removed from the count (leaving 96 tokens of *dwejra* from the entire corpus). The form *dwejra* meaning 'cottage' thus occurs approximately 0.39 times per million words.

Ġbejna 'cheeselet', the diminutive form of *ġobon* 'cheese' appeared twice in the first 10,000 results for *-ejC₃a-*, which makes up 0.02% of the matches for the search. When the entire corpus is searched for *ġbejna*, there are 154 tokens with a frequency of 0.62 occurrences per million words. None of the instances of *ġbejna* are translated, but for all of them the context makes sense to be talking about cheeselets (9).

(9) a. *. . . Chris Zahra, li jrabbi 60 nagħaġ u jipproduċi 440*
Chris Zahra to breed.3SG 60 sheep.PL and produce.3SG 440
ġbejna tan- nagħaġ kuljum minn ħalib pasturizzat. . .
cheese.dim from sheep every.day with milk pasteurized
'. . . Chris Zahra, who keeps 60 sheep, produces 440 cheeselets daily from pasteurized sheep milk. . . '

b. *. . . ir- raħħala li jkabbru n- nagħaġ u l- mogħoż biex*
DEF farmer.PL to breed.3PL DEF sheep.PL and DEF goat.PL to
jipproduċu l- ġbejna Maltija.
produce.3SG DEF cheese.dim Maltese
'. . . the farmers who raise sheep and goats to produce Maltese cheeselets.'

Searching the corpus for *-in* resulted in 570,190 matches. Of the first 10,000 results, the only diminutive found was *plattin* 'saucer', the diminutive of *plat* 'plate'. Most of the matches were with *snin* 'years', *ħin* 'time', and *ċittadin* 'citizen', none of which are diminutives. In the entire corpus, *plattin* accounted for 90 matches, or a frequency of 0.36 per million. Of the other *-in* diminutives in the grammar (Borg and Azzopardi-Alexander 1997:279), *biskuttin* 'small biscuit' returned 19 matches for a frequency of 0.08 per million words, while *berrittin* 'small cap' was not found at all.

-ina had 95,456 matches. In the first 10,000 results, *sinjorina* 'Miss/Ms., title for a young woman', *kuċċarina* '(tea)spoon', *kartolina* 'postcard', and *ċikkulatina* 'piece of chocolate' (cf. *sinjura* 'Mrs.', *kuċċarun* 'ladle', *karta* 'paper, card', and *ċikkulata* 'chocolate (bar)') all appeared. *Kuċċarina* was the most frequent of these diminutives with 655 matches (2.63 occurrences per million), followed by *sinjorina* (580 matches; 2.33 occurrences per million), *kartolina* (443 matches; 1.78 occurrences per million), and *ċikkulatina* (6 matches, 0.02 occurrences per million).

-etta had 50,795 matches. In the first 10,000 results, *trumbetta* 'trumpet', *istatwetta* 'statuette', *pipetta* 'pipette', and *favetta* 'beans' were the diminutives found. When the entire corpus was searched for each word, *pipetta* was the most frequent with 238 matches (0.95 occurrences per million), followed by *trumbetta* (106 matches; 0.43 occurrences per million), *favetta* (23 matches, 0.09 occurrences per million), and *istatwetta* (17 matches, 0.07 occurrences per million).

3.3 Discussion

Of the diminutive morphemes, the most frequent orthographic sequence is *-in*, although this conflates the diminutive, other homophonous morphemes, and monomorphemic nouns that end in *-in*. Of each diminutive word that was searched in the corpus, the most frequent by a large margin was *tfajla*, followed by *kuċċarina*. The diminutive morphemes with the largest number of words associated with it in the corpus are *-ina* and *-etta*, both with four exemplars. This would suggest that these morphemes are the most productive in terms of everyday usage. It is also important to note that they are found frequently in a formal register that is not usually associated with diminutive use, which would either suggest that the provided examples are fully lexicalized and therefore not decomposed by native speakers—and therefore also not examples of a productive morpheme—or the two morphemes are so pervasive that their usage seeps into a formal register.

On the other hand, there were no diminutive examples for the *-ejje* and *-ejja* morphemes, which could mean that these morphemes are simply not productive, or that they are productive and are simply not used as frequently as other morphemes might be in formal registers. Due to the makeup and relatively small size of this corpus, it is not clear which of these possibilities is more likely.

The continued lack of clarity about how productive diminutive morphology is in Maltese is can be partly resolved by asking speakers to derive new words following their mental morphological grammar. Thus, a more promising way to assess the productivity of the diminutive patterns is to perform a Wug Test, after Berko (1958), and ask Maltese native speakers to provide examples of small and/or cute versions of phonotactically legal nonwords ("This is a wug. This is a smaller one. What would you call it?"). Further, in doing this type of elicitation task, we can ask the question of whether Semitic-sounding nonsense words get a Semitic-sounding pattern and Indo-European-sounding nonsense words get an Indo-European-sounding affix. Twist (2006) created a list of phonotactically legal nounlike nonwords for both the Semitic- and Indo-European-derived vocabularies, so the forms could be elicited using these items that have been used successfully in previous experiments. Performing this kind of elicitation would show whether certain affixes were more productive than others, as well as whether the domain (i.e., language family) for the Semitic templatic diminutive and Indo-European suffixal diminutive is determined purely phonologically.

