In search of the minority default: the case of Arabic plurals

Sami Boudelaa (sami.boudelaa@mrc-cbu.cam.ac.uk)
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road,
Cambridge, CB2 2EF, UK

M. Gareth Gaskell (g.gaskell@psych.york.ac.uk)
Department of Psychology, University of York,
Heslington, York YO10 5DD, UK

Abstract

A Minority-default inflectional system is one in which a regular affixational process (e.g., the plural morpheme ~s in English) applies to fewer forms in the language than the irregular stem modifying process (e.g. the umlauting in "foot-feet"-like pairs). Following the work of McCarthy & Prince (1990), the plural system of Modern Standard Arabic has been cited as an archetype of a minority-default system with the regular sound plural involving fewer nominal forms than the irregular broken plural. On the basis of linguistic, statistical and distributional evidence, we argue that this assertion is wrong. We point out that while both broken and sound plural have qualitatively limited productivity, the latter is quantitatively the more productive process. Furthermore, the diversity of regularly inflected phonological forms ensures that they will be treated as the default by a connectionist model. In the light of these findings we argue that a good model of morphological processing should motivate the observation that so few of the world's languages use minority defaults.

Introduction

A major debate in psycholinguistics revolves around the question of how human language users employ finite means to produce large numbers of words and utterances. In order to deal with this generic question, several more specific questions need to be spelt out. One such specific question is whether or not the structural properties of regularly and irregularly inflected words correspond to their representational and processing properties. Focusing on the representational format would lead one to tackle the question of whether morphologically complex words are represented as full forms or as decomposed morphemes (Marslen-Wilson et al., 1994). Focussing on the processing aspect of the equation would lead one to raise the same question from a different standpoint, namely whether morphologically complex words are formed via a symbolic rule-based mechanism operating on grammatical categories or via a memory-based associative network that extracts probabilistic contingencies between them (Rumelhart & McClelland, 1986; Pinker & Prince 1988; Marslen-Wilson & Tyler, 1998).

The acquisition of the English past tense has been extensively studied in an attempt to decide between the different approaches to this problem. The literature on the subject provides at least three different models. The first, and most traditional assumes that the regular past forms in English like "walk-walked" are formed by a rule, whereas irregular past tenses like "eat-ate, give-gave" are learned individually by rote (Berko, 1958). Because it fails to explain the sub-regularities among the irregular verbs and the expansion of irregular inflection to phonologically similar nonce forms, this view has largely been superseded by a second model which claims that a rule-governed process inflects all the regular forms while an associative memory takes care of all the irregular forms. The associative memory identifies the irregular forms and blocks the default process from applying to them (Pinker, 1991; Pinker & Prince, 1988). The third model is a connectionist one, which dispenses with explicit rules and assumes that language learning is better accounted for using a single mechanism, namely a network of interconnected units (Rumelhart & McClelland, 1986). Both regular and irregular forms are inflected by this network, with responses to novel forms depending on their phonological similarity to familiar patterns (Plunkett & Marchman, 1991).

Both dual-route models and connectionist networks are able to handle an inflectional system like English because of its distributional characteristics. The English system is one in which about 95% of the forms are regularly inflected. This is an unproblematic situation for a dual-route model, which deals with the small number of irregulars via associative memory and the rest via a default rule. A connectionist network would also exhibit relative ease handling such cases. The network would store information about all forms and the preponderance of regular forms would trigger a regularisation process, by virtue of the fact that any novel form is more likely to resemble a regular form than an irregular one. Proponents of the dual-route model have argued that a dual mechanism can also deal satisfactorily with linguistic systems where the default is a minority, such as the inflection system in German (Clahsen, 1999). This is because rule-like behavior does not need to be contingent on the default pattern applying to a majority of the forms in the language. Rather, a default can be defined, the argument goes, even in terms of the least frequent patterns, because this process depends on applying the same procedure to different items bearing the same symbol (Clahsen, 1999). Conversely, a connectionist network was predicted to be unable to simulate people's regularisation of novel forms in a minority-default system like German.

