Learnability and generalisation of Arabic broken plural nouns

Lisa Garnand Dawdy-Hesterberg & Janet Breckenridge Pierrehumbert

Published online: 18 Mar 2014.

Submit your article to this journal 🗗

Article views: 2369

View related articles 🗗

View Crossmark data 🗗

Citing articles: 8 View citing articles 🗗

Routledge
Taylor & Francis Group

Learnability and generalisation of Arabic broken plural nouns

Lisa Garnand Dawdy-Hesterberg* and Janet Breckenridge Pierrehumbert

Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60202-0854, USA

The noun plural system in Modern Standard Arabic lies at a nexus of critical issues in morphological learnability. The suffixing "sound" plural competes with as many as 31 non-concatenative "broken" plural patterns. Our computational analysis of singular–plural pairs in the Corpus of Contemporary Arabic explores what types of linguistic information are statistically relevant to morphological generalisation for this highly complex system. We show that an analogical approach with the generalised context model is highly successful in predicting the plural form for any given singular form. This model proves to be robust to variation, as evidenced by its stability across 10 rounds of cross-validation. The predictive power is carried almost entirely by the CV template, a representation which specifies a segment's status as a consonant or vowel only, providing further support for the abstraction of prosodic templates in the Arabic morphological system as proposed by McCarthy and Prince.

The noun plural in Modern Standard Arabic (henceforth Arabic) provides an interesting challenge for the theory of morphological learnability. There are a large number of possible plural patterns. In addition to two regular patterns (the so-called "sound" plurals that are formed by suffixation), there are as many as 31 productive irregular patterns (McCarthy & Prince, 1990a; Wright, 1896/1988). These are called "broken" patterns because they are formed non-concatenatively by interleaving, and sometimes modifying, morphemes that do not appear in isolation in base forms. The broken patterns comprise an unusually high percentage of plurals, with as many as 41% of noun types having a broken plural (Boudelaa & Gaskell, 2002). Predicting the pluralisation of new forms is difficult, not only because so many patterns exist, but also because the relevant cognitive dimensions of pluralisation are not fully understood. Despite many attempts to discover underlying principles in the Arabic plural system, no fully predictive account exists. McCarthy and Prince's (1990a) well-known analysis of the broken plural cites prosodic structure as the main determinant of which specific broken pattern is selected. However, as we will discuss, the approach is unable to predict whether many forms will take broken or sound plurals.

The complexity of the Arabic noun plural system makes it difficult to learn. Although adult speakers can achieve perfect accuracy on a pluralisation task (Albirini & Benmamoun, 2012), Ravid and Farah (1999) found that children learning Palestinian Arabic produce many incorrect plurals at age 6, with many errors in broken pluralisation. In a study of the acquisition of Egyptian Arabic, Omar (1973) found that speakers as old as 15 made errors on common broken plurals. In comparison, children learning languages with concatenative morphology generally acquire number inflexion of nouns by age 6 (Berman, 1981; Clahsen, Rothweiler, Woest, & Marcus, 1992). The noun plural systems of Palestinian and Egyptian Arabic are minimally different from that of Modern Standard Arabic, so learnability of these systems is comparable.

These properties of the Arabic noun plural system allow us to examine a number of issues related to the learnability of complex morphological systems. Previous research has yielded important insights into the Arabic noun plural system, but these investigations have not quantitatively compared the relevance of different features to pattern learnability. Pattern learning is, at its core, generalisation across forms. Once a speaker has acquired a sufficient number of forms with a given pattern, the speaker can abstract across the shared features of those forms and generalise to new forms that share those features (e.g. Albright, 2009; Pierrehumbert, 2001). In this paper, we will examine learnability through the lens of generalisation and investigate the factors that support generalisation for the Arabic noun plural.

It is well established that morphological generalisation can occur via analogy. The similarity of a base form to one or more stored words determines the pattern for a novel-derived form (Bybee, 1985; Hay & Baayen, 2005; Rumelhart & McClelland, 1986). In general, both semantic and phonological dimensions of similarity are relevant to analogical processes. However, in the present

*Corresponding author. Email: lisah@u.northwestern.edu

study we focus on the phonological dimensions and omit examination of the semantic dimensions. Two semantic factors (gender and animacy) are already well documented as the factors determining the choice between the [-uun] and [-aat] sound plurals. Several studies have also suggested that transparently derived forms are more likely to take the sound plural, while more semantically opaque-derived forms are more likely to take a broken plural (Boudelaa & Gaskell, 2002; McCarthy & Prince, 1990a). Such findings are consistent with the theory that morphological structure is a gradient reflex of semantic and phonological similarities amongst words, and that this gradience is in turn reflected in statistical variation in the productivity of morphological patterns (Bybee, 1985; Haspelmath & Sims, 2010; Hay & Baayen, 2005). However, there are major challenges to implementing this approach for a corpus study of Arabic, as there is no generally agreed method for measuring semantic transparency or for assigning gradient structure to non-concatenative lexical derivations. Therefore, we have focused on quantifiable phonological factors.

In studies of concatenative morphology, the number of different stored lexical items exhibiting a particular pattern has been shown to affect the propensity of the pattern to be generalised, a finding referred to as the gang size effect (McClelland & Elman, 1986; Rumelhart & McClelland, 1982; Stemberger & MacWhinney, 1988). Each additional stored form adds statistical support for the pattern, so larger gangs generalise to new forms more easily. The best way to define lexical gangs for Arabic morphology is one of the questions that we take up.

Another question we address is whether non-concatenative morphological processes, in which morphological changes occur non-locally, establish different dimensions of similarity than concatenative morphological processes, in which morphemes are combined sequentially. One reason to expect such differences is the known importance of the CV template to many morphological processes in Arabic, including diminutivisation and verb measure derivation (McCarthy, 1981, 1993; McCarthy & Prince, 1990a). The CV template is the skeletal structure of the word that abstracts away from the segmental features of phonemes, retaining only their status as consonants or vowels. It is proposed as a separate representational entity in order to capture patterns of morphological generalisation (McCarthy, 1981, 1982, 1993; McCarthy & Prince 1990a). Its existence is corroborated by evidence from psycholinguistic experiments on word priming (Boudelaa & Marslen-Wilson, 2004).

In this paper, we will examine three factors that may influence morphological generalisation in Arabic: the CV template, segmental similarity and gang size. Using a comparison of several predictive analogical models, we investigate the importance of these three factors in pluralisation of the Arabic noun, gaining insight

into the mechanisms of morphological generalisation in Arabic, as well as into the overall learnability of this system. First, we will examine the Arabic noun plural system and the issues that work in this area have raised. We find support for Boudelaa and Gaskell's (2002) claim that the noun plural is not a minority default system, as the default sound plural constitutes 74% of forms by type in our corpus study. Next, we discuss predictive analogical models of morphological generalisation and define the five variations of the generalised context model (GCM; Nakisa, Plunkett & Hahn, 2001; Nosofsky, 1990) that we implement. Comparison of model results shows that the CV template and gang size carry most of the statistical power in plural selection, and that fine-grained segmental similarity makes a small additional contribution. These results indicate that the combination of coarse-grained abstraction and statistical knowledge drives generalisation in Arabic morphology.

Key terminology
CV template

The **CV template**, also referred to as the **CV skeleton**, is defined in traditional accounts of Arabic morphology as the skeletal base of the word onto which the segmental tiers of the word are associated, and contains only the units *C* and *V*, which represent consonants and vowels, respectively (Hammond, 1988; McCarthy, 1981, 1982, 1993; McCarthy & Prince, 1990a). The CV template contains one additional unit, *T*, which represents the feminine grammatical marker *taa marbuta* ("bound t"). It is generally elided in speech, but is included overtly in the CV template in this account because forms with this suffix behave differently in pluralisation than their unsuffixed counterparts (McCarthy & Prince, 1990a; Ratcliffe, 1998). Additionally, it is coded differently from the other consonants (including other instances of the segment *t*) because it has variable behaviour in relation to syllable weight, which is specific to this particular gender marker. The CV template contains no additional information on the segmental features of a form. This is in contrast to the **pattern**, which refers to the semi-filled CV template without a specific verbal root, but with the other morphemic tiers specified (McCarthy, 1981, 1985). For example, the pattern of [madrasaT] ("school") is [maC₁C₂aC₃aT], while the CV template is [CvCCvCvT] (see Figure 2).

