LANGUAGE AND COGNITIVE PROCESSES, 2002, *17* (3), 321–343

A re-examination of the default system for Arabic plurals

Sami Boudelaa
MRC Cognition and Brain Sciences Unit, Cambridge, UK

M. Gareth Gaskell
Department of Psychology, University of York, York, UK

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 and Prince (1990), the plural system of Modern Standard Arabic has been cited as an archetype of a minority-default system with the affixational sound plural involving fewer nominal forms than the templatic 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 are qualitatively productive in the sense of being subject to a number of constraints or conditioning factors, the latter is quantitatively the more productive process and involves more nominal forms. Furthermore, the diversity of the phonological forms taking a sound plural 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 limited means to produce effectively

Requests for reprints should be addressed to Sami Boudelaa, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.
Email: sami.boudelaa@mrc-cbu.cam.ac.uk

We are indebted to William Marslen-Wilson for his help with an earlier version of the paper. We also thank Matt Davis, Robert Ratcliffe, and Ewa Dabrowska for helpful comments. This article is based on a paper presented at the Twenty-second Annual Conference of the Cognitive Science Society, Philadelphia, PA.

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960143000245

unlimited combinations of words and utterances. In order to deal with this generic question, several more specific questions need to be spelt out. One such 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 (Butterworth, 1983; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Pinker 1991). Focusing 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 (Marcus, 1998; Marslen-Wilson & Tyler, 1998; Pinker, 1991; Pinker & Prince, 1988; Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986).

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 tense in English as in ''walk-walked'', is formed by a rule, whereas irregular past tenses like ''eat-ate, give-gave'' are learned by rote (Berko, 1958; MacKay, 1978). Because it fails to explain the sub-regularities among the irregular verbs and the generalisation of irregular inflection to phonologically similar nonce forms (Bybee & Moder, 1983), 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 (Clahsen, 1999; Marcus, Brinkemann, Clahsen, Wiese, & Pinker, 1995; Pinker, 1991; Pinker & Prince, 1988; Prasada & Pinker, 1993). The third and perhaps most radical model is the connectionist one, which dispenses with rules and assumes that language learning is better accounted for using a single mechanism, namely a network of highly interconnected units (MacWhinney & Leinbach, 1991; Rumelhart & McClelland, 1986). Both regular and irregular forms are inflected using a pattern associator and no separate default process is assumed to exist to deal with regular or novel forms. On this account, the network's response to a novel form depends on that item's phonological similarity to already experienced patterns (Plunkett & Marchman, 1991). It is worthy of note that non-connectionist models like the network model (Bybee, 1985), or the Analogical Model of Language (Skousen, 1989) also account for morphological processing within a single mechanism. However, here we focus on the connectionist approach to inflectional morphology, and the

debate it has sparked with the symbolic accounts in so far as the Arabic plural system is concerned.

Both dual 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 the "default" is regular both descriptively and psychologically: descriptively, because the lexicon is positively skewed towards regular forms with 95% of the verbs in the language taking the ~ed regular suffix and psychologically because speakers tend to generalise to this pattern as in "*fax-faxed*, *xerox-xeroxed*" (Marcus et al., 1995; Ullman, 1999). Accordingly, this is a relatively simple situation for a dual-route model, as it would easily deal with the low number of irregulars via associative memory and the rest via a default rule. A connectionist network will also exhibit relative ease handling such cases. The associative network will store information about all forms and the preponderance of regular forms will 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 (Rumelhart & McClelland, 1986). Proponents of the dual route model have argued, however, that a dual mechanism can also deal satisfactorily with linguistic systems where the default is a minority as is the case of the German participle ~t and the plural ~s (Marcus et al., 1995). This is because rule-like behaviour need not 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 merely depends on applying the same procedure to different items bearing the same symbol "Verb" (Clahsen, 1999; Marcus et al., 1995). Conversely, a connectionist network was predicted to be unable to simulate people's regularisation of novel forms in languages which have a minority-default.

Along with the German inflectional system, the Arabic plural is the most widely cited example of a minority default system (Hare, Elman & Daugherty, 1995; McCarthy & Prince, 1990; Pinker & Prince, 1994; Ravid & Farah, 1999). For this reason it was used as a test case by Plunkett & 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 in the present research is to take issue with the position that Arabic has a minority default plural system and to show that it hinges on an inaccurate account of the Arabic plural system. 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 (although not as extreme as the English past tense system). We conclude by discussing why minority default systems seem to be scarce across world languages.

