

## Adults assigning gender: Loanwords and L2 acquisition

Mary Ann Walter  
walterma@mit.edu  
MIT

### 0. Introduction

- In assigning grammatical gender to unfamiliar lexical items, the same L1-based biases come into play for both L2 learners and loanword borrowers.
- L2 learners must assign it in such a way as to conform to the target language (L2)
- Borrowers must modify the source language form to their L1.
- I present here a case study of each scenario:

Study 1: L2 Arabic plurals by English speakers

Study 2: Arabic loanwords into Spanish

- In both, speakers make morphophonological generalizations to aid them in the gender decision.
- However, the application of these is modulated by an effect of lexical statistics.
- This suggests a tendency toward probability-matching in adult language learners, in opposition to rule formation (as shown by Hudson and Newport 2005; Wonnacott and Newport 2005).

### 1. L2 Arabic plurals

- English speakers begin with a language largely lacking morphological gender.

Influence of L1 gender on L2 acquisition should be minimal.

The complexity of Arabic presents a major challenge.

Its gender morphology surfaces both in adjectival agreement and pluralization patterns.

#### 1.1 Arabic pluralization

##### 1) Masculine suffixation

Singular	Plural	Gloss
mutarjim	mutarjim- <i>iin</i>	translator
Tabbaax	Tabbaax- <i>iin</i>	cook

Semantically coherent – limited to human-male

But not all human-male items fall into this class

##### 2) Feminine suffixation

Singular	Plural	Gloss
‘akl	‘akl- <i>aat</i>	food
sayyaara	sayyaar- <i>aat</i>	car

60% of feminine stems have final /a/ in native input (Kouloughli 1992)

75% do in L2 input (textbook counts; Thackston 1996; Brustad *et al.* 1995, 1997)

But final /a/ may also morphologically mark unit-hood, or nothing at all

20% of ‘broken’ stems also have it (constant across input types)

Feminine stems exhibit the most phonological variability, and are the largest single class if broken stems are decomposed (Boudelaa and Gaskell 2002)

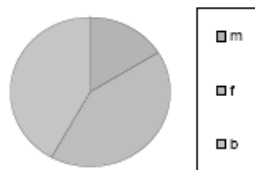
### 3) Broken internal stem modification

Singular	Plural	Gloss	Template
jariida	jaraa'id	newspaper	waraa'id
jism	'ajsaam	body	'awraad
qalb	quluub	heart	wuruud
rajul	rijaal	man	wiraad
kitaab	kutub	book	wurud

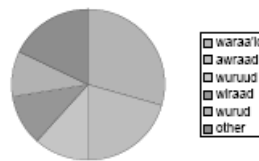
30-odd broken plural types, falling into 11 major classes  
The 5 most frequent, shown above, constitute ~80%

### 4) Plural distribution

3 major classes



Broken templates



## 1.2 Testing L2 gender assignment

### 1) Method

- Written questionnaire with singular stem (Arabic script) and English gloss given of 42 Arabic words, randomized per subject and all present in their textbook glossaries
- Items based on L1 experiment of Ravid and Farah (1999) with dialect substitutions
- Stimulus properties:

Masculine stems: 14 items

Feminine stems: 14 items, 9 with final /a/

Broken stems: 14 items, 3 with final /a/

- Informed consent obtained from 43 adult English-speaking subjects, none native speakers of Arabic or other Semitic language, grouped based on course membership:

Group 1 (n=7): completed one-month intensive Arabic course

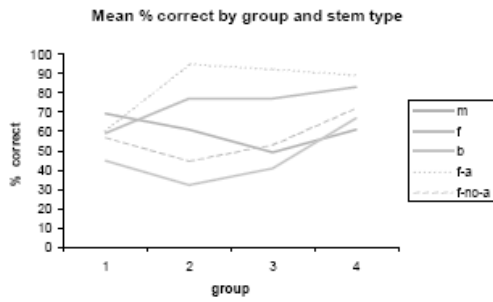
Group 2 (n=15): at end of 1-year university Arabic course

Group 3 (n=16): at end of 2<sup>nd</sup> year university Arabic course

Group 4 (n=5): at end of 4<sup>th</sup> year university Arabic course, also some immersion experience