Prosodic pattern

The prosodic pattern of the word is the moraic structure of the underlying word stem as defined in McCarthy and Prince (1990a). The prosodic patterns of the Arabic broken plural can be divided into four classes: iambic (CvCvv+), trochaic (CvCvC), monosyllabic (CvCC) and other (CvCCvC and CvCCvvC; McCarthy & Prince,
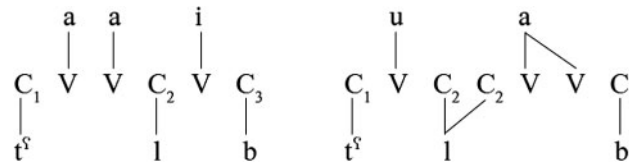
Figure 1. Tier structures of [tˤaalib] ("student") and plural [tˤullaab] ("students").

1990a). The stem can undergo initial [Ca] metathesis or suffixation, which results in a variety of surface CV templates within each class. The monosyllabic and other classes are very limited, while the iambic and trochaic classes have greater generality. In general, the iambic and trochaic classes can be distinguished by stress patterns: the trochaic stem has initial stress, while the iambic stem has non-initial stress.

Segmental similarity

The phonological composition of the word is the basis for similarity in this work. Specifically, we concentrate on similarity at the segmental level based on shared natural classes (Frisch, Pierrehumbert, & Broe, 2004) aggregated over all the segments in the word. We calculate similarity using string-edit (Levenshtein) distance (Kruskal, 1983; Levenshtein, 1966), where the optimal alignment is the alignment with the smallest number of changes between the first comparison form and the second, weighted by segmental similarity. The open question for this research is which dimensions of similarity matter to the Arabic plural system.

Analogy

Many theories of morphological generalisation have posited that generalisation to new forms can occur by analogy, wherein the new form selects a morphological pattern based on the patterns of one or more similar stored forms (e.g. Prasada & Pinker, 1993; Rumelhart & McClelland, 1986; Skousen, 1989, 1993). The number of stored forms necessary to create a successful analogy is a matter of active research.

Lexical gang effects

Current research uses the concept of a lexical gang as the support for analogy. The lexical gang is a concept from psycholinguistics, where it is defined as a group of highly similar forms where similarity is based on shared structural and segmental features (e.g. Alegre & Gordon, 1999; Daelemans, Gillis, & Durieaux, 1994; Ernestus & Baayen, 2003). Gang size is an important factor in morphological generalisation; it is easier to generalise from a larger gang, as more examples mean more support for the generalisation (e.g. Rumelhart & McClelland, 1982; Stemberger & MacWhinney, 1988). We apply this concept to Arabic,

where we define a lexical gang by the CV template of the base and inflected forms, such that word pairs which have the same singular CV template and the same plural CV template constitute a gang. The current study evaluates the roles of segmental and structural information along with gang size in morphological analogy in Arabic.

The Arabic noun plural system

The Arabic morphological system contains a number of productive non-concatenative morphological processes. This system is frequently studied because of its complexity and its distinctive use of non-local information in morphological patterning. At the surface level, a single morphological process applies to non-adjacent parts of a word, as in Figure 1.

A word formed via a non-concatenative process is comprised of interleaved morphological tiers, which are abstract morphemes that do not occur in isolation in base forms. Under the framework of prosodic morphology, there are four tiers of a word in Arabic: the verbal root, the CV template, the *x*-morpheme ([m] in Figure 2) and the vocalic melody (McCarthy, 1981, 1982). The prototypical verbal root is three consonants and carries semantic information shared by the morphological family. The CV template specifies the prosodic structure of the word and is the base to which the other tiers attach. The *x*-morpheme tier contains all consonants other than the verbal root consonants, with [t] and [m] most common due to a number of productive verbal and nominal patterns. The vocalic melody specifies the V slots in the CV template and exhibits rightward spreading (McCarthy, 1982). The pattern, consisting of the CV template plus specified vowels and *x*-morphemes, holds the majority of the morphological information, including word type, gender and animacy (see Prunet, 2006 for a review).

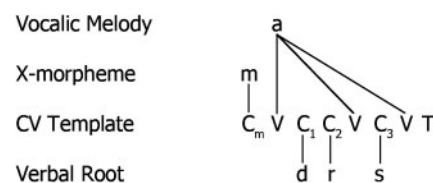The noun plural system in Arabic is frequently cited for its complex behaviour because of the large number of

Figure 2. Tier structure of the Arabic word [madrasaT] ("school").

plural patterns and the high percentage of irregular forms (Levy, 1971; McCarthy & Prince, 1990a; Murtonen, 1964). There are two sound plural suffixes, which attach to semantically distinct groups: [-uun] (which attaches to human masculine nouns) and [-aat] (which attaches to human feminine and non-human nouns). The sound plural is morphologically regular, and [-aat] deletes the stem-final feminine grammatical marker *taa marbuta* prior to suffixation.

The broken plural constitutes as many as 31 different patterns, as well as additional rare and non-productive patterns (McCarthy & Prince, 1990a; Wright, 1896/1988). The singular noun undergoes significant internal structural change in broken pluralisation due to the change of CV template. The surface form undergoes a number of complex changes, which can include consonant gemination (e.g. [tˤaalib]→[tˤullaab], "student"→"students"; see Figure 1), vowel lengthening (e.g. [rajul]→[rijaal], "man"→"men"), vowel shortening (e.g. [kitaab]→[kutub], "book"→"books"), glide deletion/substitution (e.g. [ʕaduww]→[ʔaʕdaaʔ], "enemy"→"enemies"), glide insertion and metathesis (e.g. [baab]→[ʔabwaab], "door"→"doors").

Previous accounts

In this paper, we will examine possible sources of information for pluralisation in Arabic to determine the factors that influence generalisation. Using traditional methods, Hammond (1988), McCarthy (1981, 1993) and McCarthy and Prince (1990a) concluded that the CV template of the singular noun was the main factor in pluralisation. The authors claim that it determines whether a form is eligible for broken pluralisation. If so, it also partially determines the CV template of the broken plural form (see Figure 10 and Appendix for further discussion). This approach captures a subset of the plural system extremely well, with strong predictive power for forms taking the iambic broken plural.

However, the CV template is by definition a coarse-grained representation. In many other languages, finer-grained segmental similarities play a significant role in morphological generalisation (e.g. Alegre & Gordon, 1999; Ernestus & Baayen, 2003). That is, a form undergoing inflexion or derivation will pattern like forms to which it is close in a similarity space that includes segmental features. A crude measure of the overall segmental similarity of two words is the number of segments they share (e.g. Derwing & Skousen, 1994; Skousen, 1989, 1993). However, this measure neglects the fact that some pairs of segments are more similar to each other than others. For example, /s/ is more similar to /z/ than to /l/. In their study of consonant co-occurrence restrictions in Arabic, Frisch et al. (2004) developed a gradient measure of similarity between two segments that takes into account the distinctive features of the segments

as well as the correlations amongst the features. This measure (the ratio of the natural classes shared by the segments to the number of shared plus non-shared natural classes) is validated by Frisch and Zawaydeh's (2001) experimental study on nonce verbs in Jordanian Arabic and is also adopted for Albright and Hayes' (2003) investigation of past tense formation in English. The role of fine-grained segmental similarities in noun pluralisation in Arabic has been pointed to in some of the literature. Specifically, the features of the vocalic melody of the singular and the presence of "weak" consonants ([ʔ], [w], [j]) in the stem appear to be correlated with the pattern of the plural (Levy, 1971; Ratcliffe, 1998). In this paper, our ability to examine these factors is limited by the use of unvoweled text from a corpus source. Therefore, we only explore the extent to which fine-grained segmental similarities overall add additional predictive power in a model of the Arabic noun plural system. We find that the statistical utility of segmental similarities in non-concatenative morphology is much lower than that reported in similar studies of concatenative morphology.

Gang size has also been shown to play a role in analogical processes. The applicability of an analogy depends not only on the similarity of the gang members to the novel item, but also on the number of members in the gang (Alegre & Gordon, 1999; Bybee & Moder, 1983; Prasada & Pinker, 1993). The open question in this area is the number of gang members necessary to successfully form an analogy. Bybee (1995) claims that single forms cannot form analogies productively, while other accounts claim that analogy can occur by comparison to a single form (Bauer, 2001; McCarthy, 1982). If analogy can occur to a single stored form, distance in similarity space should be the primary factor in forming a successful analogy. However, if analogy requires multiple forms to be productive, then the size of the gang will interact with distance in similarity space, such that a gang may win out over a singleton if both are equally similar to a new form undergoing analogy. The specific parameters of the tradeoff between gang size and similarity are a matter of debate, and we will return to this issue later.

