Entropy-Conditioned Syncope of Vowels in Latin and Ibero-Romance

Robert Elwell
Department of Linguistics
University of Texas at Austin
1 University Station B5100
Austin, TX 78712-0198 USA
relwell@mail.utexas.edu

December 9, 2006

Abstract

In this paper, I contrast Old Spanish and Latin, showing that post-tonic syllables are subject to vowel syncope because they in fact carry less overt information via actual utterance. This is measured in terms of entropy, using maximum likelihood estimates from a trigram model of Latin segments. For the hypothesis to be correct, the entropy of the probability distribution from the summation of all maximum likelihood estimates of vowels after a stressed syllable should be lower than that of vowels after an unstressed syllable. I discuss the methodology of my experiment as well as frame my findings in optimality-theoretic terms.

1 Introduction

Post-tonic syllabic peak deletion in Latin is a well-documented phenomenon that attributed to the historical direction which many Romance languages took–Spanish being one of these. According to an analysis by Lleó (2003), this results from a series of optimality-theoretic constraint interactions, mainly focused around the prosodic well-formedness of footed or unfooted syllables in Latin. An example of this is the change from Latin MANICA, 'sleeve', to Spanish *manga*, where the first syllable is stressed and the following syllable loses its peak, thus causing its onset to become a coda on the preceding syllable. Lleó frames this as the domination of the workhorse constraints PARSE-SYL and FOOTTROCHEE (among others) over the constraint Max-Segment:[1]

/manica/	FootTrochee	Parse-Syl	Max-Segment
a. (má.ni.ca)	*!		
b. (má.ni)ca		*!	
c. ☞ (mán.ga)			*

Table 1: Prosody-based analysis of vowel syncope. (Lleó, 2003)

While a syllable-structure-based analysis is of some theoretical interest, there are a few shortcomings in taking this approach. First, assuming a change in a constraint ranking that originally allowed post-tonic syllables over a syllabic well-formedness rule assumes a great degree of change deriving from a single reranking. However, a set of rerankings carries its own set of questions, such as whether there is any data to describe the sequence of constraint rerankings, or why so many constraints would rerank. If the change stemmed from contact, it would be difficult to believe that stress would be the first to change, as it is generally the case that single lexical borrowings will be forced to agree with the phonotactics of the borrowing language and not that of its native origins. Therefore, the clearest explanation would be that this is a case of internal phonological change.

Internal change can be best understood as the interaction of relexification with the conflict that exists between different functional principles that carry their own fixed or partly fixed constraint ranking hierarchies. These functional principles are eternally at odds at each other, and their constant reranking with respect to one another yields a cycle of "eternal optimization". (Boersma, 2003) This concept will be considered as the underlying postulation regarding what will be argued here. Functional principles that lead to fixed rankings for prosodic words will be considered as follows:[2]

(1)
1. Minimization of articulatory effort.
2. Maximization of perceptual contrast.
3. Maximization of utilization of prosodic proclivities.

The constraint rankings which each of these conflicting principals yield would interact in the manner which ultimately resulted in the change from Latin to Old Spanish. (1), in the context of the prosodic word, can clearly be understood as a case of where a greater number of syllables would increase articulatory effort. (2), for the case of the prosodic word would, be such that every segment contributes to

---

[1] For a more thorough description of the analysis, please see Lleó (2003)

[2] Adapted from the functional principles for obstruents in Boersma (2003).

discerning words. For instance, a language consisting of the two sequences ‘xyz’ and ‘wyz’ would not be a proper maximization of perceptual contrast, while one consisting of ‘xyz’ and ‘abc’ would. (3) simply refers to the importance that every utterance be feasible within the precepts of the established stress pattern.

Under this paradigm, Latin originally existed with the set of constraints that reflected (3) probably equally ranked with those reflecting (1). (2) would have been preventing deletion of vowels to maintain as much contrast as possible. The change stemmed from the continual striving for optimality, because over time, as the prosodic template became concretized, the information carried in words by post-tonic vowels was weakened, as they had the least presence in the word so as to contrast with the stressed syllable preceding it. This phonetic characteristic would lead to a minimized perceptual contrast in this vowel. Because of the emphasis on the preceding vowel, and such concepts as the obligatory contour principal mitigating the appearance of certain vowels (always stressed vowels, often long vowels), the number of different phonemic vowels was also reduced. A post-tonic unstressed vowel, therefore, is extremely more predictable than a vowel following an unstressed segment. This predictability directly correlates with the deletion of these vowels.