ASPECTS OF MODERN STANDARD ARABIC MORPHOLOGY

It is widely accepted that the Arabic language has a morphology that 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 [kalb] "*dog*" the verb [kaliba] "*get infected with rabies*" can be formed. The second part relates to verb morphology and subsumes verbs proper and derived nouns (i.e., nouns that are derived from verbs). Verbs are further divided into unaugmented and augmented verb forms. The unaugmented forms comprise the three patterns {faʕal}, {faʕil}, {faʕul}, where the letters /f, ʕ, l/ indicate the slots to be filled by the three radical letters of a root. These three forms are unaugmented in the sense that they are made up of the minimal phonetic material necessary for a form to surface (i.e., the two vowels of the word pattern plus the consonants of the root). An example of an unaugmented surface form is [katam] "*conceal*" where the abstract root {ktm} is combined with the pattern {faʕal}. Augmented forms consist of 14 patterns of which only 9 are frequent in Modern Arabic. These are {faʕʕal}, {faaʕal}, {ʔafʕal}, {tafaʕʕal}, {tafaaʕal}, {ʔinfaʕal}, {ʔiftaʕal}, {ʔifʕall}, {ʔistafʕal}. They are augmented because they contain additional consonantal and/or vocalic material over and above that required to become a surface form. For instance, the surface form [takattam] "*keep mum*", which is formed by combining the root {ktm} with the pattern {tafaaʕal}, is an augmented form as it contains the epenthetic initial syllable /ta/ and its second radical consonant /t/ is geminated. It is estimated that as many as 400 surface forms can be derived from certain roots (Xasaara, 1994). For example, combining the root {xrʒ} "*going out*" with the pattern {faʕal}, produces the form [xaraʒ] "*go out*". The same root can be further combined with as many as five augmented verb patterns yielding the following surface forms: [xarraʒ] "*move out*", [ʔaxraʒ] "*take out*", [taxarraʒ] "*graduate*", [taxaaraʒ] "*disengage*", [ʔistaxraʒ] "*extract*".

As for nominal morphology, there are about eight types of deverbal nouns such as the active participle, the passive participle, the instance noun, a noun denoting that the action expressed by the verb takes place only once, and the noun "masdar" (Holes, 1995; Wright 1995). The number of nominal word patterns is well over a hundred (El-Dahdah, 1990). A deverbal noun is a surface form consisting of a word pattern and a

root (Cohen, 1961; Hilaal, 1990). For example, the masculine active participle [xaariʒ] ''*someone who goes out*'', which comprises the word pattern {faaʕil} and the root {xrʒ} is derived from the unaugmented surface form [xaraʒ] ''*go out*''. Also, the masculine active participles [muxarriʒ], [muxriʒ], [mutaxarriʒ], [mutaxaariʒ], [mustaxriʒ] can be respectively derived from the following augmented verb forms [xarraʒ], [ʔaxraʒ], [taxarraʒ], [taxaaraʒ], [ʔistaxraʒ]. Passive participles can also be formed from these verb forms. In addition to this, an ''*instance noun*'', [xarʒa] ''*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] ''*rave*'' is derived and from the latter an active participle [mutakaalib] ''*someone who raves*'' and a ''masdar'' [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*''.

In view of the high productivity of the two components of verb morphology, it is important to consider the implication this may have for the inflectional plural system in the language.

The qualitative productivity of sound and broken plurals in MSA

Apart from case endings, Arabic nominal forms undergo various morphological alterations of which the most frequent is pluralisation. This is achieved either via suffixation or pattern modification. In the first case, known as sound pluralisation, the suffix ~uun is added to citation forms of masculine nouns (e.g., [naaʒiħ-naaʒiħuun] ''*successful*'' male) while ~aat is appended to citation forms of feminine nouns (e.g., [naaʒiħa-naaʒiħaat] ''*successful female*''). Note that while the masculine sound plural suffix ~uun is added directly to the noun stem [naaʒiħ], the feminine sound plural suffix ~aat is added to [naaʒiħa] after removing the feminine singular suffix ~a (Holes, 1995). If the noun in question does not end in ~a as is the case of [marjam], ''*proper name*'', [ramadʕaan] ''*fasting month*'', and [taʕriif] ''*definition*'', the plural suffix ~aat is directly added to yield the following plural forms respectively [marjamaat], [ramadʕaanaat], and [taʕriifaat].

In the second type of pluralisation, often referred to as broken pluralisation, the pattern of the singular noun is more dramatically altered as in [ʕunquud]-[ʕanaqiid] ''*cluster-clusters*'', and in some cases some of its consonants are lost as in [ʕandaliib]-[ʕanaadil] ''*nightingale-nightingales*'', where the radical letter /b/ of the singular is lost (Levy & Fidelholtz, 1971;

Murtonen, 1964; Ratcliffe, 1998). We refer to sound and broken pluralisation as regular and irregular inflectional processes respectively, not because we think that the former is rule-based and the latter is not. Rather we use the term *regular* as shorthand for an inflectional process involving little or no allomorphy and the term *irregular* to describe a process entailing substantial modification of the singular input. Note, however that if the term *regular* is used to denote the systematic or consistent nature of a given morphological process, then both sound pluralisation and at least some subclasses of broken pluralisation can be described as regular by virtue of their consistency (Ratcliffe, 1998).