Two previous studies have examined predictive computational models of the Arabic noun plural system that incorporate these fine-grained measures. Plunkett and Nakisa (1997) employed a multi-layered connectionist network to classify a singular form as one of the 12 broken plural types or one of the 2 sound plural types. The data-set consisted of 859 singular–plural pairs from the Hans Wehr (1976) dictionary. The broken plural accounted for 76% of the data-set, with broken patterns ranging from 1.5% to 17.5% of the data-set. The singular forms were represented as vectors containing 16 segmental features per phoneme, which were then mapped onto a left-justified VCVCVCVCVCVCV template. Unused slots were represented as an empty vector, for a total

vector of 208 features per word. The network was trained on the singular forms, using the 208-feature vectors as input units and plural type as output units. After 1000 epochs of training on the entire data-set, the classifier was able to learn 93.3% of trained forms. The authors then examined generalisation to unseen forms by training the network on 90% of the data-set and using the remaining 10% as a held-out test set. For unseen forms, the network correctly classified 63.8% of forms. Importantly, the network trained on the full data-set was able to learn both trochaic and iambic broken plural patterns with good accuracy, which suggests that both of these prosodic patterns should be productive (see Figure 11 and Appendix for further discussion). The much lower accuracy for unseen forms suggests that the plural system is learnable, but that generalisation to new forms may be a more difficult task.

Nakisa et al. (2001) employed the connectionist network from Plunkett and Nakisa (1997) in addition to a *k*-nearest-neighbours model and the GCM (Nosofsky, 1990) to compare single- and dual-route approaches. The single-route models used only the classifier in question, while the dual-route models also had a rule-based module that was triggered whenever the first module failed to reach a threshold of similarity to the test form. The data-set was much larger than in the previous study (*n* = 4771 singular–plural pairs), but contained only 11 broken plural patterns. The broken plural constituted 73.6% of the data-set. The single-route models achieved higher accuracy across the board on unseen forms (see Table 1), and the single-route GCM and connectionist network were equally accurate.

Although these studies achieved fairly good accuracy, there are empirical problems that need to be addressed. First, the data-sets represent unrealistic distributions of the sound and broken plurals, as both data-sets contain a minority of sound plurals. Although this system is frequently cited as a minority default system, (e.g. Levy, 1971; Murtonen, 1964; Wright, 1896/1988), our analysis of 6597 singular–plural noun pairs from the Corpus of Contemporary Arabic (CCA; Al-Sulaiti, 2009) finds that the sound plural is statistically dominant by type (74% of word types; see Figure 3), where a given word counts as a single type regardless of how frequently it appears in the corpus. Type count is generally viewed as most relevant to morphological productivity (e.g. Albright, 2009; Baayen & Lieber, 1991; Bybee, 1995), but we find that the
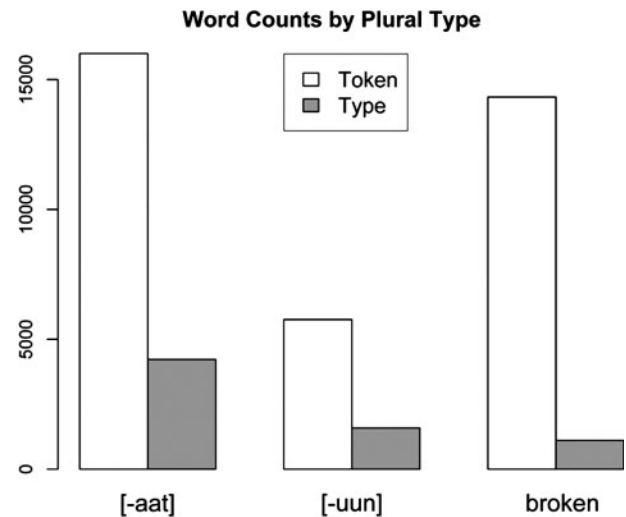
Figure 3.   Type and toke counts of sound ([-aat] and [-uun]) and broken plurals in the CCA.

dominance of the sound plural holds true even by token (61% of word tokens), where each instance of a word in the corpus counts as a token. Our finding is in line with Boudelaa and Gaskell's (2002) analysis of 1670 high-frequency nouns from the Basic Lexicon of Modern Standard Arabic (Khouloughli, 1992), which found that 59% of singular noun types take a sound plural. To the best of our knowledge, all previous support for Arabic as a minority default system has come from dictionary sources, rather than corpus data. Using dictionaries as a data source results in a data-set that is not representative of the language in actual use. For instance, dictionaries are often not up to date for colloquial usage, such as loanwords and lexical innovations. They may also contain highly specialised words and archaic forms and omit transparently derived regular forms. To address this problem, we only used forms that occurred at least once in the corpus (token frequency ≥ 1). Because of the size and corpus source of both Boudelaa and Gaskell's (2002) and our data-sets, this is strong evidence that the Arabic plural is not a minority default system. Future analyses of the Arabic noun plural must represent the sound plural accordingly.

A second empirical issue is that both studies addressed only a subset of broken plural types. Wright (1896/1988) cites 31 common broken plural patterns, but Plunkett and Nakisa (1997) examined only 12 broken plural patterns, while Nakisa et al. (2001) examined 11 patterns. In order to capture morphological generalisation for the entire system, we must examine a larger set of plural types.

A third shortcoming of these studies, while not a design flaw, is that they give little insight into the factors that govern pluralisation in Arabic. Specifically, we wish to know the extent to which fine-grained measures of similarity and gang size influence morphological generalisation in Arabic. Both of these factors play important

Table 1. Model accuracy in Nakisa et al. (2001).

	Single-route	Dual-route
k-nearest-neighbours	56.3	54.0
Generalised Context Model	70.8	57.6
Connectionist network	70.9	63.7

roles in morphological generalisation in other languages, but have not been systematically examined for Arabic.

In this study, we employ the GCM (Nosofsky, 1990) as implemented by Nakisa et al. (2001) and Albright and Hayes (2003) and address these issues with the following innovations: (1) the data-set is collected from a corpus and represents a more realistic distribution of broken to sound plurals, with 28.8% broken plurals; (2) the data-set encompasses 37 broken plural types, which includes multiple types from each of McCarthy and Prince's (1990a) four prosodic patterns and (3) we use a comparison of five variations of the GCM to determine the importance of three specific factors in plural selection.

Methodology

Data collection

The full data-set used in these simulations consists of 1945 singular–plural pairs collected from the CCA (Al-Sulaiti, 2009), which contains approximately 840,000 words from a variety of print sources. The sound singular–plural pairs were collected directly from the corpus. The broken singular–plural pairs were compiled by Mohamed (2009), and any forms that did not occur at least once in the CCA were excluded (frequency of <1.18 per million). The final data-set contains 1384 sound and 561 broken singular–plural pairs.

Each plural type (sound [-aat], sound [-uun], and broken) was separated into gangs based on singular–plural CV structures, such that a gang is a group of singulars with the same CV template that take the same plural CV template. For this study, a gang is defined as at least four forms that pattern together, as the protocol used in cross-validation (75% training set and 25% test set) requires at least four items in each set. These pairs form a total of 108 gangs, with 55 taking the [-aat] sound plural, 16 taking the [-uun] sound plural and 37 taking the broken plural. For the [-aat] sound plural, the largest gang contains 98 forms. For the [-uun] sound plural, the largest gang contains 30 forms. For the broken plural, the largest gang contains 70 forms. This gang has the iambic pattern [CVCC]->[CVCVVC] and includes forms such as [dars]->[duruws] ("lesson"), [sijn]->[sujuwn] ("prison") and [qasˤr]->[qusˤuwr] ("castle").

The CCA is written in standard Arabic orthography (so-called "unpointed" orthography), in which short vowels are omitted. In unpointed orthography, the example just discussed has the form [drs]->[drws] ("lesson"), [sjn]->[sjwn] ("prison") and [qsˤr]->[qsˤwr] ("castle"), corresponding to the unpointed template pattern [CCC]->[CCVC]. Unpointed orthography presents numerous issues for text analysis (see also Buckwalter, 1997). We will discuss our approaches to the three most important issues here.

First, the glides [w] and [j] are orthographically identical to the long vowels [u:] and [i:] and to the diphthongs [ai] and [au]. Glides in a verbal root may be realised in surface form as long vowels (either with a single vowel quality or a diphthong), as consonants, or be deleted entirely. The glides are phonologically weak in morphological processes (McCarthy & Prince, 1990a) and in phonotactics (Frisch et al., 2004). Because of the ambiguity and their weakness in morphological processes, we classified all instances of [w] and [j] in the data-set as long vowels in the CV template.

Second, the diacritic that marks geminates is frequently omitted. In our data-set, geminates are represented as CC (where C represents *consonant*) in the CV templates only when the diacritic appeared in the orthographic form. Thus, some forms may contain an unmarked geminate.