My hypothesis is that post-tonic vowels in Latin were lost because they carried little information and were by and large more predictable than other vowels. This predictability was inversely related to the carrying of new information or perceptual contrast, rendering these vowels a weak link in the prosodic word. As the stress scheme of Latin changed to that of Old Spanish, losing nonfinality, in order to maintain all syllables in the prosodic word parsed in a trochaic, binary foot, these vowels were deleted for a minimal loss in actual perceptual contrast.

2 Experiment

To make my case, I will frame my argument within the concept of entropy. Entropy is the measure of uncertainty within a probability distribution. The higher the entropy, the greater the amount of encoding necessary to properly describe each outcome within the distribution. Entropy is calculated as follows:

$$H(X) = -\sum_{x \in X} p(x) log_2 p(x)$$

Using texts from the Latin Library[3] annotated for syllable boundaries and vowel length, I will use a trigram model to calculate the maximum likelihood

---

[3] I would like to take time to extend my gratitude to Lev Blumenfeld, whose work on Latin corpora and helpfulness made my research possible.

estimate (MLE) of all possible segments in the third position of a trigram. This probability distribution is what I will calculate the entropy for. Therefore, my hypothesis is that the entropy of the probability distribution of the MLE of a vowel position following a stressed vowel will be significantly lower than the entropy of probability distribution of the MLE of a vowel position following an unstressed vowel. Maximum likelihood estimate for trigrams are calculated as follows:

$$P_{MLE}(w_3|w_1, w_2) = \frac{C(w_1, w_2, w_3)}{C(w_1, w_2)}$$

Finding the entropy of the distribution of all outcomes for the third position of a trigram, therefore, is calculated as follows:

$$H(X_{MLE}) = -\sum_{x \in X} \frac{C(w_{x-2}, w_{x-1}, w_x)}{C(w_{x-2}, w_{x-1})} log_2 \frac{C(w_{x-2}, w_{x-1}, w_x)}{C(w_{x-2}, w_{x-1})}$$

The corpus in use is 113,478 words long. For each word, I annotated the primary stress and word boundaries. The original work was annotated for length and syllable boundaries. My stressed version is only annotated for stress and length. The appearance of the Latin text is as follows:

```
#consvEtÚdinis@
#mIrÉtur@
#atrócitAs@
```

Using this intermediate corpus, I then created a set of trigrams using single segments. This yielded 860,695 trigrams at my disposal. The beginning and end characters were mainly used to separate different words to maintain correct trigram numbers; trigrams with either character in the middle position were discarded. These trigrams appear as follows:

```
# c o
c o n
o n s
```

I then performed the above calculation for four different sets of trigram sequences:

1. Post-tonic segments (V́ C _)
2. Post-tonic vowels (V́ C V)
3. Post-atonic segments (V C _)
4. Post-atonic vowels (V C V)

3 Results

The calculations yielded the following results:

	Post-tonic		Post-atonic	
Trigram	(V́ C _)	(V́ C V)	(V C _)	(V C V)
Entropy	4.316	2.728	3.441	3.824

Table 2: Entropy measurements given MLE probability distributions for trigrams.

These results confirm my hypothesis with a slight hiccough: the entropy for all segments in which the first trigram position is a stressed vowel. This is due to the fact that not only vowels can appear two segments from a stressed vowel, but onsets after a coda as well. This is also compounded by the fact that stress is quantity-sensitive. However, with a more focused scope or simply looking at post-tonic vowels (and therefore ignoring cases of quantity sensitivity at the moment), the entropy is significantly lower in comparison. In information-theoretic terms, post-tonic vowels necessitate an entire bit less to encode than post-atonic vowels.