The Arabic plural is perhaps the most widely cited example of a minority default system (McCarthy & Prince, 1990; Hare, Daugherty & Seidenberg, 1992; Ravid & Farah, 1999). For this reason it was used as a litmus test by Plun-

kett and Nakisa (1997) who found that a connectionist network can model generalisation behaviour to both regular and irregular patterns, despite the absence of a default rule. One of our aims here is to take issue with the position that Arabic has a minority default plural system, and show that it hinges on an inaccurate description of the language. In order to come to grips with this claim we will begin by laying out the morphological system of Modern Standard Arabic and argue that this language does not exhibit a minority-default, using linguistic and corpus analyses. Second, we will examine the phonological distribution of Arabic nominal forms using a more representative sample than the one used by Plunkett and Nakisa (1997). All these sources of evidence converge on the idea that the Arabic plural system has a majority default of the type learnable by a connectionist model. We conclude by considering why minority default systems seem scarce across world languages.

The Morphological System of Arabic

Traditionally Arabic surface forms are analysed as consisting of two abstract morphemes a root and a word pattern[1]. The root usually comprises three consonants and carries semantic meaning, while the word pattern contains vowels and conveys syntactic information. According to this approach, the representation of a surface form such as [nuqil] "*be moved*" will consist of the root {nql}, and the word pattern {fuʕil} where the letters "f, ʕ, l" indicate the slots into which the root consonants map.

The morphology of Arabic falls into two relatively distinct parts (Bohas & Guillaume, 1984). The first consists of primitive nouns that are thought to be unrelated to verbs, although verbs can be derived from them. For example, from the primitive noun [kalbun] "*dog*" the verb [kaliba] "*get infected with rabies*" can be formed. The second part relates to verb morphology and subsumes verbs proper and deverbal nouns. Verb morphology can be further divided into unaugmented and augmented verb forms. There are three unaugmented forms and 14 augmented forms of which only 9 are frequently used in Modern Arabic. As for deverbal nouns, there are about 10 types such as the active participle, passive participle, instance noun, manner noun, "assimilated noun", and the "masdar" (Holes, 1995).

Verb Morphology

Verb morphology with its two components is the most productive part of the language in the sense of being the main source of most of the *transparent derivatives*[3]. For example, combining the root {xrʒ} "*go out*" with the pattern {faʕal}, produces the form [xaraʒ] "*go out*". The same root can be further combined with as many as 5 augmented patterns yielding the following surface forms: [xarraʒ] "*move out*" [ʔaxraʒa] "*take out*", [taxarraʒ] "*graduate*" [taxaaraʒ] "*disengage*", [ʔistaxraʒ] "*extract*". From each of these forms a host of deverbal nouns can be derived. For example, the masculine active participle [xaariʒun] can be derived from the unaugmented surface form [xaraʒ]. Also, the following active participles can be derived respectively from each of the augmented verb forms above: [muxarriʒ], [muxriʒ], [mutaxarriʒ], [mutaxaariʒ], [mustaxriʒ]. Passive participles can also be formed from these verb forms. In addition to this, an "instance noun", a noun denoting that the action takes place only once, [xarʒatun] "*one departure*" can be obtained from the verb [xaraʒ], the noun [taxaaruʒ] can be derived from the verb [taxaaraʒ], the noun [ʔistixraaʒ] can be derived from the verb [ʔistaxraʒ] and so on. This pattern of productivity holds even for verbs that are originally derived from primitives. Thus from the primitive noun [kalb] "*dog*" the verb [takaalab] "*to rave*" is derived and from the latter an active participle [mutakaalib] "*someone who raves*" and a noun [takaalub] "*raving*" are formed. Similarly, loan words like [talifuun] "*telephone*" can be used to derive verbs such as [talfan] "*to telephone*", and an active participle like [mutalfin] "*phone-caller*".

Nominal Morphology

Arabic nouns undergo various morphological alterations of which the most frequent is perhaps pluralization. This is achieved either via suffixation or via pattern modification. In the first case, known as sound pluralization, the suffix ~uun is added to masculine nouns (e.g. [naaʒiħun-naaʒiħuun] "*successful*" male) while ~aat is appended to feminine nouns (e.g. [naaʒiħatun-naaʒiħaat] "*successful*" female). In the second, often referred to as broken pluralization, the pattern of the singular noun is dramatically altered and in some cases some of its consonants are lost (e.g. [ʕunquud-ʕanaqiid] "*cluster*" [ʕandaliib-ʕanaadil] "*nightingale*" (Murtonen, 1964; Xasaara, 1994; Holes, 1995). Sound pluralization is considered as regular inflection because it involves little or no allomorphy while broken pluralization is irregular because it is rich with allomorphic variations.