4 Elicitation task

A nonce probing task, or Wug Test (after Berko 1958), was used as an additional measure to test the productivity of the diminutive allomorphs in Maltese. This task

provides a way to assess morphological productivity, as adults are asked to apply what they already know about Maltese morphology to phonotactically legal novel words, and will ostensibly follow the morphological rules that they have in place rather than creating new and never-before-seen rules (cf. Berko 1958). The task allows subjects to use concatenative or non-concatenative morphology freely and will provide a clearer picture of what is most intuitive for Maltese speakers.

In a Wug Test, participants see an imaginary animal along with a carrier phrase. The carrier phrase names the animal and asks for the participant's input. The canonical phrase is, "This is a *wug*. Here is another one. There are two _____." The participant then must fill in the plural form of *wug*, most likely following English's productive morphological rules (add /-s, z, ɪz/, as opposed to /-ɛn/ or another plural allomorph), and also English's phonotactic rules (if the final consonant is voiced, add /-z/). When this task is used with children, it shows the developmental trajectory of employing inflectional and derivational morphology (Berko 1958). With adults, it shows the application of productive inflectional or derivational morphology to novel words. In the case of the diminutive in Maltese, where there is some question of whether a morphological diminutive is productive at all and whether speakers will use morphology congruent with the lexical items' language family, this type of elicitation task will help to answer these questions.

4.1 Method

4.1.1 Participants

34 native Maltese speakers were recruited from the University of Malta and a radio show on the station *Radju Malta* to participate in the experiment. They were paid €5 for their participation. Participants ranged in age from 18 to 69 years, with a median age of 21 years. All subjects were bilingual in Maltese and English.

4.1.2 Materials

40 words from Twist's (2006) novel word elicitation were used (see the Appendix). 20 words from the Semitic list and 20 from the Indo-European list were chosen. These words were vetted by a native Maltese speaker for this experiment and minor alternations were made to the original stimuli to ensure they were phonotactically sound according to native speaker intuitions. The method used to generate these words is outlined below (Twist 2006:91–95).

Real words selected randomly from the Semitic or Indo-European sublexicons were used as the basis for the nonsense words. For the Semitic nonwords, Twist generated a list of unattested triconsonantal roots, which were then mapped onto vocalic patterns of real Maltese words of Semitic origin. For example, the nonce root $\sqrt{xsn}$ was mapped onto the prosodic template for *toqba* 'hole' (CVCCV) to produce *xVsnV*. For the Indo-European nonwords, prosodic templates from existing Maltese words of English origin were used. Single consonants were filled in at random, while consonants in clusters were replaced by consonants of matching sonority. For example, for the real word model *drill* 'drill', *br* was used for the cluster and a random consonant (in this case, *f*) was selected to fill in the final consonant slot. This resulted in

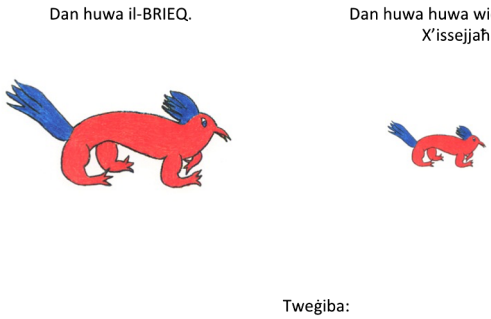

Fig. 1 Examples of imaginary animals used in the task

brVff. Then for both the Semitic and Indo-European sequences, vowels were filled in randomly. The examples above yielded *xesna* and *braff*, respectively.

40 pictures of imaginary animals (see Fig. 1) were randomly paired with each nonsense word to avoid any effect of semantic cueing or bias from the picture/word pairing. All pictures were a subset of those used in a novel animal elicitation task by Ohala (1999). The pictures are a blend of a variety of real animals; they look like possible animals, but do not strongly resemble any animals that currently exist.

Practice materials comprised four real-word examples from Maltese and two nonsense word/animal pairings. Two of the real-word examples were Semitic in origin (*ġobon*/*ġbejna* 'cheese/cheeselet'; *triq*/*trejqet* 'street/alley') and two were Indo-European (*festa*/*festina* 'party/little party'; *kuċċarun*/*kuċċarina* 'ladle/teaspoon'), and each word was paired with a picture of the item in question. The nonsense words were selected from Twist's (2006) word elicitation and did not appear in the experiment, and they were paired with two additional imaginary animal pictures that were also not used in the experiment trials. One nonsense word was modeled after a Semitic prosodic template (*brieq*), and the other was modeled after an Indo-European prosodic template (*korfa*). With these items, participants were asked to make their best guess as to how to refer to a smaller version of the provided animal. Participants were instructed to go through the practice materials at their own pace.

At the end of the experiment, participants were asked to fill out a language background questionnaire to determine their practices of language use and dominance. The questionnaire was adapted from Twist's (2006) materials.

All materials except for the verbal instructions and the language background questionnaire were presented in Maltese in order to minimize any English-based morphological bias; verbal instructions and the language background questionnaire were presented in English.