Third, because diacritics for short vowels are only rarely included, multiple words are written in the same way. For example, active and passive verbs are generally distinguished by short vowels, such as [kataba] ("he wrote") and [kutiba] ("it was written"), which both appear in unpointed text as "ktb".

In order to determine how many forms in our data-set represent multiple pointed forms, we analysed a random 10% selection of our data-set using the Buckwalter Arabic Morphological Analyzer (BAMA; Buckwalter, 2004; Dehdari, 2009). The BAMA uses a dictionary of stems, suffixes and prefixes to parse the unpointed word and returns all possible attested pointed words including grammatical class and morphological structure. The 10% selection included forms from 39 gangs, of which 5 gangs contained multiple pointed templates. Four gangs had two pointed templates and one had three pointed templates. In one gang, this ambiguity arose from a long vowel surfacing as a consonant in one case and a vowel in the other. In the other four gangs, the ambiguity arose from an unmarked geminate. In the case of the three-way split, the unmarked geminate occurred on two different consonants. In sum, 87% of the analysed gangs had only one possible pointed template. Although the gangs with multiple templates create noise in the data-set, the strong correspondence between the pointed and unpointed CV templates for nouns shows that unpointed Arabic text can be a relatively accurate source of data for linguistic analysis.

Analogical modeling

The GCM is based on similarity as modulated by the larger context of the domain; as applied to morphological systems, it reflects an analogical framework, in which a test form is predicted to take the same pattern as the most similar form(s) in the context of similarity to all forms. We will first outline the GCM as implemented by Nakisa et al. (2001) and Albright and Hayes (2003), and then describe the five GCM-based models we implement for the model comparison.

The GCM employs segmental similarity to determine which lexical gang is the best pattern for a test form. Recall that the general definition of a lexical gang is a group of phonologically similar forms that display shared morphological behaviour, while we define a gang in Arabic based on the CV template of the singular and plural forms. For instance, the irregular English past tense pattern [−ing]→[−ung], as in [string]→[strung], [swing]→[swung] and [cling]→[clung], is a fairly robust gang, in that many forms in the group have resisted analogical levelling. Speakers are willing to generalise a novel form such as "spling" to this pattern (Bybee & Moder, 1983; Prasada & Pinker, 1993). By comparison, the pattern [−eave]→[−eft], as in [leave]→[left], is a more fragile gang, as it contains a smaller number of items, and some of the forms have undergone or are undergoing levelling (e.g. [cleave]→[cleft]/[cleaved]). An example from Arabic is the broken plural gang [CVCC]→[CVCVVC], as in [qarn]→[quruwn] ("century"), [ħaqq]→[ħuquwq] ("right") and [kalb]→[kilaab] ("dog").

The GCM predicts the plural pattern for a given form based on overall similarity of a test form to each candidate gang, weighted by gang size. Similarity of the test form to each form in a candidate gang is calculated using string-edit (Levenshtein) distance (Kruskal, 1983; Levenshtein, 1966), which measures the smallest number of changes necessary to transform the test form i into the comparison form j. We employ Albright and Hayes' (2003) and Nakisa et al.'s (2001) implementation of the GCM, which incorporates segmental similarity based on the natural class theory presented in Frisch et al. (2004), such that a change from one segment to another is weighted based on the number of shared and non-shared natural classes. We set the insertion/deletion cost at 1, while the transformation cost is between 0 and 1, depending on segmental similarity. Figure 4 shows the maximal alignment for [masjid] ("mosque") and [taʕliyq] ("comment").

String-edit distance is effectively a measure of dissimilarity, in that a larger value indicates lower similarity, so the string-edit distance is converted into similarity using the following transformation (Nakisa et al., 2001; Nosofsky, 1990):

$$\eta_{ij} = \exp(-d_{ij}/s)^p$$

In Equation (1) above, d_{ij} is the dissimilarity between form i and form j as defined by string-edit distance, η_{ij} is the transformed similarity between form i and form j, and s and p are fitted parameters. The s parameter modulates the tradeoff between similarity and gang size; with a low s, the model tends to select small gangs with high similarity to the test form, whereas with a high s, the model tends towards larger but less similar gangs. The best-fit value for s with our data was 0.3. The p parameter modulates the shape of the similarity function; $p = 1$ results in exponential decay with respect to similarity distance, while $p = 2$ results in Gaussian decay (Nosofsky, 1985; Shepard, 1987). We set $p = 1$, following Albright and Hayes (2003).

The overall similarity S_{iC_J} of a test form i to a gang C_J is given by the equation:

$$S_{iC_J} = \frac{\sum_{j \in C_J} \eta_{ij}}{\sum_{k \in C_K} \eta_{ik}}$$

This is calculated by summing the similarity η_{ij} of each member j of class C_J to the test form i, and dividing by the summed similarity η_{ik} of each member k of class C_K (the class of all stored forms) to the test form i. The similarity for a gang is not weighted by token frequencies of the gang members, only by type frequencies, which is equivalent to gang size. There is ample support in the literature for type statistics driving morphological generalisation, rather than token statistics (e.g. Albright, 2009; Baayen & Lieber, 1991; Bybee, 1995), and so token frequencies are not considered in this model.

The gang with the highest similarity rating is selected as the pattern that the test form will utilise in pluralisation. The model chooses deterministically (always selecting the most probable gang) rather than probabilistically. Deterministic choosing maximises the likelihood of being correct, as it results in 75% accuracy in a 25/75% binary choice task, while probabilistic choosing will result in 62.5% accuracy. This idealisation of subjects' choice behaviour establishes an upper bound on the model performance. It is justified by tendencies towards deterministic choosing observed in many experiments (Friedman & Massaro, 1998; Goldrick & Larson, 2008).

The GCM employs equal weighting across an entire form in calculating string-edit distance. This method is not optimal for English, as the part of the stem closest to the affix attachment point has the most influence on concatenative morphological processes (Albright & Hayes, 2003; Ernestus & Baayen, 2003). In the case of English verbs, the changes from present to past tense occur primarily at the end of the word, so information drawn from the end of the stem is most relevant. However, in Arabic, the entire form is local to the changes. Thus, the use of equal weighting across a form reflects the nature of

Figure 4. String-edit distance between [masjid] ("mosque") and [taʕliyq] ("comment").

Arabic templatic morphology and is a strength of this approach.

The GCM is conceptually similar to a *k*-nearest-neighbours model, with two major differences: *k* is variable depending on gang size, and similarity to a form is modulated by similarity to all other forms. Unlike other methods such as connectionist networks, wherein hidden units and subsymbolic representations do not correspond to a specific form or process (Green, 1999; Smolensky, 1988), the GCM is highly interpretable. It is trivial to examine the mechanisms of the model by examining similarity outputs.

The classic GCM as outlined above incorporates segmental similarity and gang size as factors in plural selection. We employ this method and four modified versions of the GCM to investigate three factors: (1) featural information, which examines if the fine-grained segment-specific features are important in pluralisation; (2) single match versus gang match, which tests whether the analogy is to a single form or to an entire gang; (3) CV template, which examines if plural-by-analogy is restricted to matching singular CV templates. The first factor is implemented by considering the segmental features in similarity (*Featural*) or by considering the CV template with no additional segmental features (*Simple Template Match*). The second factor is implemented by calculating similarity of the test form to the entire gang (*Whole Gang Match*) or to only the most similar form in the gang (*Single Best Match*). The third factor is implemented by restricting the search to candidate gangs with matching singular CV templates (*Restricted*) or considering all gangs (*Unrestricted*). The latter two factors can be combined into a 2 × 2 design, shown in Table 2 in the two centre columns. The first factor can only be implemented with gang size as a factor, as two candidate gangs with matching CV templates cannot be distinguished otherwise when segmental features are not used. This factor is considered by comparing a fifth model, *Simple Template Match*, to the featural models.

The *GCM, Unrestricted* is the model implemented in Albright and Hayes (2003) and considers all gangs in the comparison set, with the winning gang weighted by both gang size and segmental similarity based on shared natural classes. The *GCM, Restricted* is implemented by restricting the comparison to gangs that match the test form in singular CV template, such that the winning gang is that

with the highest similarity rating and the same singular CV template as the test form. The *Single Best Match, Unrestricted* model considers all gangs and matches the gang containing the single most similar form to the test form. The *Single Best Match, Restricted* model considers the gangs with the same singular CV template as the test form and selects the gang containing the single most similar form. All four of these models incorporate featural information in calculating similarity. The fifth model (*Simple Template Match*) considers only the CV template and gang size, such that the winning gang is the largest gang with the same singular CV template as the test form.