McCarthy & Prince's (1990) work on the broken plural in Modern Standard Arabic has promulgated the idea of Arabic having a minority default system of pluralization. According to them the sound plural is "*systematically found only with the following short list: proper names; transparently derived nouns or adjectives such as participles, deverbals and diminutives; non-canonical or unassimilated loans and the names of the letters of the alphabet*" (McCarthy & Prince, 1990: p. 212). Phrased as such, the above claim is misleading because it fails to distinguish between qualitative and quantitative productivity. The distinction between these two aspects of productivity rests on the difference between the number and/or the strength of the constraints weighing on a particular morphological process (Aronoff & Anshen, 1998). Perhaps an English example will help to bring our point home. The suffix ~ity is qualitatively productive but quantitatively unproductive. This is because it tends to be appended preferentially to adjectives ending in suffixes like ~ible, ~able, ~ic, ~id etc. Conversely the suffix ~ness, is quantitatively productive because it is subject to fewer constraints and is not restricted to follow a

[1] Within the framework of multilinear phonology the word pattern is further broken down into a vocalic morpheme and a skeletal morpheme (McCarthy, 1981).

[2] Some authors claim that "as many as 400 different surface word forms can be derived from some triliteral verbs" (Xasaara, 1994, p. 134)

limited set of suffixes (Aronoff & Anshen, 1998). Arabic sound and broken pluralization processes lend themselves readily to a description in terms of a distinction between qualitative and quantitative productivity. Both are subject to few constraints. Sound pluralization is restricted to a set of nominal forms that must meet formal (e.g., length in syllables) and syntactic criteria (e.g. being preferably adjectives). But broken pluralization is subject to even more rigid and more numerous formal (e.g. length and syllabic structure) and syntactic criteria (e.g. being preferably a substantive). Quantitatively, however sound pluralization would not be a minority case even if it were found *only* with transparent derivatives. Transparent derivatives, as we will shortly show, correspond to the most productive part of the language. Additionally, sound pluralization affects systematically all recent loan words comprising more than three letters like [dimuqraatˀiyyun] "*democracy*", [talfazatun] "*television*".

Type Frequency of Broken and Sound Plurals

A given triliteral root in Arabic can be productively mounted on some combination of the 9 frequent augmented word patterns to create new words. For instance, the triliteral unaugmented surface form [katab] "*write*" can be combined with as many as 7 augmented forms. Conversely, the unaugmented triliteral [ʕabaθ] "*fool around*" gives rise only to one augmented form [ʕaabaθ] "*banter*". Although no systematic statistical work on the number of augmented and unaugmented verb forms is available in Arabic, one may safely hypothesise that triliteral roots can yield on average at least three surface forms. Confining our analysis to active and passive participles in the masculine and feminine forms, we can plausibly say that each of the augmented forms gives rise to at least 4 deverbal forms. There are 11978 roots of which 7597 are triliterals, 4081 are quadriliterals and 300 are quinquiliterals (Moussa, 1996). Assuming that the derivation of four masculine and four feminine deverbal surface forms from each root is not an overestimate, the triliteral roots alone will yield as many as 91164 surface forms that take a sound plural. If we consider the derivatives from quadriliteral and quinquiliterals, this estimate will increase greatly.

It is true that some transparent derivatives like "assimilated nouns" and lexicalized active participles often pluralize in the broken way. This does not mean that nouns taking a broken plural will outnumber those pluralizing regularly because for almost every assimilated noun or indeed for any other noun that has a broken plural, there is either a diminutive form, a feminine form or both, and these take a sound plural. Thus the assimilated noun [ʕaaqir] "*barren*" has the broken plural [ʕawaaqir], whereas its diminutive [ʕuwaiqir] has the sound feminine plural [ʕuwaiqiraat]. Likewise, the primitive noun [qird] "*monkey*" has a broken plural [quruud] but its feminine form [qirdatun] "*female monkey*" has a sound plural form [qirdaatun].