4.1.3 Procedure

The task was administered in a well-lit quiet room at the University of Malta. The researcher remained in the room throughout the practice session to answer any questions the participants had, and then left when the participants began the experimental task. Upon finishing the experiment, participants were asked to fill out a language background questionnaire.

The task was presented on a 13″ Macbook Pro (OS 10.8) using Matlab (version 2013B) and the Psychophysics Toolbox extension (Brainard 1997; Pelli 1997;

Kleiner et al. 2007). It was untimed and self-paced for the participants. For each trial, participants were shown a novel animal and its name, and then a scaled-down version of the same image with the text *Dan huwa huwa wieħed izgħar. X'issejjaħlu?* 'Here is a smaller one. What would you call it?' Participants then typed their response using an American QWERTY keyboard with alternate keybindings for the Maltese letters *ċ*, *ġ*, *ħ*, and *ż* (where subjects instead typed 7, 6, 4, and 2 respectively). Participants were instructed that they could use as many or as few words as they needed to name the smaller version of the animal to allow for the possibility of a lexical diminutive (e.g., "little wug"), the use of determiners (e.g., "a wug/the wug"), or a morphological diminutive (e.g., "wugling").

4.1.4 Response coding

Responses (referred to as "response type") were hand-coded as Semitic or non-Semitic to match the word type, which was also either Semitic or Indo-European (or non-Semitic). Non-Semitic responses were classified as those where the sequence of letters in the original word provided was preserved and an additional sequence of letters was added to one edge of the word. For example, *fintinu* and *fintina* were given as responses to *fint* and were classed as non-Semitic, and such responses are referred to interchangeably throughout the rest of the paper as "Indo-European affixes", "Indo-European responses", or "affixes". Semitic responses were classified as responses where changes in the sequence of letters in the original word occurred, or where segments were deleted or added in the middle of the word in addition to or instead of adding segments to a word edge. For example, *tmajdi* and *tmejdu* were both given as responses to *tamdi*, which could be construed as having the root consonants $\sqrt{tmd}$, and were classed as Semitic. Such responses are referred to interchangeably throughout the rest of the paper as "Semitic patterns", "Semitic responses", or "patterns". Although participants were free to provide any type of diminutive response they chose, all participants used morphological diminutives except for a single data point, where the participant provided one lexical diminutive ("little wug") among the rest of their morphological diminutive responses. As there was only one such response in the entire dataset, it was treated as an anomaly and excluded from further analysis and discussion.

Certain of the allomorphs were more common than others throughout the data. For the Semitic patterns, *-ejje-* was the most common pattern that was attested in the grammar (Borg and Azzopardi-Alexander 1997), but other common responses from participants that were not attested in the grammar included *-ijji-*, *-ejC₃u-*, *-wejC₃(C₃)a-*, *-ajC₃u-*, and *-ijC₃a-* (see Table 3 below for sample responses). The corpus analysis in Sect. 3 above showed that *-ejje-* is a common pattern that is used with words that are not necessarily diminutive, but the frequency of this sequence may influence speakers to use it more often with novel words. *-(i)nu* and *-ina* were the most common Indo-European affixes used, and were attested in the grammar (Borg and Azzopardi-Alexander 1997). Due to the variety of morphemes provided even within individual participants, the only dependent variable considered in the statistical analysis was the response type: whether the response was a Semitic-type response or an Indo-European-type response.

Table 3 Non-exhaustive list of examples of common responses to nonsense words

Word/Response Type	ldir (Semitic)	girma (Semitic)	qarr (IE)	mirx (IE)
Semitic responses	ldrejjen	grejmu	qrejru	mrejxu
	ldejru	girmejja	qjarru	mrejxa
	ldwejjer	gwirma	qwejjar	mwirrex
	ldejjer	girmajna	qrwejjel	imrejxu
Indo-European responses	ldiru	girminu	qarrinu	mirxinu
	ldirina	girmina	qarra	mirxina
	ldirettu	girmettu	qarret	mirxett
	ldiret	girmita	qarrin	mirxet

Table 3 shows a selection of responses provided to four of the nonsense words that participants saw. When the words are given Semitic responses, the linear order of the consonants in the word is changed. For example, in *ldir* → *ldejjer*, the attested diminutive pattern *-ejje-* is inserted between the second and third root consonants, as would be expected from the examples provided by Borg and Azzopardi-Alexander (1997). In the Indo-European responses, the singular is retained in its entirety: for example, in *ldir* → *ldirina* the diminutive can be decomposed as *ldir* + *-ina*. In addition to the diminutive morphemes attested in the grammar and in the corpus analysis above, participants also used *-u*, *-ettu*, *-ett*, and *-in* in their concatenative responses, and tended to insert glides in their non-concatenative responses. All participants used at least one concatenative diminutive, and five participants did not use any non-concatenative diminutives.

4.1.5 Language background questionnaires

Because Maltese speakers are proficient in at least the two national languages, Maltese and English, and because bilingualism is suggested to have an effect on metalinguistic tasks, particularly when speakers are dominant in one language over the other (Galambos and Hakuta 1988; Campbell and Sais 1995; Gathercole 2007; Paradis 2010), the results are also analyzed for the impact of language dominance. While results from previous studies focusing on language dominance in psycholinguistic tasks are not always consistent, the primary hypothesis that falls out of the prior studies is that at minimum, bilingual speakers will be more sensitive to variables within the experiment than monolingual speakers would be, and speakers who are dominant in one language, their responses will mirror that language. That is, Maltese-dominant speakers should use more Semitic diminutives and English-dominant speakers should use more Indo-European diminutives.