4 Theoretical Ramifications

The results show that there is more involved in the syncope of post-tonic vowels in Latin than simply the interaction of stress constraints. Interestingly enough, the Lleó analysis actually is just shy of presenting this concept. Consider the following constraint it presented:

(2) "MAX-MORPHEME: An input morpheme has a correspondent in the output."

This prevents the deletion of a post-tonic vowel if it is itself a morpheme, such as in the case of nominal gender endings (cf. MALU > *malo* vs. MALE > *mal*). While it would be difficult to actually quantify the amount of entropy involved in these situations without a large corpus of interlinear glosses and a great deal of effort, it can be assumed that morphological data contributes an amount of perplexity to the vowel that would protect it from being deleted due to low entropy. This assumption would actually render the constraint MAXMORPH unnecessary.

In order to properly frame the findings here, it is important first to describe the state of Latin before the changes to Old Spanish. These will be framed using the principals described in 1. The constraints to consider in this case are as follows:[4]

---

[4] Constraints from Kager (1999), based on Mester (1994).

(3) NONFINALITY: Do not parse the word-final syllable. Parsed word-final syllables incur a penalty.

(4) FTBIN: Feet are binary. Assign a penalty to any non-binary feet.

(5) TROCHEE: Feet are trochaic. Assign a penalty to non-trochaic feet.

(6) PARSE-SYL: All syllables should be parsed into feet. Assign a penalty to unparsed syllables.

(7) WSP: Heavy syllables are stressed.

(8) RIGHTMOST: The head foot is the rightmost in the prosodic word. Assign a penalty if the rightmost syllable is not the source of main stress.

(9) MAXSEG: A segment in the input should be reflected in the output. Assign a penalty to any segment in the input not reflected in the output.

The definition of (9) will serve its purpose here, but will be revised at a later point to better reflect its relation with high-entropy positions. Here, we will assume the following ranking:

(10) NONFINALITY, MAX-SEG $\gg$ WSP, FTBIN, TROCHEE $\gg$ PARSE-SYL

Focusing on the more pertinent constraints within this ranking, MANICA is evaluated as follows:

		manica	NONFINALITY	MAX-SEG	FTBIN	PARSE-SYL
a.		(má.ni.ca)	*!		*	
b.		(má.ni)(ca)	*!		*	
c.		(mán.ga)	*!	*		
d.		(mán)<ga>		*!		*
e.		(mán)(ga)	*!	*	*	
f.	☞	(má.ni)<ca>				*

Table 3: An evaluation of MANICA in Latin using updated constraints

The demotion of NONFINALITY below PARSE-SYL caused a problem that needed to be resolved by reducing the number of syllables to return to prosodic normalcy within the established stress hierarchy. However, given the current definition of MAX-SEG, this is impossible, and ultimately will lead to the same optimal candidate as before (see Table 4).

Because of this issue, and considering the results of the experiment described above, MAX-SEG should be redefined as follows:

/manica/		MAX-SEG	FTBIN	PARSE-SYL	NONFINALITY
a.	(má.ni.ca)		*!		*
b.	(má.ni)(ca)		*!		*
c.	(mán.ga)	*!	*		*
d.	(mán)<ga>	*!		*	
e.	(mán)(ga)	*!	*		*
f.	(má.nic)	*!	*	*	
g. ☹	(má.ni)<ca>			*	

Table 4: Nonfinality demotion given the current definition of MAX-SEG

(11) MAX-INFO: A segment in the input, given a relatively pertinent degree of phonological information, should be reflected in the output. Assign a penalty if an information-carrying segment is not reflected in the output.

With this new constraint in lieu of MAX-SEG, we will arrive at the following evaluation:

/manica/		MAX-INFO	FTBIN	PARSE-SYL	NONFINALITY
a.	(má.ni.ca)		*!		*
b.	(má.ni)(ca)		*!		*
c. ☞	(mán.ga)				*
d.	(mán)<ga>			*!	
e.	(mán)(ga)		*!		*
f.	(má.ni)<ca>			*!	
g.	(má.nic)	*!			

Table 5: Evaluation of MANICA given updated stress and proper MAX constraint.

Note here that (má.nic) was not a proper candidate because post-atonic vowels in fact carry a greater degree of information contributing to the formation of the word perceptually.