The type of pluralization taken by a particular nominal form may be driven by semantic considerations as well. Many active participles, derived from roots mounted on the unaugmented pattern, like [kaatib] may pluralize regularly or irregularly depending on whether they function as a substantive or as an adjective. Used as a substantive to denote a permanent activity or quality, they form a broken plural. Thus when the token [kaatib] is used in the sense of "*author*", it has the broken plural [kuttaab]. By contrast, when it is used in the sense of "*someone who writes*", it pluralizes regularly as [kaatibuuna].

In order to support our claim statistically, we analysed all nouns listed in the "Basic Lexicon of Modern Standard Arabic" (henceforth BLMSA), which consists of the 3000 most frequent words in the language (Khouloughli, 1992). The BLMSA is based on a statistical analysis of more than 200,000 words drawn from newspapers and literary work throughout the Arab world. The author reports a total of 1670 nominal forms (i.e. nouns and adjectives).[4] Of these, 666 tokens are explicitly listed as taking a broken plural and 610 as taking a sound plural (215 masculine and 395 feminine). For the remaining 394 words, the author lists either the plural form (sound or broken) with no mention of the singular or vice versa. The 394 words divide into 357 singular forms for which the corresponding sound plural is not listed, 11 sound plural forms without their relevant singular forms, 20 singular forms without their corresponding broken plurals, and 6 broken plurals for which the corresponding singulars are not listed. Possibly the author lists only the singular or the plural of these forms because the other is not one of the 3000 most frequent words of the language. However, this does not mean that they would be *hapax legomena* in a larger database if this were available. Indeed many of the unlisted words like [murabbaʕaat] "*squares*" and [ʔaaliha] "*gods*" the respective sound and broken plural forms of the listed singular forms [murabbaʕ] "*square*" and [ʔilaah] "*god*" are part of the familiar repertoire of words that can be encountered even in children's books.

In sum, of the 1670 most frequent nominal forms of the language almost two thirds, 978 nouns, pluralize via suffix addition and the remaining forms take a broken plural. This is important for two reasons. First, testing a few random samples taken from the BLMSA shows that it has an average coverage of 75 to 95% of any Modern Arabic text. So if the BLMSA is representative, we can infer that about 56% of Arabic words are nouns (i.e. lexical nouns and adjectives) and most critically that about 59% of all nouns of the language take a sound plural while only 41% take a broken plural. Because BLMSA is a sample of the most frequent words, it is likely that lower frequency nouns are even more skewed towards the regular plural.

In view of this, it seems untenable to consider Modern Standard Arabic as an example of a minority-default system. Just why this stance has come to be held is an offshoot of Arabic lexicographers' work that lists only the broken plural forms because they are unpredictable.

In this section, we have laid out linguistic and corpus-based evidence that the Arabic plural system is not a minority-default. The affixational process involves far more words than the templatic processes, although the proportion

[4] The remaining 1330 items listed in the BLMSA comprise verbs and the closed classes of particles, prepositions and conjunctions.

is still not as high as the English past tense system, with 95% regulars (Daugherty & Seidenberg, 1992).

The Phonological Distribution of Sound and Broken Plurals

The supposed status of the Arabic plural as a minority default system has resulted in claims that it cannot be accommodated by a connectionist model. Plunkett and Nakisa (1997) examined this claim using statistical analyses and connectionist simulations. They noted that a minority default is not necessarily a problem for a connectionist account provided there is an even distribution of regulars and relatively tight clustering of irregulars in the phonological space spanned by the uninflected forms (cf. Hare, Elman & Daugherty, 1995). In cases where irregulars share strong phonological resemblances, but the minority of regulars vary widely in their phonological form, a multi-layered connectionist network can develop "distributional default" behaviour. Although the irregulars may be dominant in number, they are concentrated in relatively small pockets of the network's input space, and so are unlikely to be similar to novel items. Instead, most novel inputs will be more similar to a regular item, and so will be inflected in the same way leading to default behaviour.

Plunkett and Nakisa (1997) examined the phonological distribution of Arabic singulars in this respect using a set of nouns drawn from the Wehr Arabic Dictionary (Wehr, 1976). On the basis of statistical analyses of the distribution of singulars in phonological space, they argued that the Arabic plural system does not provide a basis for developing a distributional default. Instead of evenly spanning the phonological space, the sound plurals appeared to be even more phonologically coherent than many of the broken plural sets. A connectionist network trained on the singular to plural mapping for these items would therefore be unlikely to develop behaviour resembling a default rule.