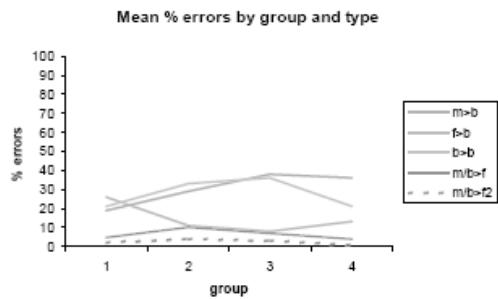
- Instructed to write the Arabic plural of each word, and to guess if they were not sure

## 2) Results – correct



No overall effect for group level; more proficient learners still at only ~70% accuracy  
 Better performance on feminine stems  
 But cf separation between feminine with and without final /a/  
 Better performance is driven by presence of the final /a/ cue for feminine

## 3) Results – errors



m>b: Tabbaax \*Tabbaa'ix (√ Tabbaax-iin)

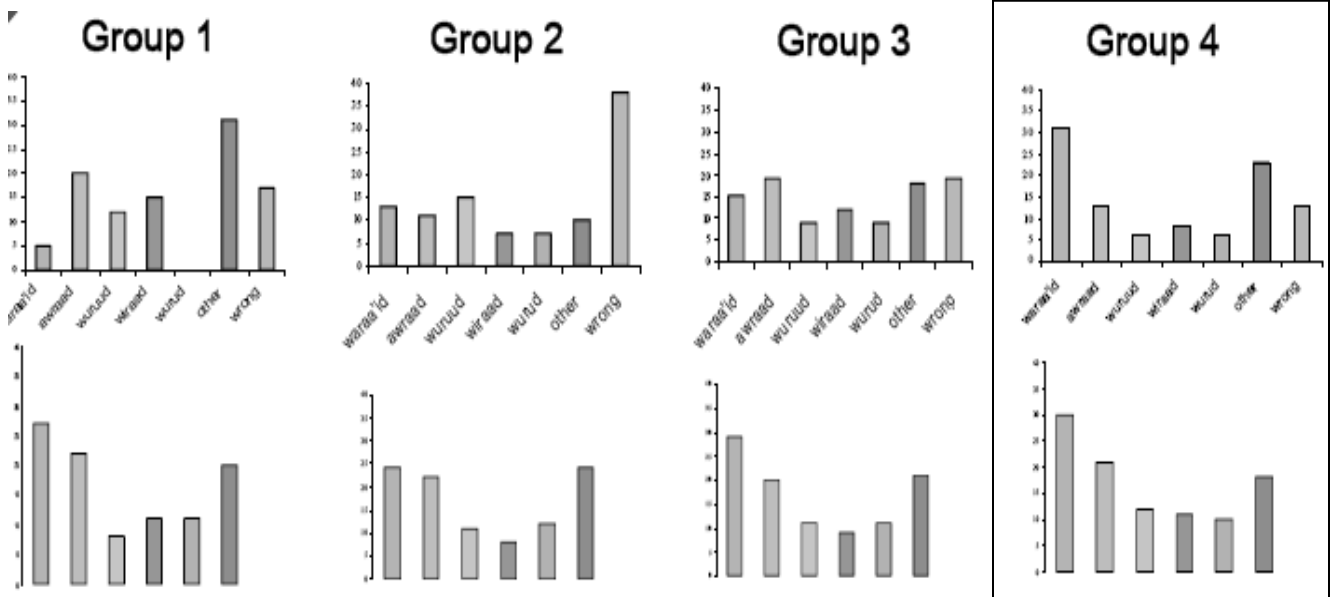
f>b: 'akl \*'ukuul (√ 'akl-aat)

b>b: jariida \*'ajraad (√ jaraa'id)

m/b>f: jism \*jism-aat (√ ajsaam)

Feminines still advantaged (due to final /a/)  
 Most errors involve an incorrect broken production from a masculine or different-broken-class stem  
 What are these erroneous broken outputs like, and how are they distributed?

### OUTPUT DISTRIBUTION



### INPUT DISTRIBUTION

Most errors do fall into one of the pre-existing broken plural classes, rather than being entirely erroneous. By the proficiency level obtained by Group 4 members, the output distribution closely mirrors the input. Speakers do not produce correct plurals at the individual item level, but

- they know the relative frequency of occurrence of each class,
- and structure their guesses accordingly

### 1.3 Discussion

- Speakers make the morphological generalization that a final /a/ implies feminine gender, and that lack of final /a/ means not feminine gender
- The relative robustness of this generalization leads to an advantage for feminine stems
- In the absence of this cue, speakers model their output on the statistical distribution of gender in the lexicon
- This behavior contrasts strongly with that of L1 Arabic learners –
  - Their performance on feminine stems rapidly reaches ceiling (~3 years)
  - Errors overwhelmingly involve overregularization to feminine