Based on the data from the language background questionnaires, subjects were separated into four groups: Maltese dominant ($n = 18$), slightly Maltese dominant ($n = 10$), slightly English dominant ($n = 2$), and English dominant ($n = 3$). One speaker was a speaker of the Gozitan dialect of Maltese; as Gozitan is substantially different from the main dialect of Maltese (Borg 1996), this participant was excluded from the analyses.

Participants were classified into groups in accordance with the following guidelines. If subjects had their schooling in Maltese 80% or more of the time, then they were classed as Maltese dominant. If subjects had their schooling in Maltese at least 60% of the time, spoke to most people in Maltese, and watched television and read for fun in Maltese at least 50% of the time, they were also classed as Maltese dominant. If they had their schooling in Maltese at least 60% of the time, but spoke to most people in English, and watched television and read for fun less than 50% of the time in Maltese, then they were classed as slightly Maltese dominant.

If subjects had their schooling in English 80% or more of the time, then they were classed as English dominant. If subjects had their schooling in English at least 60% of the time, spoke to most people in English, and watched television and read for fun at least 50% of the time, they were also classed as English dominant. If they had their schooling in English at least 60% of the time, but spoke to most people in Maltese, and watched television and read for fun less than 50% of the time in English, then they were classed as slightly English dominant.

As the groups were quite small, participants were classified as only Maltese- or English-dominant for the purposes of statistical analyses.

4.2 Results

To ensure that no items were unduly driving the effect, the number of response types for word type (Fig. 2) and each item (Fig. 3) were calculated. Participant proportion of responses was calculated as well (Fig. 4). These graphs show that overall, participants responded using Indo-European suffixes, but were slightly more likely to use a Semitic pattern with a Semitic-sounding nonsense word.

Figure 3 shows the number of response types for each nonsense word. Overall the Semitic- and Indo-European-sounding words had similar numbers of Semitic- and Indo-European responses.

Figure 4 shows the number of response types each participant provided. Notice that not every participant provided 40 total responses, as they were allowed to skip items. Additionally, data from subject 7 was excluded as they were the Gozitan speaker.

The data were analyzed in R (version 3.2.4, R Core Team 2016) using binomial generalized linear mixed models in the lme4 package (Bates et al. 2015). The maximal model used binary independent variables word type (Semitic/non-Semitic) and language dominance (Maltese/English) as fixed effects. Subjects, items, and images were random effects, and fully expanded as would allow for convergence as justified by the design (Barr et al. 2013).[4] The maximal model was then compared to other models, eliminating the non-significant interaction between word type and language dominance ($p < 0.2$) and the non-significant main effect of language dominance ($p < 0.4$), resulting in the best fitting model showing an effect of only word type.

[4]The maximal model was computed as follows: *glmer(ResponseType ~ WordType * Dominance + (1 | subject) + (1 | item) + (1 | image), data = data)*. Stepwise model comparison was performed to arrive at the best fitting model, which is coded as the following: *glmer(ResponseType ~ WordType + (1 | subject) + (1 | item) + (1 | image), data = data)*. *p*-values were then computed using the anova function and comparing the best fitting model to a depleted model with WordType removed.

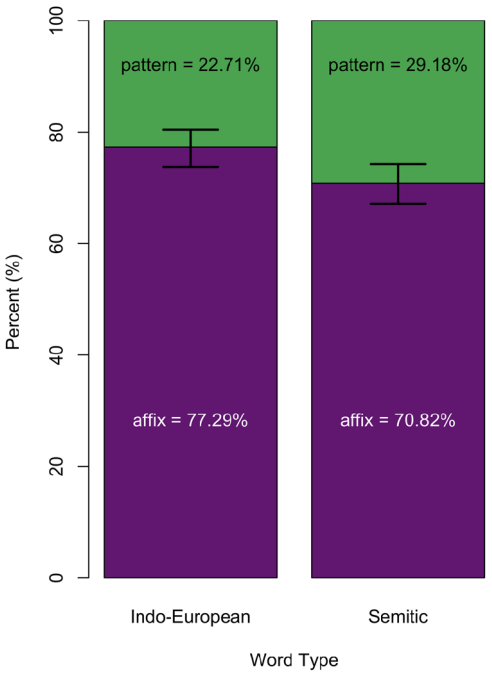

Fig. 2 Bar graph showing proportion of Indo-European and Semitic responses to each word type. Error bars indicate 95% confidence intervals