McCarthy and Prince's (1990) work on the broken plural in Modern Standard Arabic has promulgated the idea of Arabic having a minority default system of pluralisation. According to them "*essentially all canonically-shaped lexical nouns of Arabic take broken plurals*", while 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 in this way, the above claim is misleading because it does not distinguish between quantitative and qualitative productivity. Fundamentally, productivity is a graded phenomenon that is subject to different conditioning factors; and the distinction between its qualitative and quantitative aspects helps capture this gradation. By describing a morphological process as being qualitatively productive we mean that it is subject to various conditioning factors which can be phonological, semantic, syntactic or even pragmatic in nature (Anshen & Aronoff, 1999; Aronoff, 1976; Baayen, 1992; Bauer, 1983). Nevertheless, being subject to conditioning factors does not rule out the possibility of a particular inflectional operation being productive for a constrained set of forms (Aronoff & Anshen, 1998; Ratcliffe, 1998). A quantitatively productive process on the other hand is one that is subject to fewer constraints and consequently applies to larger numbers of items in the language. An English example may help bring this point home: The suffix ~ity and its rival suffix ~ness are respective instances of qualitative and quantitative productivity. Indeed, the suffix ~ity, used to convert adjectives into nouns is qualitatively productive in the sense that it applies to the majority of adjectives ending in suffixes like ~ible, ~able, ~ic, ~id. Conversely its rival suffix ~ness is quantitatively productive because it is subject to fewer constraints (Aronoff & Anshen, 1998).

Now going back to the plural system of MSA, the picture that emerges once we consider the characteristics of sound and broken pluralisation is arguably one of qualitative productivity in both cases. As correctly pointed out by McCarthy and Prince (1990), sound pluralisation (particularly the

masculine plural) is restricted to a set of nominal forms that must meet formal and probably syntactic criteria. Being so restricted, sound or suffixal pluralisation is qualitatively productive. Broken pluralisation is also qualitatively productive since it tends to apply mostly to short primitive nominal forms comprising two to three consonants and to lexicalised derivatives. For example, the plural template [fuʕʕaal] requires the singular nominal form to be lexicalised and to have the pattern [faaʕil] before it can apply to it (Ratcliffe, 1998). In sum, both sound and broken pluralisation are subject to a host of conditioning factors that make them qualitatively productive. It remains to be determined, however, which of these two types of pluralisation is quantitatively the productive process.

The quantitative productivity of sound and broken plurals in MSA

A given triliteral root in Arabic can be productively mounted on some combination of the nine frequent augmented word patterns to create new words. For instance, the triliteral unaugmented surface form [katab] "*write*" can be combined with seven augmented forms, whereas 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, it follows that each of the augmented forms gives rise to at least four deverbal forms. Being transparently derived, all of these forms will take a sound plural.

According to Moussa (1996), who recorded the roots of Arabic in Taj Al-Arous, one of the major dictionaries of the language, there are 11,978 roots of which 7597 are triliterals, 4081 are quadriliterals and 300 are quinquiliterals. Assuming that the derivation of at least four deverbal surface forms from each root is not an overestimate, the triliteral roots alone would yield $7597 \times 3 \times 4 = 91{,}164$ surface forms that take a sound plural. If we consider the rest of the derivatives from quadriliteral and quinquiliterals, this estimate will quickly increase. A possible counter-argument might be that the predicted figure is inflated because some derivatives like assimilated nouns often pluralise in the broken way and that transparent derivatives take a broken plural once they are lexicalised. This argument can be easily countered if we take into account the fact that for almost every assimilated noun or 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 [qirda] "*female monkey*" has a sound plural form [qirdaat]. Accordingly, there is a form that pluralises in the suffixed regular way for almost every form that has an irregular broken plural, but the reverse is not true. In addition to this, the type of pluralisation taken by a particular nominal form may be driven by semantic considerations as well. Many active participles (e.g., [kaatib] "*write*r") that are derived from an unaugmented verb form may pluralise 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 pluralises regularly as [kaatibuun]. This demonstrates that the number of surface word forms taking a sound plural in MSA outnumbers that of surface forms taking a broken plural. Accordingly sound pluralisation is quantitatively the productive process in MSA despite its being subject to conditioning factors. We now turn to present some statistical evidence to corroborate this claim.