## 2. Arabic loanwords

### 2.1 Spanish gender

- Spanish speakers natively have masculine and feminine grammatical gender categories
- Feminine often associated with final /a/, but may lack it, and some masculine forms have it
- Articles and adjectives agree:

- 1) a. la profesor-a guap-a  
the-F professor-F goodlooking-F
- b. el profesor guap-o  
the-M professor goodlooking-M  
'the goodlooking professor'

- The singular definite article is subject to a productive hiatus-resolving alternation (Harris 1987):

**Singular definite article allomorphy:** la (F) el (M) / \_\_ [N á...]

- 2) a. el águ-a \*la águ-a  
the-M water  
'the water'
- b. el águ-a suci-a \*el águ-a sucio  
the-M water dirty-F  
'the dirty water'
- c. la mism-a águ-a \*el misma águ-a, \*el mismo águ-a  
the-F same-F water  
'the same water'

- This alternation formerly applied to all initial /a/ words, not just stressed (Penny 2000)
- Many loanwords entered from Arabic during this time (pre-1500)

- Due to imperfect knowledge of the source language, many were borrowed with the Arabic definite prefix /a(l)- still attached, and are therefore /a/-initial in Spanish
- This set of loanwords is the object of investigation here

## 2.2 Spanish loanwords from Arabic

- 453 such /a/-initial loanwords into Spanish from Arabic were identified in an etymological dictionary (Corominas and Pascual 1997)
- The grammatical gender for 438 of them was established using the Real Academia Espanola (RAE) dictionary and examining corpus examples (Davies)

### Masculine

- Masculine forms vary in their phonological form at word end
- Many end in phonotactically licit phonological consonants from the Arabic forms
- Other consonants are followed by epenthetic /e/, the normal epenthetic vowel in Spanish

### Feminine

- Feminine constitutes 41% of the set with corpus attestations (n=245), and 40% of the RAE set (n=438)
- These percentages accord with those of /a/-initial forms in the Spanish lexicon as a whole (Davies; Sebastián *et al.* 2000 for synchronic lexicon):

Gender of /a/-initial Spanish words				
	1200s	1300s	1400s	Now
N	154	119	198	--
%F	40	49	46	41

- Feminine forms overwhelmingly end in /a/ in Spanish (96 out of 100 from Davies)
- Of those, most of the Arabic source words also end in /a/
- 10 exceptions out of 100 Davies feminine forms, 16 from Davies+RAE's 175, listed below:

Exceptions to Arabic final /a/		
<i>Spanish</i>	<i>Arabic</i>	<i>Gloss</i>
azotea	suTayH	building covering; human head
atalaya	Talaayic	tower or elevated place
arracada	qarraT	earring
almarada	mixraz	iron weapon point or needle
alhóndiga	funduq	market, warehouse
alforja	xurj	type of carrying bag
alcarraza	karraaz	cooling container
alcaparra	kabar	type of plant
aduana	diwaan	customs
adárgama	darmak	type of flour
ajaquefa	shiqaaF	upper part of a building; iron decoration
albenda	band	patterned white hanging cloth
alharma	Harmal	type of plant
almijara	ma'jal	oil deposit
atafea	TafaH	indulgence to excess
almartaga	martak	?

- Most require an epenthetic final vowel for Spanish use, but not all
- One (*alharma*) gets final /a/ via deletion of the final consonant rather than epenthesis

- No obvious semantic femininity or other commonality unites this class of exceptions
- Arabic gender is also disparate – of the 10 Davies forms, one is unknown, one collective, three feminine, five broken.

### 2.3 Discussion

- Spanish speakers tend to class Arabic words as feminine with final /a/, masculine otherwise
- Like English learners of Arabic, they seem to hypothesize that final /a/ forms must be feminine, and non-final /a/ forms cannot be
- The main class of exceptions *introduces* a final /a/ for some feminine forms
- This is surprising, given the obligatory use of the masculine article with these /a/-initial nouns in some contexts
- It cannot be explained either through ‘normal’ phonology, or noun semantics

**Proposal:** Final /a/ is used in place of epenthetic /e/ so that the percentage feminine in the lexicon remains constant.