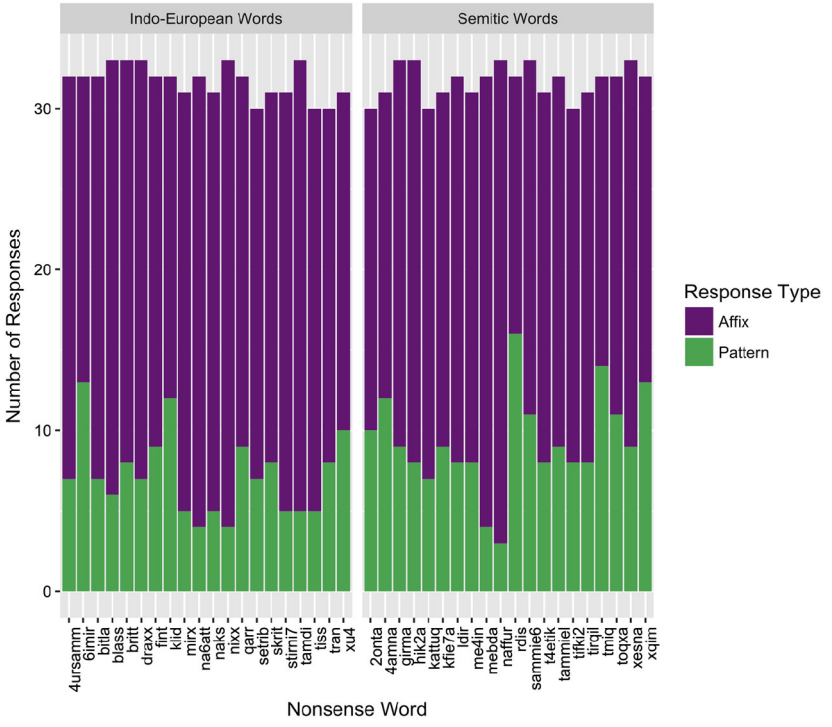

Fig. 3 Bar graph showing number of Indo-European and Semitic responses to each nonsense word

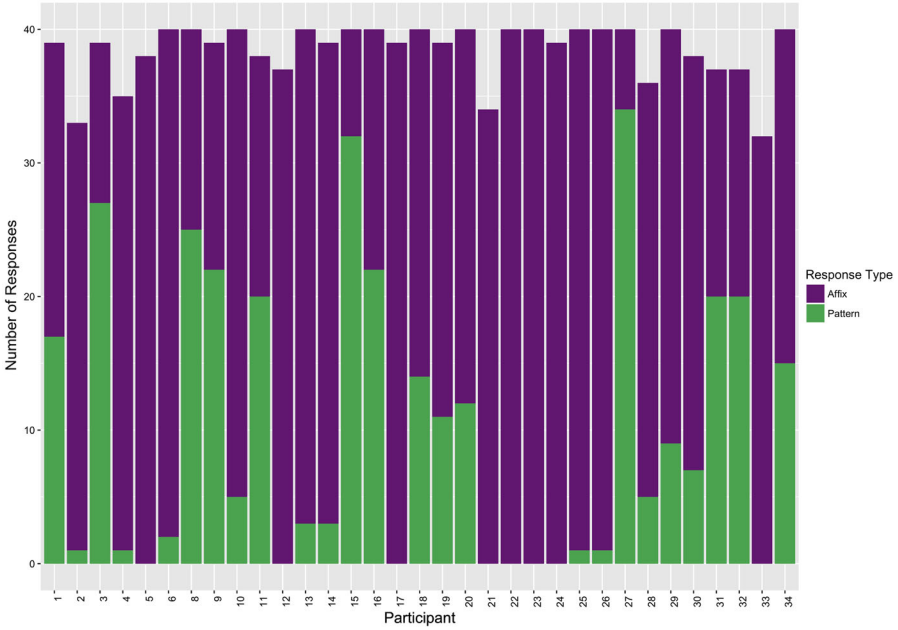

Fig. 4 Bar graph showing the number of Indo-European and Semitic responses provided by each participant

The analyses found a main effect of word type on participants' responses ($\beta = 0.55$, SE $= 0.23$, $\chi^2(1) = 5.13$, $p < 0.05$). This indicates that the type of word—whether the nonce word followed a Semitic or non-Semitic CV structure—had an effect on whether the participants responded using a Semitic pattern or an Indo-European affix. This suggests, following Twist's (2006) work, that participants are more likely to respond to a novel word that is similar to an existing Semitic word with a Semitic pattern, and are more likely to respond to a novel word similar to an existing Indo-European word with an Indo-European affix. A Hosmer-Lemeshow Goodness of Fit test (hoslem.test function in the ResourceSelection package; Lele et al. 2017) indicated that the observed values did not differ significantly from the expected values ($\chi^2(8) = 11.96$, $p < 0.2$). See Table 4 for a summary of the model values.

5 General discussion & conclusion

The given pattern of results suggests that (A) morphological decomposition must be accounted for in a model of the Maltese speaker's lexicon, and (B) some members of the Maltese-speaking population has access to two morphological systems in their grammar that are sensitive enough to produce morphology that matches a word's origin, even though that type of morphology is rarer and less productive. Speakers were also not constrained to use particular affixes or patterns, which led to considerable variation in the diminutive allomorphs produced in this task.