STATISTICAL INVESTIGATION OF THE MSA PLURAL SYSTEM

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).[1] Of these, 666 word forms 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 352 singular forms that take a sound plural, 16 sound plural forms, 20 singular forms that take a broken plural, and 6 broken plural forms. Possibly the author lists only the singular or the plural of these forms because the unlisted form is not one of the 3000 most frequent words of the language. However, this does not mean that they

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

would be *hapax legomena* in a larger database if this were available. Indeed, for 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.

Some of the words that Khouloughli (1992) did not list happened to be part of a 1200 word set used in a familiarity judgement pretest that the first author conducted for another study. In this pretest 15 native users of Arabic were asked to rate words for their familiarity on a one to five point scale, with one being rare and five very familiar. The results showed that the set of words which Kouloughli did not list, and which happened to be in the pretest, were actually given high familiarity ratings. To take an example the unlisted broken plural [ħuðʕuuðʕ] "*chances*" received an average rating of 4.22 on a 5-point scale with 1 being *least familiar* and 5 *most familiar*. Similarly, the unlisted sound plural forms [ʔiħtifalaatun] "*celebrations*", and [ʔiħtiyaaʒaat] "*needs*" were respectively rated 4 and 4.50.

Therefore, in the BLMSA the total number of nominal forms taking a sound plural is 978 (i.e., the 610 forms for which both singular and sound forms are listed, the 352 singular forms for which the corresponding sound plural forms are not listed, and the 16 sound plural forms for which the relevant singular forms are not given). In other words, about 59% of the 1670 most frequent nominal forms pluralise via suffix addition and the remaining forms, around 41%, take a broken plural. Although the BLMSA is a relatively limited database containing only the 3000 most frequent words of the language, the figures it offers with respect to nouns taking a sound or a broken plural are reliable. This is so because when the representative aspect of the BLMSA was tested by assessing the average number of words it covers in randomly chosen samples of Arabic texts, the results revealed that it contained "*practically the entirety of the grammatical tools used in texts written in Modern Standard Arabic, as well as a widely representative sample of the lexical units most frequently met in non technical modern texts*" (Kouloughli, 1992, p. 12).

So if the BLMSA is representative, we can infer that about 59% of all nominal forms of the language take a sound plural while only 41% take a broken plural. This is likely to be an underestimate of the prevalence of the sound plural. The BLMSA is a sample of the most frequent words, so given that irregulars are generally common words it is likely that lower frequency nouns are even more skewed towards the regular plural.

A further source of evidence is Murtonen's (1964) extensive study of broken plural formation in Arabic. This study consisted of a statistical survey of all the nominal forms taking a broken plural listed in Lane's

dictionary, excluding those beginning with the glides /w, j/.[2] Murtonen (1964) reports the existence of 9540 nominal forms taking a broken plural. This figure is an order of magnitude less than our estimate of the number of nominal forms pluralising regularly (over 90,000) even when we limit ourselves – as noted above – to the active and passive participles derived from triliteral roots. Including all the nominal forms beginning with the glides /w, j/ would definitely have raised the overall number of forms with an irregular plural, but never to the extent of outnumbering those pluralising regularly.

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 of their high degree of unpredictability. The Wehr Arabic Dictionary (Wehr, 1976) on which McCarthy and Prince (1990) and Plunkett and Nakisa (1997) have relied is no exception. It hardly lists any of the sound plurals of derivatives. Other recent frequency lists of Arabic typically count masculine sound plural and feminine sound plural as different tokens related to the same singular type, but consider broken plural forms as different types with respect to their singular forms and list them (Abdah, 1979).

In this section, we have laid out linguistic and corpus-based evidence that the plural system of Modern Standard Arabic is not a minority-default. The affixational process involves far more words than the templatic processes, although the proportion of regulars in the system is still not as high as the English past tense system, with 95% regulars (Daugherty & Seidenberg, 1992). We now turn to an investigation of the distributional structure of MSA nominal forms applying the method of principal components analysis to the set of material collected from the BLMSA.

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 (Pinker & Prince, 1994). Plunkett & 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

[2] Murtonen (1964) does not state the reasons why nominal forms beginning with a glide are excluded from his survey.

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 pluralisation 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 (their simulations used a dataset with just 24% sound 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).