Plunkett and Nakisa (1997) also showed that despite the absence of the conditions necessary for developing default behaviour, a connectionist model was able to learn and generalise the pluralization task rather well. In fact generalisation (i.e., performance on untrained patterns) in the network was superior to a dual-route model irrespective of the division of labour between the two routes. In effect, the network was performing adequately with neither a majority nor a minority default.

The work of Plunkett and Nakisa (1997) is important because it marks out the conditions necessary for default-like behaviour in a connectionist model of morphological processing. The behaviour of a connectionist system does not just depend on the numbers of regular and irregular items. It also depends on the distribution of these items in phonological space. However, with respect to the specific case of Arabic, there are still many unanswered questions. Since the data-source used by Plunkett and Nakisa (1997) has, as we have argued, a bias in the proportions of sound and broken plurals, the detailed predictions made in their paper may be unfounded. We have already argued that sound plurals are in the majority in Arabic, but this is not enough to demonstrate that a connectionist system will learn to treat them in a default-like way. The phonological properties of a representative sample of the language must also be examined in order to assess the basis for a distributional default. If it turns out that both sound and broken plural classes are phonologically well defined and compact, then a "no default" system would be predicted on the basis of Plunkett and Nakisa (1997).

The 1670 nominal forms were classified by plural type, and the 16 categories that contained 10 or more members were used in the analyses and these amounted to 1491 items. Of these, 972 took the sound plural (273 masculine forms and 699 feminine forms). The remaining 519 items were members of 14 broken plural subtypes, containing between 13 and 121 nouns). In order to examine the phonological similarities between the members of these groups, each singular form was translated into a featural code based on a slight modification of the template system of Plunkett and Nakisa (1997). First, the phonemic transcriptions for the singular forms were aligned to an 18-slot template consisting of alternating consonants and vowels. The slots were filled from left to right, with consonants placed in consonant slots and vowels in vowel slots. When a word contained two consonants or vowels in a row, this procedure led to an empty slot between them, but it also ensured that as far as possible the representations reflected similarities between words by comparing like with like. For example, the representation of /jurħun/ "*scar*" in the template was `jur-HUn----------`. The slot-based phoneme representations were then translated into featural representations in order to capture similarities between different phonemes. The outcome of this transformation was an 18 slot x 20 features (360 dimensional) vector for each singular form. Taking the dataset as a whole, the vectors span a 360 dimensional space, in which each word form is a point. The issue we address is how the different plural classes are distributed in this multidimensional space.

Principal components analysis takes a set of points in a high dimensional space and determines a smaller set of orthogonal vectors within this space that captures the greatest variation between the points. The original points can be projected on to these principal components to extract a low dimensional plot preserving the most important information from the high dimensional space and eliminating redundant dimensions. Figure 1 plots the positions of the different plural subtypes in the space defined by the first three principal components. For the sample used by Plunkett and Nakisa (1997), the sound plurals occupied relatively restricted positions in the space. For our sample, the sound plurals are fairly ubiquitous. There are many completely empty regions of the space, corresponding to phoneme combinations that are in some way badly formed, but most of the occupied regions are occupied by sound plurals, whereas the broken plurals sets are generally more coherent. Plunkett and Nakisa (1997) quantified their observations by calculating a coherence measure for each plural subtype. However, this measure is less valuable for our dataset (containing plural types of greatly varying size) because it is confounded with set size, such that larger sets will be rated as more coherent purely because of their size.

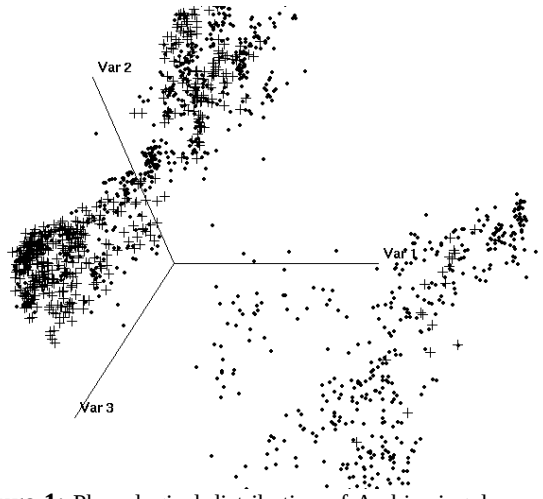

Figure 1: Phonological distribution of Arabic singulars across a plane through the first three principal component. Pluses mark broken plurals, dots mark sound plurals.