Table 4 Random effects, fixed effects, and goodness of fit measures from best-fitting model

Random effects:				
Groups	Variance	Std. Dev.		
Image	0.03415	0.1848		
Item	0.29042	0.5389		
Subject	5.56499	2.3590		
Fixed effects:				
	Estimate	Std. Error	Z value	Pr(> |z|)
Word type	0.5508	0.2324	2.370	0.0178
Goodness of Fit:				
AIC	BIC	log likelihood	deviance	
1036.3	1062.0	−513.2	1026.3	

Speakers having two morphological systems is consistent with previous research, both in forming denominal verbs (Twist 2006) and in forming verbal nouns (Spagnol 2011). Twist (2006) found that an Indo-European affix was used, even when the nonsense verb was modeled after words in the Semitic sublexicon. She also found that a small subset of subjects was more likely to respond with Semitic morphology when they were presented with a Semitic-sounding word and with Indo-European morphology when they were presented with an Indo-European-sounding word. These results are tempered by Spagnol's (2011) findings, where he finds that templatic verbs in Maltese are generally not productive; new verbs (borrowed or otherwise) tend not to be formed with templatic morphology, and templatic verbs are a closed class in Maltese. However, these verbs still fit into this split between the sublexicons, as verbal nouns derived from templatic verbs tend to be formed with root-and-pattern morphology, while those derived from concatenative verbs tend to be formed with a suffix. Templatic verbs are also typically Semitic in origin, while concatenative verbs are typically Indo-European in origin. A further reason that some subjects would be more likely to respond using templatic or concatenative morphology could be due to language dominance, based on psycholinguistic studies of metalinguistic awareness in bilinguals (Galambos and Hakuta 1988; Campbell and Sais 1995; Gathercole 2007; Paradis 2010), but no effect of language dominance was found in the present study.

Multiple models may account for the observed morphological variation. An analogical model (Krott et al. 2001) would allow for the degree of variation observed in the elicitation task in the number of allomorphs produced by the speakers. Under this model, the allomorphs used would be determined by the allomorph used with a similar existing word. That is, if an existing word used root-and-pattern morphology and sounded similar to a novel word, then a speaker would be more likely to use root-and-pattern morphology with the novel word rather than concatenative morphology. Forming words via analogy is supported in dual-route models examining past tense formation in English (e.g., Pinker and Prince 1988; Prasada and Pinker 1993), and is also intuitively comfortable. An analogical model is potentially more flexible than a rule-based system particularly when the use of an allomorph is inconsistent, seen in the results of this study.

Distributed Morphology (Halle and Marantz 1993; Harley and Noyer 1999) is also a viable framework to situate these results. Words requiring a diminutive would be marked with an underspecified feature, and filled in by the phonological mechanism at the level of Vocabulary Insertion. This would account for either the use of a morphological diminutive or a lexical diminutive, with additional constraints to determine which allomorph is the best fit. The results suggest that the Indo-European allomorph is the default or "Elsewhere" form, but a Semitic pattern may be triggered by contextual characteristics, such as phonological structure or CV skeleton (after Arad 2003). Following Bobaljik's (2000) framework, we may assume that Vocabulary Insertion is cyclic, starting from the bottom of the hierarchical structure and moving up. The root is at the bottom of the hierarchical structure and so gets inserted first, which allows its phonological characteristics to affect the selection of either the Indo-European or Semitic allomorphs further up the structure.

A final model that could account for these findings is a co-phonology analysis, after Inkelas and Zoll (2007). In this model, two grammars are hypothesized, each with a distinct ranking of the constraints relevant to determining the form of the output. For speakers who use both Semitic and Indo-European morphology, the grammar used to form each diminutive would have a different constraint ranking dependent on whether an input item was classed as Semitic or non-Semitic. This dual grammar results in a more Semitic-sounding diminutive form for words that are similar to the Semitic sublexicon, and an Indo-European-sounding diminutive form for words similar to the Indo-European sublexicon. However, to accurately predict the speakers' performance, having just two constraint rankings may not be enough due to extensive within-speaker variation (for example, one speaker provided *-ina*, *-inu*, and *-ette*, in addition to others as reasonable Indo-European forms for different stimuli).

An alternative is that this kind of co-phonology is implemented via class feature. Both the Indo-European suffix and the Semitic pattern would be associated with the feature [+diminutive], but the Semitic pattern would require an additional trigger from the speaker's grammar. Individual vocabulary items or encyclopedia entries may have a [+Semitic] feature that must be matched by a morphological derivation, which would lead to speakers matching that feature if it was included in their grammar. Other triggers may include the phonological representation or the CV structure of the root.

The spirit of these analyses is similar; namely, the input is received by a system, which selects allomorphs within certain parameters in order to best fit the phonology of a root word along with any semantic or syntactic features required by the context. If a Semitic allomorph were present in a speaker's grammar and was phonologically appropriate for the root word, the system would select it. Otherwise, either in an "Elsewhere" case or in the case of a speaker only having Indo-European allomorphs in their grammar, the default affix would be applied to the word. In the case of having multiple appropriate allomorphs in one sublexicon (exemplified by the speaker who used *-ina*, *-inu*, and *-ette* for Indo-European suffixes above), a mechanism that selects from the appropriate allomorph from the speaker's grammar within the relevant sublexicon is necessary. This mechanism may be mediated by a co-phonology hierarchy (after Inkelas and Zoll 2007), a feature geometry (Harley 1994), or a Universal Hierarchy of Features (Noyer 1992), depending on which theoretical framework is pursued.