The core of the four featural models is the string-edit distance module (Albright, 2003), which calculates string-edit distance between two forms. We modified this script to incorporate the similarity-transformation equation (Nakisa et al., 2001; Nosofsky, 1990) and gang restrictions for our models. The segmental similarity matrix was created with the Segmental Similarity Calculator (Albright, 2006) using the natural class features for the Arabic consonant inventory from Frisch et al. (2004). Because this inventory includes only consonants, three vowel-specific features were added: [+high], [+front] and [+back], which are sufficient to distinguish the three vowels in the Arabic inventory, [a], [i] and [u]. In addition, the vowels were marked as [−consonant], [+sonorant], [+continuant] and [+voice] and were unmarked for all other features.

We implemented 10-fold cross-validation, which is a standard method of ensuring that the results of a model are replicable and generalisable (Breiman, Friedman, Olshen, & Stone, 1984; Mosteller & Tukey, 1968; Stone, 1974). This was performed by iterating the testing procedure for each model 10 times, each time with a randomly selected 25% of the data-set used as the test group and the remaining 75% used as the comparison group. The accuracy is the number of times the model selects the correct plural gang for a test form. For forms taking the sound plural, the model is considered accurate if it selects any sound gang with the appropriate suffix, whereas for forms taking the broken plural, the model must select the correct broken gang. As discussed above, all models choose deterministically.

We established a model baseline in order to interpret the results. For each test form, a random gang, weighted by gang size, was selected as the predicted winner.

Table 2. Names and factors of the five models.

	Featural, restricted	Featural, unrestricted	Template only
Single best match	Single best match, restricted	Single best match, unrestricted	–
Whole gang match	GCM, restricted	GCM, unrestricted	Simple template match

Results

Figure 5 shows the accuracy for the five models, averaged across 10 iterations. All statistical results reported here use two-sided paired *t*-tests with Bonferroni correction. Restricted featural models perform significantly better than unrestricted featural models, $t(19) = 11.56$, $p < .01$. Whole gang match featural models perform significantly better than single best match featural models, $t(19) = 8.79$, $p < .05$.

The best-performing model is the *GCM, Restricted* (65.97%). It performs significantly better than the second-best-performing model, the *Simple Template Match* (63.94%), $t(9) = 5.06$, $p < .01$. There is no significant difference between the second- and the third-highest performing models, the *Simple Template Match* and the *Single Best Match, Restricted* (63.23%), $t(9) = 1.39$, $p = .19$. Notably, the *GCM, Unrestricted* performs significantly worse than the third-ranked model, *Single Best Match, Restricted*, $t(9) = 2.96$, $p < .05$. No model performs worse than the random baseline condition.

Figure 6 shows the accuracy for each plural type for the five models. The sound [-aat] plural is classified accurately most frequently, with individual model accuracy ranging from 71% to 82%. The sound [-uun] plural (35–52%) and broken plural (30–37%) are classified accurately less frequently. The pattern of results is similar for each plural type, with the *GCM, Restricted* the highest-performing model for all types.

Model Accuracy for All Models
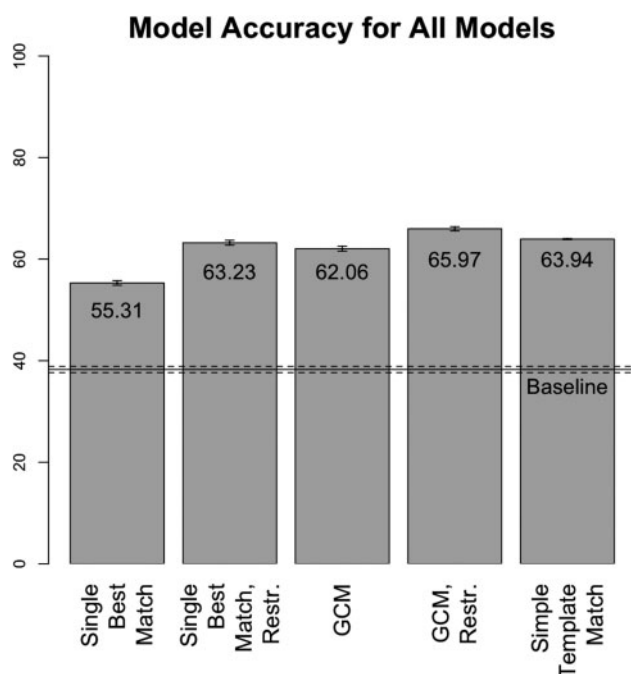
Figure 5. Model accuracy for all models. Solid line is the random baseline. Error bars show SE for each model and dashed lines show SE for the random baseline.

Figure 7 shows the accuracy for the five models when only forms taking broken plurals are considered, and the model selects from only broken plural gangs. Accuracy for this broken-plural-only set is much higher than accuracy for all types, despite a much lower baseline. The overall pattern of results is similar to the results for the entire set, with restricted featural models performing significantly better than unrestricted featural models, $t(19) = 14.19$, $p < .001$, and whole gang match featural models performing significantly better than single best match featural models, $t(19) = 4.53$, $p < .01$. Likewise, the *GCM, Restricted* is the highest-performing model of the five. However, unlike for the entire set, the *GCM, Restricted* does not perform significantly better than the *Simple Template Match*, $t(9) = 1.69$, $p = .12$.

A breakdown of errors made by each model is shown in Figure 8. Overall, the majority of errors for all models stem from classifying a broken plural as a sound plural or vice versa. Errors made by confusing one sound plural for the other are frequent, while errors made by confusing broken plurals are uncommon. The most accurate models (*GCM, Restricted* and *Simple Template Match*) are the least likely to make sound-to-sound and broken-to-broken errors.

General discussion

The major goal of this work is to investigate the learnability of the noun plural system in Arabic. Specifically, we examine the influence of the CV template, segmental features and gang size on morphological generalisation in the Arabic noun plural system. We have shown that all the three of these factors are relevant for the noun plural system, albeit to varying degrees. Taken together, these factors are highly successful in predicting the plural template for unseen forms. Future work might yield improved accuracy by including features we omitted, in particular semantic and short vowel features.

Overall, these results strongly support McCarthy and Prince's (1990a) claim that the CV template is the major factor in the productivity of noun pluralisation patterns in Arabic. We find that considering only gangs with matching singular CV templates improves accuracy by a substantial margin (restricted vs. unrestricted models). This claim depends on the abstraction of the CV template as developed in McCarthy (1981) and illustrated in Figures 1 and 2. McCarthy (1993) and McCarthy and Prince (1990a, 1990b) further support this theory with their analyses of the Arabic diminutive and verbal noun systems, claiming that the CV template is the driving factor in these morphological processes as well. Boudelaa and Marslen-Wilson (2004) also provide psycholinguistic evidence supporting the abstraction of the CV template in Arabic. They found that word recognition in a lexical decision task is facilitated when the prime and target share
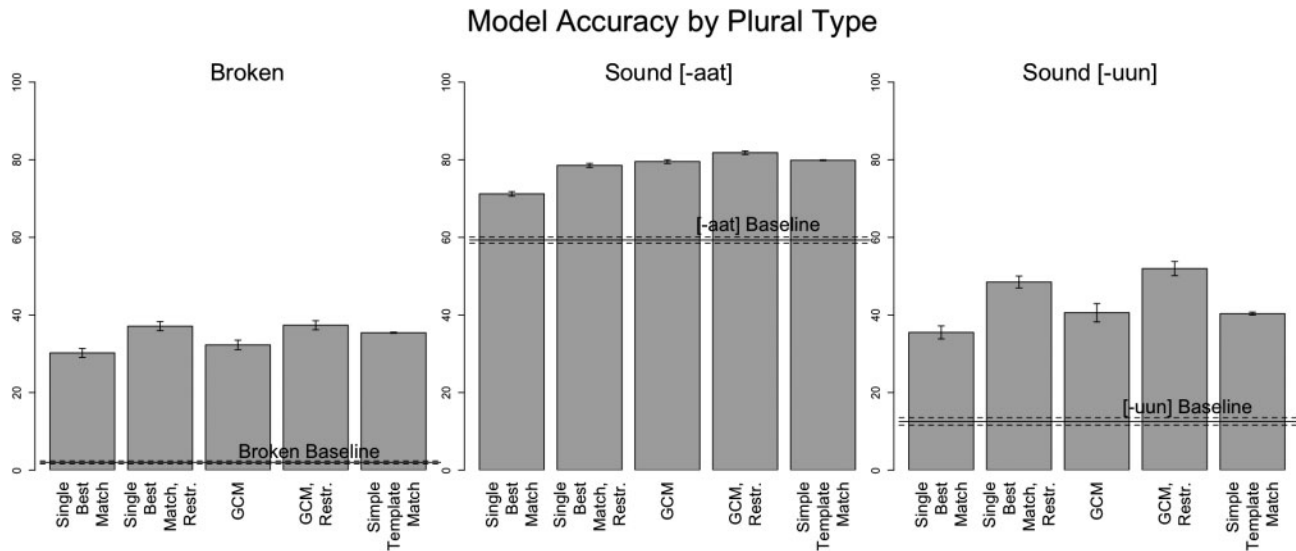
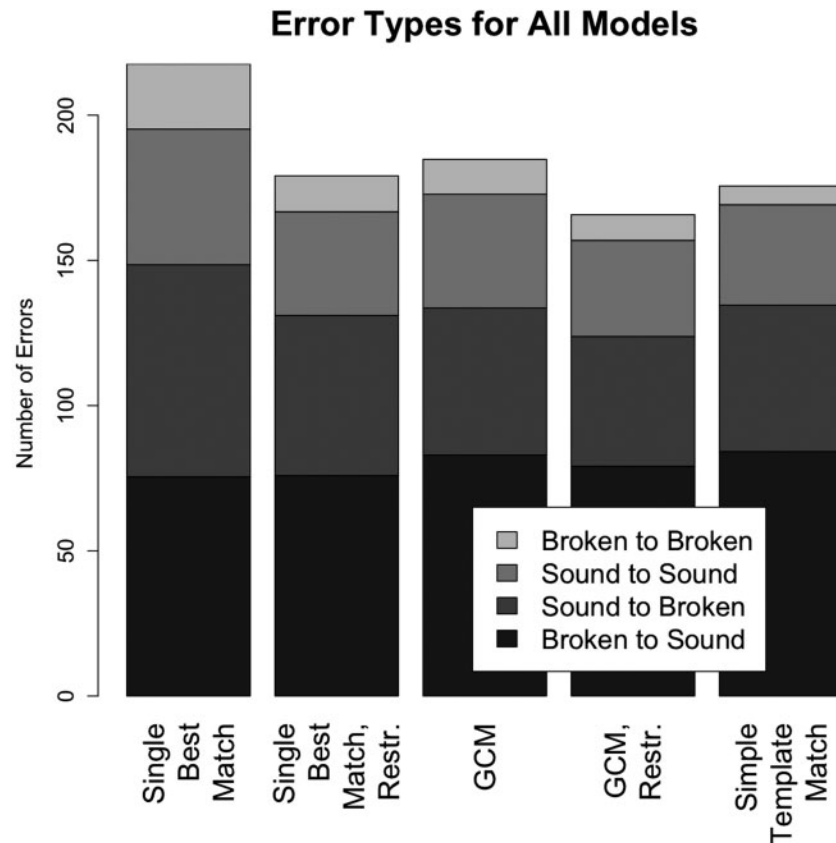**Error Types for All Models**