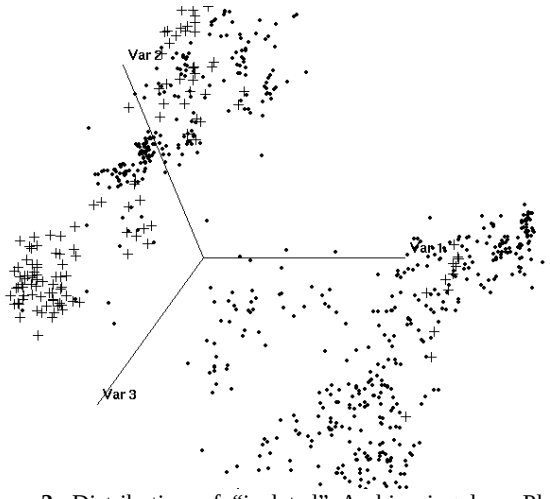

Figure 2: Distribution of "isolated" Arabic singulars. Pluses mark broken plurals, dots mark sound plurals.

Instead, we looked at the relative isolation of the regular and irregular groups as a whole. Put simply, for the regulars to act as the distributional default in a connectionist model there should be a high chance that a randomly chosen non-word will be most similar to one of the existing regulars, and therefore will be processed in the same way. Each word in the language will have it's own "sphere" of influence in the phonological space—if any novel form falls in this area, it will be closest to that point and will tend to be inflected in the same way.

The most influential items in the language will be the ones with the largest area of influence. We can analyse these areas by calculating, for each word in the language, the distance from the nearest neighbour (both of the same class and of any class). The class that exerts the most influence will be the one that has the most isolated members, because these words will have the greatest influence in terms of generalisation to novel forms. This analysis shows that not only are there more sound plurals in Arabic, but they are more spread out in the phonological space, and so have a greater sphere of influence. Sound plurals differ from their nearest neighbour by 4.9 features on average, whereas broken plural differ by 3.7. This advantage is independent of the number of items in each plural class. When nearest neighbour distances are broken down by overall class, the combined effect of numerical dominance and greater area of influence becomes clear, sound plurals differ from their nearest broken plural by 12.2 features on average, whereas broken plurals differ from their nearest sound plural by 6.0 features on average. This statistic implies that it is easy to find sound plurals that are unlike any broken plural but difficult to find broken plurals that are unlike any sound plural. This finding is confirmed in Figure 2, which plots only the singular forms that are 8 or more features different from their nearest neighbour of the opposite class (68% of the sound plurals, and 25% of the broken plurals). The broken plurals are quite closely packed in tight pockets of the space, whereas the sound plurals are more spread out. This is exactly the state of affairs required for distributional default behaviour to develop in a connectionist model.

General Discussion

Much of the evidence relating to the debate between symbolic and connectionist accounts has stemmed from the study of the English past tense, in which regulars are numerically dominant. Proponents of the symbolic account have challenged the ability of connectionist models to deal with inflectional systems in which the default inflection is a minority. Modern Standard Arabic and German were taken as instances of languages that do not depend on the regular pattern involving the majority of forms. Connectionist simulations of minority default behavior (Hare et al., 1995; Plunkett & Nakisa, 1997) have refined the debate, by showing that minority default systems are not necessarily problematic for a connectionist model. If the distribution of regulars is sufficiently broad, then a connectionist model can develop default-like behavior (Hare et al., 1995). Even in the case where regulars are more tightly clustered, a connectionist model can learn the mapping, and perform generalization, although the regular will not become a true default (Plunkett & Nakisa, 1997). These studies emphasize the importance of phonological distribution in the analysis of linguistic systems, alongside the numerical information.