Analysis of Arabic nouns

The 1670 BLMSA nominal forms were classified by plural type, and the 16 categories that contained 10 or more members were used in the analyses (1491 items; see Table 1). 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 /ʒurħ/ "*scar*" in the template was jur-H_____. In fact, most of the representations only used empty slots at the end of the word. The slot-based phoneme representations were then translated to featural representations in order to capture similarities between different phonemes (see Appendix A for details of the phonemes and features used). The outcome of this transformation was an 18 slot $\times$ 20 features (360 dimensional) vector for each singular form. Taking the data set as a whole, the vectors occupy a 360 dimensional space, in which each

TABLE 1
Frequency of different plural types. MSP and FSP stand for masculine and feminine sound plural respectively. Broken plurals are transcribed using the appropriate word pattern and numbered 1–14

Plural Type	*Frequency*	*Label*
~uun	273	MSP
~aat	699	FSP
/ʔafʕaal/	121	BP1
/fuʕuul/	79	BP2
/mafaaʕil/	56	BP3
/fawaaʕil/	40	BP4
/fiʕaal/	36	BP5
/faʕaaʔil/	33	BP6
/fuʕaalaaʔ/	32	BP7
/fuʕal/	25	BP8
/ʔafʕila/	24	BP9
/fuʕul/	17	BP10
/fuʕʕaal/	15	BP11
/mafaaʕiil/	15	BP12
/fiʕal/	13	BP13
/ʔafʕilaaʔ/	13	BP14

word form is a point. The issue we address is how the different plural classes are distributed in this multidimensional space.

Principal components analysis (PCA) takes a set of points in a high dimensional space and determines the smaller set of orthogonal vectors within this space that captures the greatest variation among 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 a plane

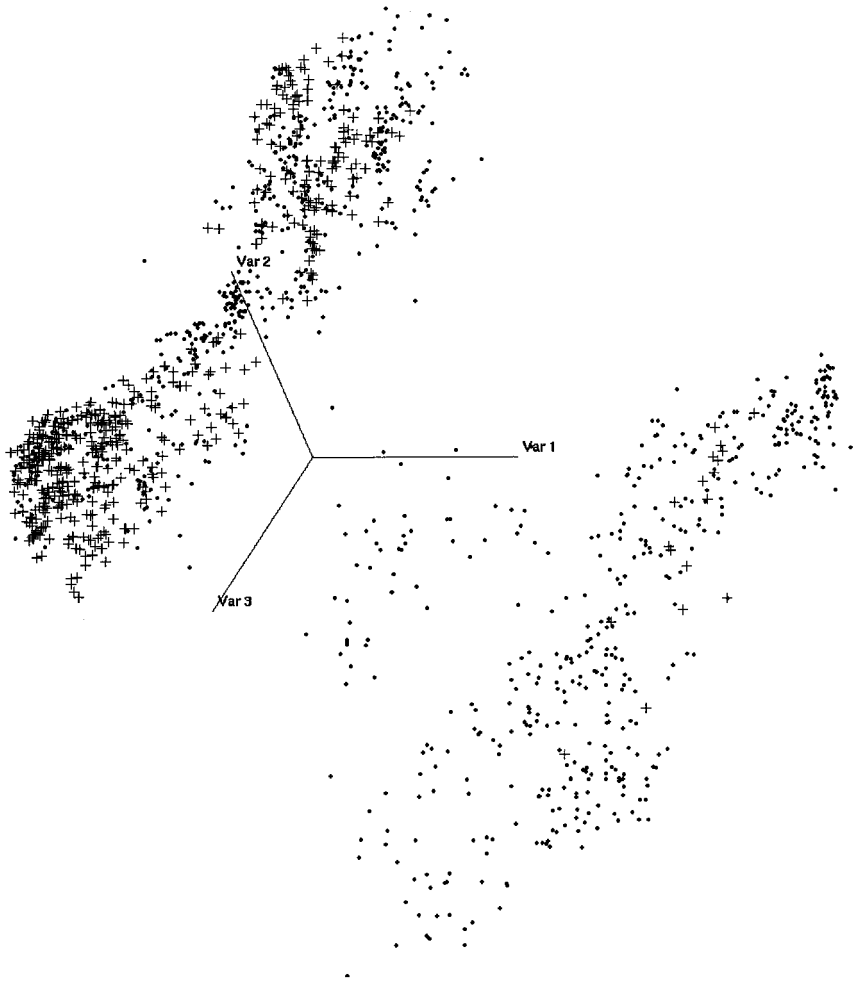

Figure 1. Phonological distribution of Arabic singulars across a plane through the first three principal components. Pluses mark broken plurals, dots mark sound plurals.

through the space defined by the first three principal components (i.e., the three dimensions capturing the greatest variation). 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 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 regions that are occupied at all 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 generally be rated as more coherent purely because of their size.[3] Instead, we looked at the relative isolation of the regular and irregular groups as a whole. Simplifying the situation somewhat, for the regulars to act as the distributional default in a connectionist model there should be a high chance that a randomly chosen nonword will be most similar to one of the existing regulars, and therefore will be processed in the same way. Each word in the language (point in the space) will have its 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.[4]

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

[3] The coherence measure used by Plunkett and Nakisa (1997, pp. 820–821) was based on the ratio of nearest neighbours within and between sets. Larger sets are more likely to have a member near any given point and will therefore lead to lower nearest neighbour distances. Monte Carlo analyses of randomly generated sets showed high coherence values for large sets and low coherence values for small sets.