As previously stated, this new constraint and its position in the hierarchy can also adequately account for the MALU > *malo* vs. MALE > *mal* issue. Before this can be accounted for, another constraint must be posited:

(12) STW: A stressed syllable is heavy. Assign a penalty to any stressed syllable that is not heavy.

Given this ranked above NONFINALITY, the following tableaux can be seen in Tables 6 and 7. Considering the fact that MAX-INFO accounts for the deletion (or lack thereof) of certain vowels, and given these cases, the actual ranking of Max-Seg can be seen in (13).

/malu/	Max-Info	FtBin	STW	Nonfinality
a. ☞ (má.lo)			*	*
b. (mál)	*!			*

Table 6: Preservation of masculine *-o* in Old Spanish due to Max-Info

/male/	Max-Info	FtBin	STW	Nonfinality
a. (má.le)			*!	*
b. ☞ (mál)				*

Table 7: Deletion of final *-e* given STW and low entropy of post-tonic vowels

(13) Max-Info $\gg$ STW $\gg$ Max-Seg

5 Conclusion

This has been an occasion where raw numerical calculation has managed to bring a degree of insight into several different areas of study and arrive at new conclusions for them all. First, and most important, there is a clear relationship between what can and cannot be deleted from a word, and this can be quantified in terms of entropy. Furthermore, segments themselves have a non-semantic amount of meaning for encoding semantic forms, and given the well-known noisy channel model (Shannon, 1948), certain segments can be dropped at a minimal loss for carriage of this information. The proof here is not only in the calculations within this article, but the fact that the change happened at all, and that a speaker could drop these vowels and a hearer would hear this form and still understand it as the same semantic denotation.

Another item to note from this article is that computational methods are useful in the study of historical linguistics, and may yield possible explanations where there were previously none. Given a corpus large enough, therefore, much information can be gleaned regarding the change of language over time, and what may have spurred those changes. Surely, cases of syncope in other languages will reflect the findings here, and historical linguists have the technology and often the corpora to confirm this.

Furthermore, historically speaking, we can see that historical change can be viewed in the lens of the eternal strive for optimality within phonological encoding of semantic information. Learning a grammar, therefore, means innately drawing conclusions about the probability of prediction of segments. As speaker groups learn how information is encoded through speech, they also learn ways in which they can say the same things with greater ease. This is also the view of phonological change via relexification in many ways, and thus bolsters many such arguments greatly.

Finally, at least one constraint, if not many, many more, exist in the optimality-theoretic universal grammar that hinge on how individuals can or cannot understand words in the "noisy channel" that is the air itself that lies between the speaker and hearer. Possibly, then, well known cases of opacity can be framed in this context. In general, while many such constraints exist in OT, they are often disparate, generally lacking strong, established theoretical grounding, and unquantifiable. Here, it is not any of those cases.

The analysis given here leaves very much open to future investigation. Note the wording necessitated for the constraint described in 11–"a relatively pertinent degree" of information. This is to capture the categoricalness of constraint violation, with the understanding that the same constraints must be violated equally cross-linguistically, as they must exist universally. Because this is quantifiable, however, this may open up future discussion as to whether thresholds against violations are language-specific. Perhaps there is a quantifiable amount of violation per language that is allowable until a penalty is assigned. This could, in fact, save many of the insights of the theory that have been threatened by such criticisms as seen in McMahon (2000), and other arguments that language-specific constraints must exist.

References

P. Boersma. The odds of eternal optimization in Optimality Theory. In *Optimality Theory and language change*, pages 31–65. Kluwer Publishers, 2003.

R. Kager. *Optimality Theory*. Cambridge University Press, 1999.

C. Lleó. Some interactions between word, foot, and syllable structure in the history of Spanish. In D. Eric Holt, editor, *Optimality Theory and Language Change*, pages 249–283. Kluwer Academic Publishers, 2003.

April McMahon. *Change, Chance, and Optimality*. Oxford University Press, 2000.

R.A. Mester. The quantitative trochee in latin. *Natural Language and Linguistic Theory*, 12:1–61, 1994.

C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–432, 623–656, 1948.