

## Gradient Gender Assignment in Spanish

*Mary Ann Walter*  
*walterma@mit.edu*  
 MIT

### 0. Introduction

- Underlying representations (URs) of linguistic forms may be underdetermined by evidence available to speakers, who then face a challenge in selecting a form.
- Spanish gender assignment to [a]-initial unfamiliar words presents such a challenge.
- I report the results of two studies on this case:
  - Gender assignment to loanwords (diachronic data)
  - Gender assignment to nonce words (synchronic data)
- The results exclude any role for Lexicon Optimization (LO).
- They require a more nuanced view of markedness/default gender.
- I propose an account in the spirit of Rice’s (2006) analysis of OT gender assignment.
- The gradience is captured by the Gradual Learning Algorithm (GLA; Boersma and Hayes 2001, Hayes and Londe 2006), via stochastic ranking of markedness constraints against each gender.

### 1. OT gender assignment

- Rice (2006) develops a theory of optimal gender assignment employing language-specific gender assignment constraints:
  - German gender assignment constraints:
  - \*-E M, N: A noun ending in schwa is assigned neither masculine nor neuter gender.
  - \*GE- M, F: A noun beginning in the morpheme Ge- is assigned neither masculine nor feminine gender.
  - \*SUPERORDINATE M, F: A noun denoting a superordinate is assigned neither masculine nor feminine gender.
- These are ranked as a bloc, the members of which are crucially non-ranked with respect to each other.
- This means that in cases of conflict, the gender violating the fewest of these wins.


Gemüse		GENDER FEATURES					
		*-E⇒M,N	*GE-⇒M,F	*SUP⇒M,F	*NEUT	*FEM	*MASC
a.	der Gemüse	*	*!	*			*
b.	die Gemüse		*	*!		*	
 c.	das Gemüse	*			*		

Figure 1: German gender assignment in cases of conflicting principles (from Rice 2006)

- The bloc dominates markedness constraints against each gender, also language-specifically ranked.
- When assignment constraints are balanced, the least marked gender wins.  
 No gender principles apply: default gender wins.

	Weg	GENDER FEATURES			*NEUT	*FEM	*MASC
		*-E⇒M,N	*GE⇒M,F	*SUP⇒M,F			
☞	a. der Weg					*	
	b. die Weg				*!		
	c. das Weg			*!			

Figure 2: German gender assignment with balanced principles, no violations (from Rice 2006)

Multiple but balanced gender principles apply: least marked gender not ruled out by a higher-ranked principle wins.

	Straße	GENDER FEATURES			*NEUT	*FEM	*MASC
		*-E⇒M,N	*GE⇒M,F	*SUP⇒M,F			
	a. der Straße	*!				*	
☞	b. die Straße				*		
	c. das Straße	*!		*			

Figure 3: German gender assignment with balanced principles, with violations (from Rice 2006).

- This account incorporates gradience to the extent that any OT analysis does – that is, the violable constraint rankings lead to grammatical tendencies rather than categorical absolutes.
- Further modifications are necessary to generate the full range of gradient data presented below.
- I do this by marrying the above account with the GLA.
- First, we will review the properties of grammatical gender in Spanish.

## 2. Gender in Spanish

- Spanish has masculine and feminine grammatical gender categories.
- Feminine often associated with final /a/, but may lack it, and some masculine forms have it.
- Additional principles associating feminine gender with final [d], [n] and [s] result in 95% correct gender assignment (Eddington 2004).
- Semantic principles have also been proposed, e.g. masculine for geographical and temporal divisions (overview in Bergen 1978).
- Masculine gender appears to be the least-marked/default category, despite parity in frequency (Prado 1982, Smith *et al.* 2003).

### Agreement morphology

- 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):

<b>Singular definite article allomorphy:</b> la (F)    el (M) / _ [N á...
---

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

- This alternation formerly applied to all initial /a/ words, not just stressed (pre-1500; 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.

### 3. Spanish loanwords

- 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

Figure 4

- 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	shiqaaif	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	?

Figure 5

- 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 inherent semantic femininity unites this class of exceptions.
- Nor does a semantic or other phonological Spanish gender principle apply.
- Arabic gender is also disparate – of the 10 Davies forms, one is unknown, one collective, three feminine, five broken.

Synchronic test

- As an additional test of whether some unidentified factor predisposes these exceptional items to feminine gender, the Arabic source forms were produced by an Arabic speaker and native speakers of Spanish asked to transcribe them with a definite article to indicate gender.
- N=5 (one other speaker failed to include the articles as instructed).
- One speaker classified 6 of the items as feminine.
- All but that subject did so for only two or three items.  
Two items (Spanish *alharma* and *atalaya*) are particularly likely to be assigned feminine, but not the group as a whole.