[4] This is a greatly simplified analysis of the functioning of a connectionist model, since the degree of influence will depend on the consistency between multiple neighbours and frequency of presentation. As a first approximation, however, it is valid. It should also be noted that this characterisation is compatible with the ''gang effect'' of irregular clusters (e.g., keep-kept, sleep-slept, weep-wept). These clusters are influential as a group within a tightly constrained phonological space. However, the influence of any individual item within these groups is limited. What is required for a distributional default is a wide distribution of one class of words (Plunkett & Nakisa, 1997).

neighbour by 4.9 features on average, whereas broken plurals 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.[5] 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. Novel items bearing a phonological similarity to the clusters of irregular nouns will tend to be inflected in the same way, whereas all other novel forms will take the sound plural. We should stress that the situation in Arabic is less extreme than in English, and that Arabic irregulars should be substantially more influential than their English counterparts. This facet is reflected by the sections of phonological space in Figure 2 that are dominated by broken plurals. These clusters will lead to some productivity and generalisation, analogous to the qualitative productivity of broken plurals referred to above (although our pure phonological analysis will fail to capture some of the constraints involved in this productivity). Nonetheless, our basic claim remains that Arabic pluralisation cannot be counted as a minority default system.

Relation to Hebrew data

The analyses of sphere of influence emphasise an important point about phonological distribution of morphological systems. The singulars that take the sound plural are generally more sparsely distributed than the singulars taking the broken plural, but this difference is quite small. The dominance of the sound plural only becomes clear when the size of the two plural groups is taken into account. Berent, Pinker, and Shimron (1999) have argued on the basis of distributional data that nominal inflection in Hebrew is incompatible with a connectionist approach. Although the

[5] Examples of near neighbours (minimal pairs) are [saaħa] "*play ground*" and [saaʕa] "*watch*", or [ʔamiir] "*prince*" and [ʔamiin] "*trustworthy*". [tiknuluuʒiaja] "*technology*", [ʕusbuuʕijj] "*weekly*", [ʔimbriɟaalijj] "*imperial*", and [ʔistraatiʒijj] "*strategic*" are examples of particularly isolated singular forms according to these analyses (differing from their nearest neighbour by more than 20 features).

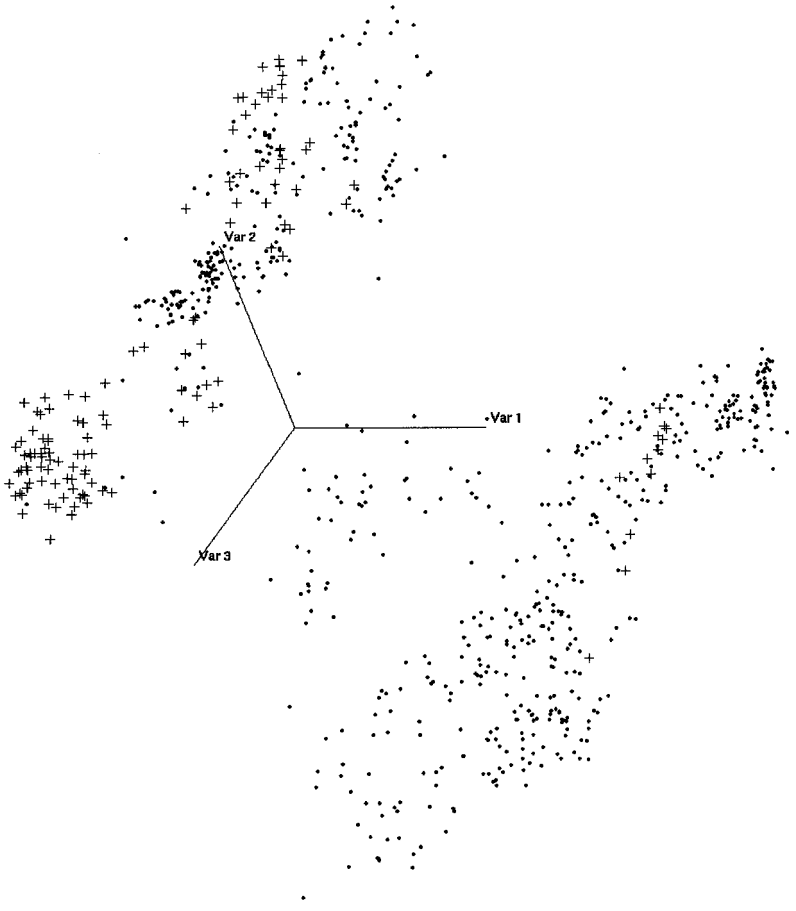

Figure 2. Distribution of ''isolated'' Arabic singulars. Pluses mark broken plurals, dots mark sound plurals. Only singulars that differ from their most similar neighbour in the opposite class by 8 or more features are plotted.

majority of Hebrew nouns are pluralised regularly, Berent et al. (1999) argued that the regulars are not sufficiently diffuse in terms of their phonological distribution for a connectionist account to treat them as a distributional default. However, in our view their arguments are flawed on a number of counts.