As this split between Semitic and Indo-European allomorphs may be induced by nonsense words as well as real words of Maltese, this points to the sets of allomorphs not being lexicalized. Theories that can account for this phenomenon must therefore be able to take phonotactics into consideration and cannot be blind to the phonological makeup of a word prior to selecting the appropriate allomorph. Phonotactic constraints are considered in the probabilistic rule-based frameworks (Pierrehumbert 2001; Albright and Hayes 2003) and are thus promising options for differentiating between Semitic and Indo-European allomorphs based either on the CV skeleton of a given word, real or otherwise, and also provide an elsewhere condition for speakers who are either not sensitive to a difference in the phonotactics or origin of a word, or for words with marginally informative phonotactics where either a Semitic or Indo-European affix is plausible.

Stochastic models use the probability of segment sequences to predict the likelihood of a given affix. Using this type of model, featural similarity of segments can be preserved, Obligatory Contour Principle (OCP) violations will be minimized, and a selection of affixes can be applied based on a variety of fine-grained rules (Pierrehumbert 2001). These models perform more accurately for novel words in English than purely analogical models (Albright and Hayes 2003), and would ostensibly be able to select among the correct possible vowels in the Maltese Indo-European affixes. The preservation of featural similarity in individual segments and OCP constraints is particularly important, as Maltese borrowings from Italian that violate the OCP for the Semitic sub-lexicon will not be incorporated and will take Indo-European morphology. Borrowings from Italian that conform to the OCP will be assimilated into the Semitic sub-lexicon and will have templatic morphology available to them despite not originating from Semitic (Frisch et al. 2004).

The benefit of these models is that, if Maltese data mimics the English data from Albright and Hayes (2003), Indo-European suffixation can be considered the default diminutive morphological system, and the multiple rules will be able to generate the other forms dependent on the similarity of each word to either the Indo-European or Semitic sublexicon. However, it is unclear exactly how the models would be able to account for the non-concatenative morphology that is provided by the participants. It is possible that the variegated similarity used by analogical models, discarded by Albright and Hayes, may be better able to account for the use of non-concatenative templates. Any theory used to describe Maltese morphology must be able to take the variability of affixes in both sublexicons into consideration as well as frequency effects of both the base and the affix or template (e.g., Hay and Baayen 2002) that also affects this type of morphological variation (Hay and Baayen 2005; Krott et al. 2001).

To form a more complete picture of the morphological system in Maltese, further investigations in the mental representations of morphemes is necessary, particularly with regards to the dual lexicon and dual grammar. For example, further examination of phonological cues that speakers use to select between morphological systems would elucidate triggers for using one system or the other. Research involving the prevalence, form, and productivity of Maltese broken plurals or other types of evaluative morphology would help to expand the literature overall with potential ties to previous work on Arabic broken plurals (e.g., McCarthy and Prince 1990;

Hammond 1988; Dawdy-Hesterberg and Pierrehumbert 2014). This type of work would also expand our knowledge of the productivity of Semitic morphology in Maltese, where results are currently mixed. Research on Maltese with its dual lexicon, dual morphological system, and variable productivity within its morphological systems would be beneficial not only for theories of morphology but also to theories of language processing. Given that Maltese has two productive morphological systems, albeit systems that vary in terms of their productivity within the population, this poses a potential problem for theories of morphology that do not account for possible phonological effects or effects of language dominance affecting the processing of language input.

Acknowledgements This work was funded by grants from the University of Arizona Social and Behavioral Sciences Research Institute and the University of Arizona Graduate and Professional Students Council. The author wishes to thank the following individuals for their comments and expertise: Lauren M. Ackerman, William Cotter, Ray Fabri, Kenneth I. Forster, Luke Galea, Heidi Harley, Ingo Plag, Adam Ussishkin, Andrew Wedel, Samantha Wray, two anonymous reviewers, and the Institute of Linguistics at the University of Malta.

Appendix: Stimuli used in the novel word elicitation task

Practice nonsense words:
- brieq (Semitic)
- korfa (Indo-European)

Practice real words:
- ġobon / ġbejna (‘cheese / cheeselet’, Semitic)
- festa / festina (‘party / little party’, Indo-European)
- kuċċarun / kuċċarina (‘ladle / teaspoon’, Indo-European)
- triq / trejqet (‘street / alley’, Semitic)

Semitic nonsense words:
- sammieġ
- hikża
- tifkiż
- tmiq
- xesna
- mebda
- ldir
- toqxa
- tħetik
- tirqil
- kfieċa
- girma
- żonta
- rdis
- tammiel
- kattuq

naffur
xqim
ħamna
meħin

Indo-European nonsense words:

xuħ
britt
naks
mirx
stirnić
ġimir
nixx
bitla
tamdi
setrib
tiss
qarr
draxx
ħursamm
fint
tran
klid
blass
skrit
naġatt

References

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, *90*, 119–161.

Aquilina, J., & Cassar-Pullicino, J. (1957). *Lexical material in Maltese folklore*. Msida: Malta University Press.

Arad, M. (2003). Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language and Linguistic Theory*, *21*(4), 737–778.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge: MIT Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bobaljik, J. D. (2000). The ins and outs of contextual allomorphy. In K. K. Grohmann & C. Struijke (Eds.), *University of Maryland working papers in linguistics* (Vol. 10, pp. 35–71).