Portuguese data

- Corresponding Portuguese borrowings exist for 11 of the forms.
- Of these, 8 are feminine in Portuguese, 2 are masculine, and one is indeterminate but probably masculine.
- Portuguese has a smaller but still sizable corpus of Arabic (mostly [a]-initial) loanwords (n=297, based on da Cunha's (1982) etymological dictionary).

Gender of Arabic-to-Portuguese loanwords		
	#	%
Masculine	165	56
Feminine	118	40
Variable	13	4

Figure 6

- Once again, a subclass of feminines exists (n=17) that lacks final [a] in the source form.
- The overall percentage of feminine items declines from 40% (above) to 36% without them, 34% if they are classified instead as masculine.
- The Spanish pattern is replicated, albeit with not entirely independent data.

Conclusions

- Exceptional gender assignment occurs in a way that cannot be motivated by Lexicon Optimization or gender markedness/default, but does fall in with statistical patterns in the lexicon.
- We will now move to an experimental study that attempts to replicate this finding synchronically.

#### 4. Spanish nonce words

- Informed consent was obtained from Spanish speakers (North Mexican dialect; n=22; 12 female, 10 male; dominant or balanced bilinguals with English).
- Subjects were presented with written nonce word stimuli in a gender-neutral context and asked to complete the frame statement with a gender-inflected form.
- Stimuli (n=32) are bisyllabic or trisyllabic and always begin with a stressed [a] vowel..
- Five forms end in [a] and five in [o], as distractors quasi-deterministic with respect to gender.
- The remaining 22 forms all have final [e], which is masculine-biased (89%) but not deterministic.
- A post-test attempted to elicit the article allomorphy before stressed [a], and it was shown not to be in effect for these speakers.

##### Results

- Forms ending in [a] and [o] are essentially deterministic with respect to gender, as expected (95% feminine and masculine respectively).
- As expected, forms ending in [e] show more gradience and are classified as feminine 32% of the time.
- This level once again approximates lexical statistics, though not with the same precision as the loanword data.
- Beneath this approximation of the subject group as a whole, there is considerable variation in individual subject data.

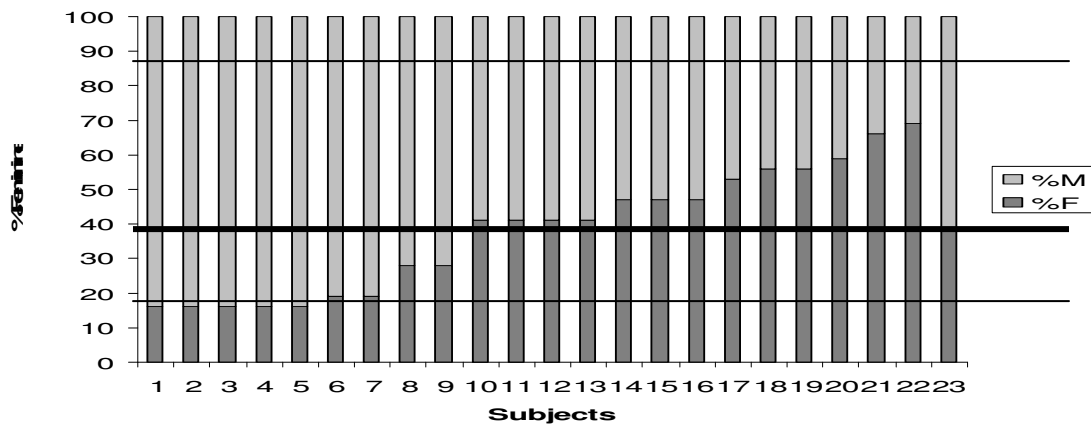


Figure 7: % Feminine gender by subject.

- A minority of subjects do seem to uniformly apply masculine gender, but most vary within the minimum and maximum bounds set by the fixed distractor percentages; the resulting group mean obscures this variation.
- Item analysis also shows variation, with the feminine classification rates of individual [e]-final stimulus tokens ranging from 16% to 59%. Most (14 of 22 items) fall in the 27% to 36% range.
- This suggests that once again, inherent properties of the nonce words (or loanwords) cannot be driving the exceptional feminine gender cases.

#### 5. The proposal

- Typical gender assignment of Spanish can be captured by a Rice-style analysis.

##### 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.

- Final [a] forms to which the F=/a/ gender principle applies are assigned feminine gender, and all others are masculine.

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

Figure X: Normal feminine gender assignment.

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

Figure X: Normal masculine gender assignment.