First their distributional analyses were based on a rather limited representation of phonological similarity. Their classification involved calculating the proportions of regular and irregular (masculine sounding) nouns in each of the mishkalim[6] in the language. They found that almost

[6] Mishkalim (singular: mishkal) are the Hebrew equivalents of Arabic word patterns.

all of the 91 mishkalim contained both regular and irregular nouns, and that most mishkalim were dominated by regulars. This was taken as evidence that regulars were more coherently clustered than irregulars (because mishkalim containing regulars are normally dominated by regulars whereas mishkalim containing irregulars are normally not dominated by irregulars). However, this type of distributional analysis is very restricted, compared to our approach. The space defined by membership of the mishkalim has in effect 91 points that can be legally occupied (i.e., a noun can be a member of one of the 91 mishkalim). In contrast, the phonological space we analysed contained 360 binary dimensions, corresponding to 2^{360} (roughly 10^{108}) potential positions that can be occupied by a noun. Admittedly, the majority of these will not be phonologically well formed, but even if only one in 10^{50} of these points is well formed this still leaves a vast range (10^{58}) of potential noun positions. In short, in order to get a good idea of the distribution of items in phonological space, one needs a good representation of phonological space. Simple classification into word patterns does not provide that.

A second point about the Berent et al. (1999) analysis is that their measure of coherence was confounded by group size. Of the 1971 nouns they analysed, 90% were regular, which meant that the majority of the mishkalim were guaranteed to be dominated by regulars. So simply showing that regulars dominate most mishkalim tells us nothing about the relative coherence of regulars and irregulars as Berent et al. (1999) wished to claim.

A final point about the Hebrew analysis was that Berent et al. (1999) were operating on the assumption that distribution alone is important in deciding whether a connectionist model can accommodate a particular morphological system. Although recent research has emphasised the importance of distribution (e.g., Hare et al., 1995; Plunkett & Nakisa, 1997), this has not had the effect of replacing type frequency as the sole relevant factor. Instead they are both significant factors in determining the response of a connectionist model to a particular system. This is underlined in the nearest neighbour analyses above, where slightly sparser distribution of regulars combined with a numerical dominance to produce a large difference in the average similarity between an Arabic singular and its nearest neighbour from the opposite class. Thus, the strong numerical dominance for regulars in the case of Hebrew nouns cannot be disregarded, as Berent et al. (1999) appear to do.

SUMMARY AND GENERAL DISCUSSION

Our starting point was the distinction between symbolic and connectionist accounts of generalisation and how these embody different approaches to

human cognition, leading to divergent predictions about the language processor. Much of the evidence relating to this debate 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 behaviour (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 behaviour (Hare et al., 1995). Even in the case where regulars are more tightly clustered, a connectionist model can learn the mapping, and perform generalisation, although the regular will not become a true default (Plunkett & Nakisa, 1997). These studies emphasise the importance of phonological distribution in the analysis of linguistic systems, alongside the numerical information. Our main point, however, was to argue that the MSA 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 both broken and sound plurals are qualitatively productive in the sense that they will apply to a particular singular form only if it satisfies certain constraints. For example, broken pluralisation requires the singular input to be a lexical noun having a specific shape or word pattern. Masculine sound pluralisation requires the singular input to be masculine and animate. Second, the empirical investigation of the most frequent nominal forms collected from the BLMSA demonstrates that regular sound pluralisation involves almost twice as many word forms as the irregular broken plural. The sound plural does not have a low type frequency and is the quantitatively productive pluralisation process. 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, although the irregular classes will be more influential in Arabic than in English.

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 principled 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 behaviour. So far as we know only German and Arabic are cited as current examples of such systems. We have argued that this is not

true of Arabic, and Bybee (1995, 1999) offers an account that questions the claim for German (see also Clahsen, 1999, and subsequent commentaries). 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 appear to be scarce? Unless the apparent scarcity of minority defaults systems is due to some kind of sampling bias, the challenge for symbolic models is to provide a learning model that can explain why majority defaults are preferred.

Connectionist models, meanwhile, can accommodate minority defaults, but are less at ease with them, since they require the regulars to have sufficient variety in their phonological form 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, modelled by connectionist learning networks. Put simply, the development of the language was one of regularisation, 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, it seems, the Arabic plural – do not employ a minority default.