Figure 8.　Error types for all models.

features does not significantly increase accuracy for the broken-plural-only set.

Finally, our error analysis shows that the majority of errors come from selecting a sound plural for a form that takes the broken plural (*broken-to-sound*) or selecting a broken plural for a form that takes the sound plural (*sound-to-broken*). Few errors stem from selecting the wrong broken plural type for a form that takes the broken plural (*broken-to-broken*), which is in line with our conclusion above that the shape of the broken plural is more predictable than whether a form takes the sound or broken plural. The pattern of errors for our models is similar to errors made by children learning Palestinian Arabic in Ravid and Farah (1999), as shown in Figure 9, with age indicated as year;month. In particular, children in the 3;4–4;1 group bear the strongest resemblance to our data, with broken-to-sound errors accounting for the majority of all errors, and broken-to-broken errors accounting for a small percentage. Like our model, the children made many sound-to-sound errors, suggesting that they have not yet learned the highly reliable gender and animacy features governing [-uun] versus [-aat]. In contrast to our results, children make few sound-to-broken errors. This difference may reflect a limitation in the treatment of lexical gangs by the GCM. Because

membership in a lexical gang requires identity in both the singular and the plural templates, the GCM model divides the sound plural into 71 different gangs. Each

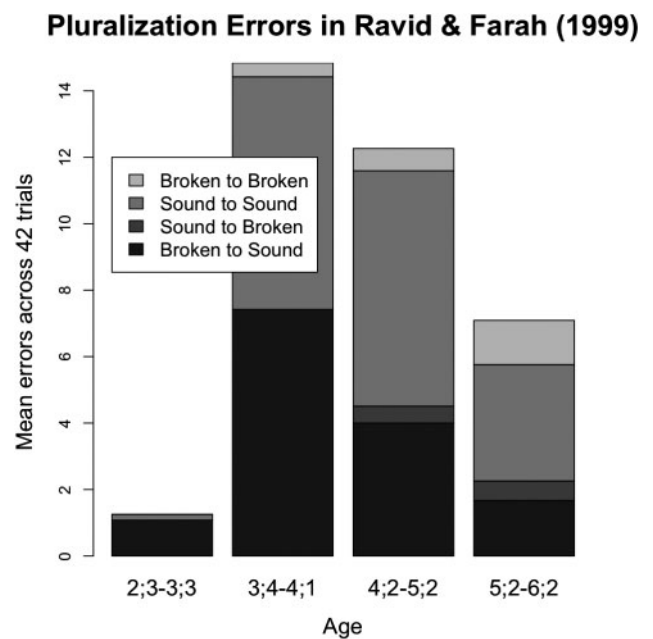**Pluralization Errors in Ravid & Farah (1999)**

Figure 9.　Errors types for children in Ravid and Farah (1999).

individual sound plural gang is therefore a weaker competitor to the closest broken plural gang than all the sound plurals would be if acting together as one large gang. In summary, the GCM has a limited ability to capture default morphological patterns when these involve heterogeneous word pairs.

The high degree of abstraction required to posit the CV template is at odds with many current models of morphological generalisation, which calculate lexical gang membership on the basis of fine-grained phonological similarity (Bybee & Moder, 1983; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Prasada & Pinker, 1993; Rumelhart & McClelland, 1986; Skousen, 1989, 1993). However, a growing body of literature suggests that statistical pattern generalisation can include coarse-grained pattern abstraction, which in this analysis is the intersection of gang size and the CV template.

In a study of long-distance consonantal harmony patterns, Heinz (2010) showed that a precedence learner model can learn patterns based on the order of the sounds regardless of distance. That is, long-distance dependencies can be explained by general principles of phonotactic learning. This case is similar to the verbal root in non-concatenative morphology, as the morpheme appears non-contiguously in the surface form and can be identified by the order of consonants rather than the distance between them.

In a study of phonological generalisation for different sized vocabularies, Pierrehumbert (2001) found that even highly regular patterns cannot be acquired if type frequency is too low (see also Bybee, 1995). The need for robustness across individual vocabularies necessitates that phonological patterns are, to some extent, coarse-grained. These representations are better able to capture generalisations over variable vocabularies and increasing sample sizes while satisfying the tendency towards short phonological descriptors.

These studies exemplify the tendency towards coarse-grained representations in order to capture phonological and morphophonological patterns. These coarse-grained representations arise from statistical inference over the lexicon and are activated in word formation processes. The current study extends these findings to non-concatenative morphology and shows that the abstract CV template plays a crucial role in pattern learnability and generalisation for the noun plural. Fine-grained features play a significant, if small, role in this system. Learning non-concatenative morphological systems depends on generalising across multiple levels of granularity, and future work in this area should examine both coarse- and fine-grained generalisations.

Acknowledgements

References

Al-Sulaiti, L. (2009). *Corpus of Contemporary Arabic* [dataset]. Retrieved August 25, 2009, from http://www.comp.leeds.ac.uk/eric/latifa/CCA_raw_utf8.txt

Albirini, A., & Benmamoun, E. (2012). Aspects of second-language transfer in the oral production of Egyptian and Palestinian heritage speakers. *International Journal of Bilingualism*. Advance online publication. doi:10.1177/1367006912441729

Albright, A. (2003). Whole word alignment and similarity calculator (Version 4/13/2003) [Perl script]. Retrieved from http://www.mit.edu/~albright/software/Alignment.zip

Albright, A. (2006). Segmental similarity calculator (Version 1/28/2006) [Perl script]. Retrieved from http://www.mit.edu/~albright/software/SimilarityCalculator.zip

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, *26*, 9–41. doi:10.1017/S0952675709001705

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*, 119–161. doi:10.1016/s0010-0277(03)00146-x

Alegre, M., & Gordon, P. (1999). Rule-based versus associative processes in derivational morphology. *Brain and Language*, *68*, 347–354. doi:10.1006/brln.1999.2066

Baayen, H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, *29*, 801–843. doi:10.1515/ling.1991.29.5.801

Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.