- If the markedness constraints \*F and \*M are stochastically ranked, however, some exceptional additions to the feminine class are predicted, just as we observe.
- Hayes and colleagues posit initial low rankings for language-specific constraints like the gender principle ones, which then rise over the course of the learning process.
- Their rankings eventually overtake those of the gender markedness constraints, but until then the relative ranking of \*F and \*M will be determined based on observation of their respective frequency in the lexicon (here, roughly 40/60%).
- A ranking like the one below, expected 40% of the time, generates the exceptions observed in the ratios observed:<sup>1</sup>

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

Figure X: Exceptional feminine gender assignment.

- I assume that winning outputs are consistently selected thereafter via something like the USELISTED constraint(s) proposed by Zuraw (2000), in order for it to keep prevailing for that individual and propagate through the speech community.
- Different lexical exposure prior to the final ranking predicts some variation between individuals' \*M/\*F ranking.
- This, in conjunction with performance factors, may account for the between-subject variation seen in the experiment in Section 4.

<sup>1</sup> This actually overgenerates exceptional forms. To the extent that competing variants exist in a population, among which the masculine form necessarily outnumbers the feminine one, this undershoot is expected.

## 6. Conclusions

- Historical and experimental data converge to show that Spanish speakers do not assign grammatical gender to unfamiliar forms by relying on previously-proposed grammatical principles such as LO or categorical constraint ranking.
- Instead they do it gradiently, in accordance with probabilistic ratios already present in their lexicons.
- I do not conclude from this that it must therefore occur outside the grammar (as Vaux and Nevins (2005) do for similar phenomena).
- Rather, I present a grammatical account couched within a model of OT that incorporates gradience.

## 7. References

- Bergen, J. (1978). A Simplified Approach for Teaching the Gender of Spanish Nouns. *Hispania* 61:865-876.
- Boersma, P. & B. Hayes. (2001). Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.
- Corominas, J. & J.A. Pascual. (1997). *Diccionario Crítico Etimológico Castellano e Hispánico*. Madrid: Gredos.
- da Cunha, A. G. (1982). *Dicionário Etimológico Nova Fronteira da Língua Portuguesa*. Rio de Janeiro: Editora Nova Fronteira.
- Davies, M. <http://www.corpusdelespanol.org>.
- Eddington, D. (2004). *Spanish Phonology and Morphology: Experimental and Quantitative Perspectives*. Amsterdam: John Benjamins.
- Harris, J. (1987). Disagreement Rules, Referral Rules, and the Spanish Feminine Article *el*. *Journal of Linguistics* 23:177-183.
- Hayes, B. & Z. Londe. Stochastic Phonological Knowledge: The Case of Hungarian Vowel Harmony. *Phonology* 23:59-104.
- Penny, R. (2000). *Variation and Change in Spanish*. Cambridge: Cambridge University Press.
- Prado, M. (1982). El Género en Español y la Teoría de la Marcadez. *Hispania* 65:258-266.
- Real Academia Española. (2001). *Diccionario de la lengua española*. 22<sup>nd</sup> edition. Madrid. <http://www.rae.es>.
- Rice, C. (2006). Optimizing Gender. *Lingua* 116:1394-1417.
- 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).
- Smith, P. & A. Nix, N. Davey, S. Lopez Ornat, D. Messer. (2003). A Connectionist Account of Spanish Determiner Production. *Journal of Child Language* 30:305-331.
- Vaux, B. & A. Nevins. (2005). Formal and Empirical Arguments for Morpheme Structure Constraints. Talk given at the annual meeting of the Linguistic Society of America, Oakland, CA.
- Zuraw, K. (2000). *Patterned Exceptions in Phonology*. PhD dissertation, UCLA.

## Appendix: Spanish experimental materials.

### Frame sentence:

Hay más alfes allá. Por favor dame \_\_\_\_\_ (uno/una).

‘There are more alfes over there. Please give me one.’

### Stimuli:

Bisyllabic [e]-final		Trisyllabic [e]-final		[a]-finals	[o]-finals
alfe	arpe	áculé	ánabe	albra	acho
alme	arre	álcaché	ántere	ampa	alco
ambe	arze	álfique	árcite	anfa	anso
anche	aspe	ámbroje	árupe	arga	apro
anger	azque	ámolve	ázquete	asca	arfo
anve		ámpone			

### Post-test:

- 1) \_\_\_ águila aquí es débil. \_\_\_ grandes águilas allí son fuertes.
- 2) \_\_\_ agua aquí es dulce. \_\_\_ agua allí no es \_\_\_\_\_ para beber.

- 1) (The) eagle here is weak. (The) big eagles over there are strong.
- 2) (The) water here is sweet. (The) water over there isn't (good) for drinking.