Manuscript received June 2000
Revised manuscript received May 2001

REFERENCES

Abdah, D.A. (1979). *Frequent words in Arabic*. Riyad University, Saudi Arabia (in Arabic).

Anshen, F., & Aronoff, M. (1999). Using dictionaries to study the mental lexicon. *Brain and Language*, 68, 16–26.

Aronoff, M., & Anshen, F. (1998). Morphology and the lexicon. In A. Spencer & A. Zwicky (Eds.), *The handbook of morphology*. Oxford: Blackwell Publishers.

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

Baayen, H. (1992). Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1991* (pp. 109–149). Dordrecht: Kluwer.

Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.

Berent, I., Pinker, S., & Shimron, J. (1999). Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition*, *72*, 1–44.

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

Bohas, G., & Guillaume, J.-P. (1984). *Etudes des Théories des grammairiens Arabes: I Morphologie et Phonologie*. Damascus, Syria: Publication de l'Institut Français de Damas.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (pp. 257–294). London: Academic Press.

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

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

Bybee, J. (1999). Use impacts morphological representation. *Behavioural and Brain Sciences*, *22*, 1016–1017.

Bybee, J., & Moder, C. (1983). Morphological classes as natural categories. *Language*, 59, 251–270.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioural and Brain Sciences*, *22*, 991–1013.

Cohen, D. (1961). Le vocabulaire de base sémitique et le classement des dialectes du sud. *Semitica*, *11*, 55–84.

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.

El-Dahdah, A. (1990). *A dictionary of universal Arabic grammar*. Librairie du Liban.

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.

Hilaal, Y. (1990). Deriving surface forms from roots and word patterns. *Linguistica Communicatio*, *1*, 77–80 (in Arabic).

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

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

Levy, M., & Fidelholtz, J. (1971). Arabic broken plurals, rule features and lexical features. *Glossa*, *5*, 57–70.

MacKay, D. (1978). Derivational rules and the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *17*, 61–71.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*, 121–157.

Marcus, G. (1998). *The Algebraic mind: reflections on connectionism and cognitive science*. Cambridge, MA: MIT Press.

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

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

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

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.

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

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

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

Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S.D. Lima, L.R. Corrigan, & G.K. Iverson (Eds), *The reality of linguistic rules*. Amsterdam/ Philadelphia: John Benjamins.

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., & Marchman, V. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, *8*, 1–56.

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

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

Ratcliffe, 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.

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.

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

Ullman, M. (1999). Acceptability ratings of regular and irregular past tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes*, *14*, 47–68.

Wehr, H. (1976). *Arabic–English dictionary: The Hans Wehr dictionary of modern written Arabic*, 3rd edition, Ithaca, N.Y.

Wright, W. (1995). *A grammar of the Arabic language*. Cambridge: Cambridge University Press.

Xasaara, M. (1994). *Translation and linguistic development*. Damascus, Syria: Al-alhaali li-ttibaʕa wa-annashr wa-ttawziiʕ.

APPENDIX A

Featural representation used to encode the set of Arabic phonemes

	Labial	Labiodental	Dental	Interdental	Alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal	Nasal	Emphatic	Consonantal	Voiced	High	Low	Back	Front	Long	Short
i	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1
u	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1
a	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1
I	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0
m	1	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0
b	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
f	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
£	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
^	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
t	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
d	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
s	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
z	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
$	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0
j	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0
T	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0
Z	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0
D	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0
S	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0
l	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
n	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0
r	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0
k	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
q	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
x	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
G	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
&	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0
?	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
h	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
w	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
y	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0

APPENDIX A—*cont.*

The ASCII symbols used to represent some of the Arabic phonemic sounds

ASCII symbols	*IPA equivalent*	*Example word*
U	/uu/	[quruudun] "*monkeys*"
A	/ææ/	[xæælin] "*deserted*"
I	/ii/	[diinun] "*religion*"
T	/tˤ/	[tˤiinun] "*clay*"
Z	/dˤ/	[dˤariibatun] "*tax*"
D	/ðˤ/	[ðˤillun] "*shade*"
S	/sˤ/	[sˤumuudun] "*resistance*"
H	/ħ/	[ħablun] "*rope*"
&	/ʕ/	[ʕadlun] "*fairness*"
G	/ɣ/	[ɣ]imdun] "*dheath*"
?	/ʔ/	[ʔamiirun] "*prince*"
j	/ʒ/	[amratun] "*ember*"
$	/ʃ/	[ʃayxum] "*fat*"
£	/θ/	[θamanun] "*price*"
^	/ð/	[ðahabun] "*gold*"