Berman, R. A. (1981). Children's regularizations of plural forms. *Stanford Papers and Reports on Child Language Development*, *20*, 34–44. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED210906

Boudelaa, S., & Gaskell, M. (2002). A re-examination of the default system for Arabic plurals. *Language and Cognitive Processes*, *17*, 321–343. doi:10.1080/01690960143000245

Boudelaa, S., & Marslen-Wilson, W. (2004). Abstract morphemes and lexical representation: The CV-skeleton in Arabic. *Cognition*, *92*, 271–303. doi:10.1016/j.cognition.2003.08.003

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.

Buckwalter, T. (1997). Issues in Arabic morphological analysis. In A. Soudi, A. van den Bosch, & G. Neumann (Eds.), *Arabic computational morphology: Knowledge-based and empirical methods*. (vol. 38, pp. 23–41). Dordrecht: Springer.

Buckwalter, T. (2004). Buckwalter Arabic Morphological Analyzer (2.0) [Perl script]. Philadelphia, PA: Linguistic Data Consortium.

Bybee, J. L. (1985). *Morphology: A study of the relation between form and meaning*. Amsterdam: John Benjamins.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455. doi:10.1080/01690969508407111

Bybee, J., & Moder, C. (1983). Morphological classes as natural categories. *Language*, *59*, 251–270. Retrieved from http://www.jstor.org/stable/413574

Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. (1992). Regular and irregular inflection in the acquisition of German

noun plurals. *Cognition*, *45*, 225–255. doi:10.1016/0010-0277(92)90018-D

Daelemans, W., Gillis, S., & Durieaux, G. (1994). The acquisition of stress, a data-oriented approach. *Computational Linguistics*, *20*, 421–451. Retrieved from http://dl.acm.org/citation.cfm?id=204923

Dehdari, J. (2009). AraMorph Fast (1.2.1) [Perl script]. Retrieved from http://aramorph.sourceforge.net/

Derwing, B., & Skousen, R. (1994). Productivity and the English past tense: Testing Skousen's analogy model. In S. Lima, R. Corrigan, & G. Iverson (Eds.), *The reality of linguistic rules, studies in language companion* (pp. 193–218). Amsterdam: John Benjamins.

Ernestus, M., & Baayen, H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*, 5–38. doi:10.2307/4489384

Friedman, D., & Massaro, D. (1998). Understanding variability in binary and continuous choice. *Psychonomic Bulletin & Review*, *5*, 370–378. doi:10.3758/BF03208814

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, *22*, 179–228. doi:10.1023/B:NALA.0000005557.78535.3c

Frisch, S., & Zawaydeh, B. (2001). The psychological reality of OCP-place in Arabic. *Language*, *77*, 91–106. doi:10.1353/lan.2001.0014

Goldrick, M., & Larson, M. (2008). Phonotactic probability influences speech production. *Cognition*, *107*, 1155–1164. doi:10.1016/j.cognition.2007.11.009

Green, C. (1999). Are connectionist models theories of cognition? In W. Maiers, B. Bayer, B. Duarte Edgalhado, R. Jorna, & E. Schraube (Eds.), *Challenges to theoretical psychology* (pp. 139–146). North York, NY: Captus Press.

Hammond, M. (1988). Templatic transfer in Arabic broken plurals. *Natural Language & Linguistic Theory*, *6*, 247–271. doi:10.1007/BF00134231

Haspelmath, M., & Sims, A. (2010). *Understanding morphology* (2nd ed.). London: Hodder Education.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, *9*, 342–348. doi:10.1016/j.tics.2005.04.002

Heinz, J. (2010). Learning long-distance phonotactics. *Linguistic Inquiry*, *41*, 623–661. doi:10.1162/LING_a_00015

Khouloughli, D.-E. (1992). *Basic lexicon of Modern Standard Arabic*. Paris: L'Harmattan.

Kruskal, J. (1983). An overview of sequence comparison: Time warps, string edits, and macromolecules. *SIAM Review*, *25*, 201–237. doi:10.1137/1025045

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*, 707–710. Retrieved from http://profs.sci.univr.it/~liptak/ALBioinfo/files/levenshtein66.pdf

Levy, M. (1971). *The plural of the noun in Modern Standard Arabic* (Unpublished doctoral dissertation). University of Michigan, Ann Arbor.

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, *29*, 189–256. doi:10.1006/cogp.1995.1015

McCarthy, J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry*, *12*, 373–418. Retrieved from http://www.jstor.org/stable/info/4178229

McCarthy, J. (1982). Prosodic templates, morphemic templates, and morphemic tiers. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (vol. I, pp. 191–223). Dordrecht: Foris.

McCarthy, J. (1985). *Formal problems in Semitic phonology and morphology*. New York, NY: Garland.

McCarthy, J. (1993). *Template form in prosodic morphology*. Paper presented at the Papers from the Third Annual Formal Linguistics Society of Midamerica Conference, Bloomington, IN.

McCarthy, J., & Prince, A. (1990a). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory*, *8*, 209–283. doi:10.1007/BF00208524

McCarthy, J., & Prince, A. (1990b). Prosodic morphology and templatic morphology. In M. Eid & J. McCarthy (Eds.), *Perspectives on Arabic linguistics II* (pp. 1–53). Amsterdam: Benjamins.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0

Mohamed, E. (2009). *List of Arabic broken plurals* [dataset]. Retrieved November 20, 2009, from http://jones.ling.indiana.edu/~emadnawfal/arabicPlural.txt

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., pp. 80–203). Reading, MA: Addison-Wesley.

Murtonen, A. (1964). *Broken plurals: Origin and development of the system*. Leiden: E.J. Brill.

Nakisa, R., Plunkett, K., & Hahn, U. (2001). Single- and dual-route models of inflectional morphology. In P. Broeder & J. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 201–222). Cambridge, MA: MIT Press.

Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception and Psychophysics*, *38*, 415–432. doi:10.3758/BF03207172

Nosofsky, R. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, *34*, 393–418. doi:10.1016/0022-2496(90)90020-A

Omar, M. (1973). *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.

Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, *16*, 691–698. doi:10.1080/01690960143000218

Plunkett, K., & Nakisa, R. (1997). A connectionist model of the Arabic plural system. *Language and Cognitive Processes*, *12*, 807–836. doi:10.1080/016909697386691

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*, 1–56. doi:10.1080/01690969308406948

Prunet, J. (2006). External evidence and the Semitic root. *Morphology*, *16*, 41–67. doi:10.1007/s11525-006-0003-5

Ratcliffe, R. R. (1998). *The 'broken' plural problem in Arabic and comparative Semitic: Allomorphy and analogy in non-concatenative morphology*. Amsterdam: John Benjamins.

Ravid, D., & Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language*, *19*, 187–206. doi:10.1177/014272379901905603

Rumelhart, D., & McClelland, J. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*, 60–94. doi:10.1037/0033-295X.89.1.60

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In J. McClelland, D. Rumelhart, & P. R. Group (Eds.), *Parallel distributed processing: Explorations in the*

Language, Cognition and Neuroscience

1281

microstructure of cognition (vol. 2, pp. 216–271). Cambridge, MA: MIT Press.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323. doi:10.1126/science.3629243

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

Skousen, R. (1993). *Analogy and structure.* Dordrecht: Kluwer.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–74. doi:10.1017/S0140525X00052432

Stemberger, J. P., & MacWhinney, B. (1988). Are inflected forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 101–116). San Diego, CA: Academic Press.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Royal Statistics Society*, *B36*, 111–147. doi:10.2307/2984809

Wehr, H. (1976). *Arabic–English dictionary: The Hans Wehr dictionary of modern written Arabic* (3rd ed.). Ithaca, NY: Spoken Language Services.

Wright, W. (1896/1988). *A grammar of the Arabic language* (3rd ed.). Cambridge: Cambridge University Press.

Appendix

McCarthy and Prince's (1990a) well-known treatment of the broken noun plural in Arabic proposes that the CV template of the singular noun is the most important factor in plural formation. Specifically, they make four major claims regarding the distribution of noun plurals. First, the authors claim that "essentially all canonically-shaped lexical nouns of Arabic take broken plurals" (p. 212). Second, they claim that the sound plural is systematically applied only to transparently derived nouns and adjectives, proper names, non-canonical or unassimilated loans, and the names of the letters of the alphabet. Third, the authors posit that only the iambic broken plural is productive, and that it is dominant or a significant competitor in 9 of 10 canonical singular noun classes. Finally, the authors claim based on these three sources of evidence that the iambic broken plural is the default method of pluralisation. This may seem at odds with the many claims that the sound plural is the default (Levy, 1971; Murtonen, 1964; Nakisa et al., 2001; Plunkett & Nakisa, 1997; Wright, 1896/1988); however, the authors cite the broken plural's wide applicability across canonical noun types rather than distributional statistics as evidence for its proposed default status. We examine these claims in detail here.