Bolozky, S. (1994). On the formation of diminutives in modern Hebrew morphology. *Hebrew Studies*, *35*, 47–63.

Borg, A. (1996). On some Mediterranean influences on the lexicon of Maltese. In J. Lüdke (Ed.), *Romania Arabica* (pp. 129–150). Tübingen: Narr.

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

Borg, C., Fabri, R., Gatt, A., & Rosner, M. (2011). *MLRS Corpus, v2. 0 BETA*. http://mlrs.research.um.edu.mt/corpusquery/malti02/.

Boudelaa, S., & Marslen-Wilson, W. D. (2004b). Allomorphic variation in Arabic: implications for lexical processing and representation. *Brain and Language*, *90*, 106–116.
Boudelaa, S., & Marslen-Wilson, W. D. (2011). Productivity and priming: morphemic decomposition in Arabic. *Language and Cognitive Processes*, *26*(4), 624–652.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.
Campbell, R., & Sais, E. (1995). Accelerated metalinguistics (phonological) awareness in bilingual children. *British Journal of Developmental Psychology*, *13*(1), 61–68.
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.
de Jong, N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, *15*(4/5), 329–365.
Deutsch, A., Frost, R., & Forster, K. I. (1998). Verbs and nouns are accessed differently in the mental lexicon: evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1238–1255.
Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, *22*(1), 179–228.
Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 829–856.
Frost, R., Deutsch, A., & Forster, K. I. (2000). Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology; Learning, Memory, and Cognition*, *26*(3), 751–765.
Galambos, S. J., & Hakuta, K. (1988). Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children. *Applied Psycholinguistics*, *9*, 141–162.
Gathercole, V. C. M. (2007). Miami and North Wales, so far and yet so near: a constructivist account of morphosyntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism*, *10*(3), 224–247.
Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from the building 20* (pp. 111–176). Cambridge: MIT Press.
Hammond, M. (1988). Templatic transfer in Arabic broken plurals. *Natural Language and Linguistic Theory*, *6*, 247–270.
Harley, H. (1994). Hug a tree: deriving the morphosyntactic feature hierarchy. In A. Carnie & H. Harley (Eds.), *MIT working papers in linguistics* (Vol. 21, pp. 289–320). Cambridge: MIT Working Papers in Linguistics.
Harley, H., & Noyer, R. R. (1999). Distributed morphology. *Glot International*, *4*(4), 3–9.
Hay, J., & Baayen, R. H. (2002). Parsing and productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology 2001* (pp. 203–235). Dordrecht: Kluwer Academic Publishers.
Hay, J., & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Science*, *9*(7), 342–348.
Inkelas, S., & Zoll, C. (2007). Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics*, *45*(1), 133–171.
Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language*, *72*(3), 533–578.
Kempe, V., Brooks, P. J., Gillis, S., & Samson, G. (2007). Diminutives facilitate word segmentation in natural speech: cross-linguistic evidence. *Memory & Cognition*, *35*(4), 763–773.
Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36. Abstract.
Krott, A., Baayen, R. H., & Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, *39*(1), 51–93.
Lele, S. R., Keim, J. L., & Solymos, P. (2017). Version 0.3-3 [Software]. Resource selection (probability) functions for use-availability data. https://github.com/psolymos/ResourceSelection.
MATLAB (Version 2013B) [Software], Natick, MA: The MathWorks, Inc.
McCarthy, J. J., & Prince, A. S. (1990). Foot and word in prosodic morphology: the Arabic broken plural. *Natural Language & Linguistic Theory*, *8*, 209–283.
Mifsud, M. (1995). *Loan verbs in Maltese: A descriptive and comparative study*. Leiden: Brill.
Noyer, R. R. (1992). *Features, positions and affixes in autonomous morphological structure*. Ph.D. dissertation, Massachusetts Institute of Technology.

Ohala, D. (1999). The influence of sonority on children's cluster reductions. *Journal of Communication Disorders*, *32*, 397–422.
Paradis, J. (2010). Bilingual children's acquisition of English verb morphology: effects of language exposure, structure complexity, and task type. *Language Learning*, *60*(3), 651–680.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Pierrehumbert, J. B. (2001). Stochastic phonology. *Glot International*, *5*(6), 195–207.
Pinker, S., & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193.
Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*(1), 1–56.
R Core Team (2016). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.
Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review*, *11*(6), 1090–1098.
Schneider, K. P. (2003). *Diminutives in English*. Tübingen: Max Niemeyer Verlag GmbH.
Spagnol, M. (2011). *A tale of two morphologies: verb structure and argument alternations in Maltese*. Ph.D. dissertation, University of Konstanz.
Twist, A. E. (2006). *A Psycholinguistic investigation of the verbal morphology of Maltese*. Ph.D. dissertation, University of Arizona.
Vella, I. (2015). English Maltese dictionary. Retrieved from http://www.englishmaltesedictionary.com.
Watson, J. C. E. (2006). Arabic morphology: diminutive verbs and diminutive nouns in San'ani Arabic. *Morphology*, *16*, 189–204. https://doi.org/10.1007/s11525-006-9103-5.