- As we have seen, the percentage of feminine forms among loanwords (40-41%) matches that of the lexicon now and at the time of borrowing (40-49%)
- Without these unexpectedly ‘feminized’ exceptions, that would not be the case
- Feminine percentages would drop to 38% (if they were not borrowed at all) or 36% (if borrowed as masculine)

OT Constraints:

NOCODA: Forms should not surface with consonant(s) in the syllable coda.

F=/a/: Feminine grammatical gender and final /a/ should be associated.

FAITH: Shorthand against epenthesis, vowel quality changes, etc

\*F: Forms should not surface with feminine grammatical gender.

\*M: Forms should not surface with masculine grammatical gender.

- \*F and \*M ranked stochastically in the sense of Boersma and Hayes (2001), in percentages that reflect the state of the lexicon

1) The normal masculine case

/atab/	NOCODA	F=/a/	FAITH	*F 59%	*M 41%
a. atab-F/M	*!				
b. atabe-F		*!	*	*	
c. atabe-M			*		*
d. ataba-F			*	*!	
e. ataba-M		*!	*		*

2) The normal feminine case

/ataba/	NOCODA	F=/a/	FAITH	*F 59%	*M 41%
a. atab-F/M	*!				
b. atabe-F		*!	*	*	
c. atabe-M			*!		*
d. ataba-F				*	
e. ataba-M		*!			*

### 3) The exceptional feminine case

/atab/	NOCODA	F=/a/	FAITH	*M	*F
a. atab-F/M	*!				
b. atabe-F		*!	*	*	
c. atabe-M			*	*!	
d. ataba-F			*		*
e. ataba-M		*!	*		*

- Computation for each item performed only once per speaker (possibly enforced by a form of Zuraw's (2000) USE-LISTED constraint)
- Result must then be propagated via its status as the first (possibly only) adaptation of the form, or via the social influence of that speaker.

### 3. Conclusions

- Based on these data sets, gender assignment by adults depends on both morphophonological properties of lexical items (final /a/) *and* on the proportional distribution of grammatical gender in the lexicon
- Spanish speakers mirror the lexical statistics of their native lexicon in adding borrowed items to it
- English speakers (who are attempting to acquire a second (L2) lexicon and lack an L1 distribution in any case) instead mirror the lexicon they aim to acquire
- Both patterns join a growing body of research suggesting an influential role for probability-matching in adult language, which differs from child acquisition patterns

### References

- Boersma, P. & B. Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.
- Boudelaa, S. & M.G. Gaskell. 2002. A re-examination of the default system for Arabic plurals. *Language and Cognitive Processes* 17, 321-343.
- Brustad, K., M. al-Batal, & A. al-Tonsi. 1995. *Al-Kitaab fii tacallum al-cArabiyya: A textbook for Arabic Part One*. Washington D.C.: Georgetown University Press.
- , 1997. *Al-Kitaab fii tacallum al-cArabiyya: A textbook for Arabic Part Two*. Washington D.C.: Georgetown University Press.
- Corominas, J. & J.A. Pascual. 1997. *Diccionario Crítico Etimológico Castellano e Hispánico*. Madrid: Gredos.
- Davies, M. <http://www.corpusdelespanol.org>.
- Harris, J. 1987. Disagreement Rules, Referral Rules, and the Spanish Feminine Article *el*. *Journal of Linguistics* 23:177-183.
- Hudson Kam, C.L., & E. L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1, 151-195.
- Kouloughli, D.-E. 1992. *Basic lexicon of Modern Standard Arabic*. L'Harmattan: Paris.
- Penny, R. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press.
- Ravid, D. & R. Farah. 1999. Learning about noun plurals in early Palestinian Arabic. *First Language* 19, 187-206.
- Real Academia Española. 2001. *Diccionario de la lengua española*. 22<sup>nd</sup> edition. Madrid. <http://www.rae.es>.
- Sebastián, N. & F. Cuetos, M. A. Martí, M. F. Carreiras. (2000). LEXESP: LéxicoInformatizado del Español. Edición en CD-ROM. Barcelona: Edicions de la Universitat de Barcelona (Col·leccions Vàries, 14).
- Thackston, W.M. 1996. *The vernacular Arabic of the Lebanon*. Cambridge, MA: Department of Near Eastern Languages and Civilizations, Harvard University.
- Wonnacott, E., & E.L. Newport. 2005. Novelty and regularization: The effect of novel instances on rule formation. In A. Brugos, M.R. Clark-Cotton, and S. Ha (eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.
- Zuraw, K. 2000. *Exceptions and Regularities in Phonology*. UCLA dissertation.