We will first examine the distribution of broken and sound plurals for canonical and non-canonical nouns. The canonical noun is defined by prosodic constraints, where the minimal word is bimoraic (McCarthy & Prince, 1990a). Further constraints on syllable structure in Arabic thus yield 10 classes of canonically shaped singulars, with five CV templates that are either unsuffixed or suffixed by the feminine marker *taa marbuta* (+aT)

We examined the full set of broken plurals (*n* = 561) from our data-set and a random selection of 10% of the sound plurals (collapsing [-uun] and [-aat]; *n* = 138). The percentages of broken and sound plurals across noun classes are estimated from this 10% sample. We used the BAMA (Buckwalter, 2004; Dehdari, 2009) to retrieve the full pointed form for each singular and plural. The singulars were then classified as canonical or non-canonical and sub-classified by canonical noun class.

First, we find that all non-canonical noun types in our sample take sound plurals (*n* = 90). This finding is in accord with McCarthy's claim that the canonicity of the singular is a major factor in broken pluralisation. However, we find that the sound plural occurs in 9 of the 10 canonical noun classes, with 6 classes having a majority of sound plurals and 2 additional classes having at least 40% sound plurals, as shown in Figure 10. Overall, we estimate that 58% of canonical nouns take a sound plural.

Second, we examine the claim that the sound plural is restricted to transparently derived nouns and adjectives, proper names, noncanonical or unassimilated loans, and the names of the letters of the alphabet. We find many cases of the noun types listed above taking the broken plural, as well as noun types outside this set taking the sound plural. For example, there are many classes of derived forms that take the broken plural, including the participles CaaCiC (*n* = 27) and CaaCiCaT (*n* = 17), as well as the nouns of place maCCaC (*n* = 45) and maCCiC (*n* = 17) and the instrument miCCaC (*n* = 16). We also find underived basic terms taking sound plurals (although the classification of these forms is somewhat difficult, as verb roots are sometimes derived from underived nouns, e.g. [kalaba] "to become mad or rabid" from [kalb] "dog"). For example, underived nouns of the form CvCC (*n* = 3 in 10% sample), CvCCaT (*n* = 5 in 10% sample) and CaCaCaT (*n* = 4 in 10% sample) can take sound plurals.
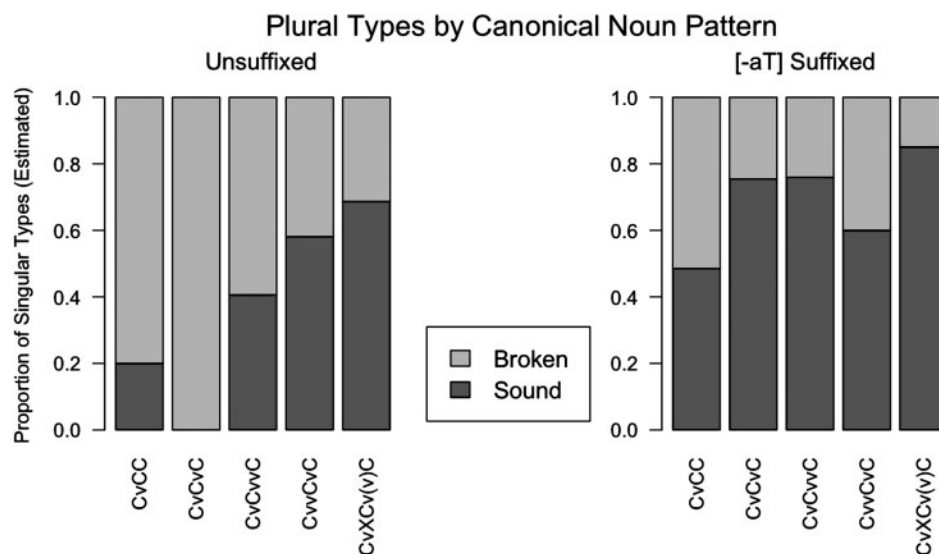

Figure 10. Proportion of broken and sound plurals by canonical noun class.
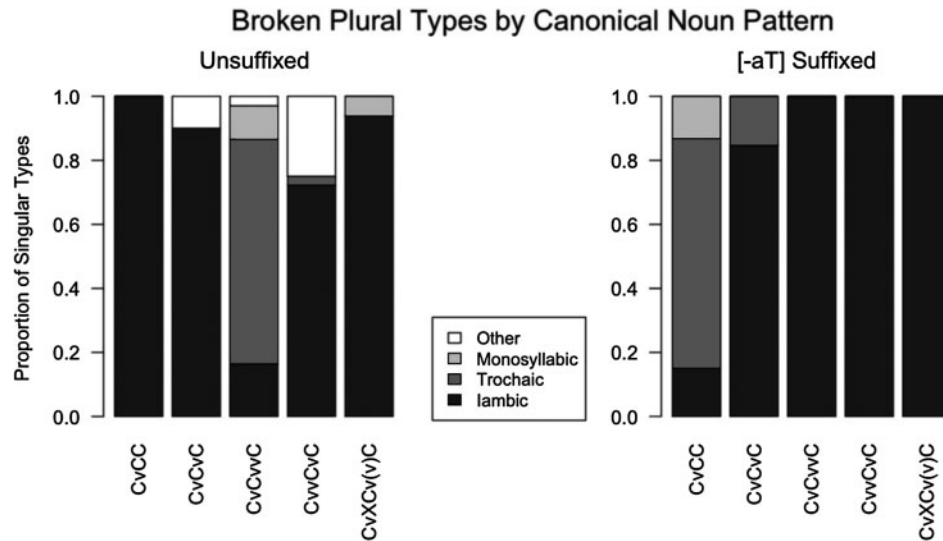
Figure 11.　Proportion of broken plural types by canonical noun class.

Third, McCarthy and Prince claim that the iambic plural is the only truly productive broken plural type. Recall that the broken plurals fall into four prosodic patterns: iambic (CvCvv+), trochaic (CvCvC), monosyllabic (CvCC) and other (CvCCvC and CvCCvvC). The authors examine broken plural types within each canonical noun class and find that the iambic plural is used exclusively in two classes, in over 90% of forms in three classes, and between 20% and 50% of the time in three classes.

We find similar results with our broken plural set, with seven classes having over 80% iambic plurals (Figure 11). In one additional class, the iambic accounts for 72% of plurals. However, in the remaining two classes (CvCvvC and CvCC+aT), the trochaic plural is dominant, with 71.7% and 70.4% of forms taking it, respectively. This is not out of line with McCarthy and Prince's findings; they reported that 50–61% of CvCvvC nouns take the trochaic plural (depending on voweling) and that 75% of CiCCaT and CuCCaT nouns take the trochaic (CaCCaT is reported to take the trochaic very rarely). Nonetheless, the dominance of the trochaic plural in these classes suggests that it may be productive, albeit with limited scope.

Note also that CvvCvC nouns take the "other" plural CuCCaaC 25% of the time. McCarthy and Prince remark that this is common for lexicalised active participles of the form CaaCiC. While we cannot comment on the degree of lexicalisation of the nouns in our data-set, 33% of the CaaCiC singulars take the "other" plural, which suggests that it may have some generality within this niche.

Finally, the authors claim on the basis of their first three claims that the iambic plural is the default form of pluralisation for Arabic nouns. Our analyses here suggest that, while the iambic plural is the dominant broken plural type for most canonical noun classes, it is nonetheless in the minority compared to the sound plural. There are a number of factors at work in determining whether a singular noun will take a broken or sound plural, the most important of which is the CV template of the singular. This is evident in the importance of canonicity in plural selection, with only canonical noun types eligible for broken pluralisation. However, our model results have shown that gang effects and segmental features of the singular are also influential in this choice. These factors likely account for the specific selection of broken versus sound plurals in classes where they are in strong competition. An additional factor that may be relevant is the degree of lexicalisation of the noun; however, this factor is difficult to estimate and as such is not discussed here.

Within broken plurals only, McCarthy and Prince are relatively accurate regarding which broken plural pattern a singular noun is likely to take. Nonetheless, the general dominance of the sound plural suggests that it is the default mode of pluralisation, although the broken plural (and in particular the iambic) is statistically robust within specific canonical noun classes. Additionally, we predict that the trochaic plural is productive within two specific classes, and that the "other" plural has some generality for singulars of the form CaaCiC. These predictions need to be tested empirically, but the statistics shown here suggest that speakers may be willing to generalise trochaic and "other" plurals to new forms within these classes.