Our main point in this paper was to argue that the Arabic plural system is not a minority default, with regular sound plural applying to fewer forms than the idiosyncratic broken plural. Three sets of arguments were brought to bear on our claim. First, we have shown that while both broken and sound plural are qualitatively productive, only the latter reflects quantitative productivity. Second, the empirical investigation of the most frequent nominal forms collected from BLMSA demonstrates that sound pluralization involves almost twice as many word forms as broken pluralization. The sound plural does not have a low type frequency. Third, analyses of similarities in phonological space showed that the distribution of Arabic nominal forms follow much the same pattern as that of English verbs.

Our analysis raises a set of problems relative to current models of human language productivity. Symbolic models are perfectly compatible with languages exhibiting a minority default inflectional system, but do not provide a princi-

pled explanation for the scarcity of these cases. This follows from the assumption that the human cognitive processor manipulates symbols and does not need a majority of forms to show a rule-based behavior. So far as we know only German and Arabic are cited as current examples of such systems. As it is demonstrated above Arabic is not and Bybee (1995) offered an account that questioned the claim that German is a minority default. Note however, that from the perspective of language change we do not exclude the possibility of a linguistic system passing through a minority default inflectional system. Rather, our point is: if minority default systems are as natural and as easy to handle as symbolic models would have it, then why do they seem to be scarce?

Connectionist models, meanwhile, have responded to the challenge of the minority default. These systems are less at ease with a minority default system, since they require the regulars to have sufficient variety in their phonological form if they are to be treated as the default case. But more critically, they also offer an explanation for the lack of minority defaults in most modern languages. Hare and Elman (1995) used connectionist networks to model the diachronic changes in the verb system of Old English, which at some stage is likely to have been a minority default system. Developments in the structure of language were assumed to be the product of imperfect learning from generation to generation, modeled by generations of connectionist networks. In essence, the development of the language was one of regularization, with regulars becoming more and more dominant in each successive generation. Thus, minority defaults can be learned by a connectionist network as long as certain distributional conditions are met. Even when those conditions are met, however, the state of the language is somewhat unstable, with a diachronic movement towards majority default likely in the long term. This fits in with the observation that the vast majority of linguistic systems—including the Arabic plural—do not employ a minority default.

Acknowledgments

We thank William Marslen-Wilson and Matt Davis for valuable comments on an earlier version of this article.

References

Aronoff, M. & Anshen, F. (1998). Morphology and the lexicon. In A. Spencer & A. Zwicky (eds.), *The Handbook of Morphology*, Blackwell Publishers.

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

Bohas, B. & Guillaume, J. P. (1984) *Etudes des Théories des grammairiens Arabes: I Morphologie et Phonologie*: Damas, Syria.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425-455.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991-1060.

Daugherty, K. & Seidenberg, M. (1992). Rules or connections? The past tense revisited. *Proceedings of the 14th annual Conference of the Cognitive Sciences Society*, 259-264.

Hare, M. & Elman, J. (1995). Learning and morphological change, *Cognition, 56*, 61-98.

Hare, M. Elman, J. & Daugherty, K. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes, 10*, 601-630.

Holes, C. (1995). *Modern Arabic*. Longman, London-NY.

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

Marslen-Wilson, W. D., Tyler, L.K., Waksler, R. and Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3-33.

Marslen-Wilson, W. & Tyler, L. K. (1998). Rules, representations and the English past tense. *Trends in Cognitive Sciences, 2*, 428-435.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry, 12*, 373-418.

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

Moussa, A. H. (1996) Database for major Arabic dictionaries. *Proceedings of the 5th International Conference and Exhibition on Multilingual Computing*, Cambridge.

Pinker, S. & Prince, A. (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73-193.

Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron: implications for child language acquisition. *Cognition, 38*, 3-102.

Plunkett, K. & Nakisa, C. (1997). A connectionist system of the Arabic plural system. *Language and Cognitive Processes, 12*, 807-836.

Ravid, D., Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language, 19*, 187-206.

Rumelhart, D. & McClelland, J. (1986). On learning the past tense of English verbs. In D. Rumelhart & J. McClelland (eds*) Parallel distributed Processing, Vol. 2*. Cambridge MA: MIT Press.

Xasaara, M. (1994). *Translation and Linguistic Development*. Damascus. (in Arabic).