Automated Deduction in Historical Phonology

陳朝陽
Chen Zhaoyang
zc23@illinois.edu

November 30, 2022

TL; DR

I wrote a small piece of program (~ 1,000 loc) that implements a couple dozen major phonological changes from Latin to Modern Spanish. The program is able to derive a good amount of Modern Spanish words from their Latin etymon.
This presentation is about how it is done.

Reviving a field that has been silent for a while

Early attempts of works alike goes back the late '60s[1]. Previous works on automated forward reconstruction has been done on Romance languages a couple times e.g. for Old French[2], and Spanish[3].
As of now, this field has been silent since the late '90s.

---

[1] Patrick Sims-Williams. "Mechanising Historical Phonology". In: *Transactions of the Philological Society* 116.3 (2018), pp. 555–573; Raoul N. Smith. "Automatic Simulation of Historical Change". In: *COLING 1969* (1969). URL: https://aclanthology.org/C69-0901.pdf.

[2] Sarah K. Burton-Hunter. "Romance Etymology: A Computerized Model". In: *Computers and the Humanities* 10.4 (1976), pp. 217–220. URL: https://www.jstor.org/stable/30199805.

[3] Charles L. Eastlack. "A Program to Simulate Systematic Sound Change in Ibero-Romance". In: *Computers and the Humanities* 11.2 (1977), pp. 81–88. URL: https://www.jstor.org/stable/30199864.

Why do it again?

An essentially identical project was done from the 1980s to early 2000s by S. Lee Hartman and is still accessible online as of now[4].
So why do it again?
The answer is that, sadly, this is field largely inactive as of now; many works were done in a time when publishing programs were not feasible, thus many works were simply lost.
Luckily, computing has made enough progress so that this work could be done rather quickly in a high-level programming language.

[4] Steven Lee Hartman. *Phono (Version 4.0): Software for Modeling Regular Historical Sound Change*. 2003. URL: https://langnhist.weebly.com/files/theme/ver40.pdf (visited on 10/18/2022).

The Standard ML Programming Language

This small system is written entirely in Standard ML.
SML is a relatively small and concise programming language that has been primarily used in programming language implementation and automated theorem proving.
Its *Heimatland* was University of Edinburgh, frist implemented in the LCF (logic of computable functions) theorem prover[5] by Robin Milner and his collegues, whose predecessor is the Stanford LCF system, written in LISP, also by R. Milner.

[5] David MacQueen, Robert Harper, and John Reppy. "The History of Standard ML". In: *Proc. of the ACM on Programming Languages* 4 (2020). HOPL. DOI: https://doi.org/10.1145/3386336.

Implementing a Phonology

This project has a simple two-layer structure: the first layer that defines ways of constructing the *statics* of a phonology – namely the segmental inventory, syllable structure, and the phonological word of the Spanish language and her predecessors – and the second layer that defines ways of constructing the *dynamics* of a phonology – namely sound changes and how to compose them into chain shifts.

The Statics

Representing Features, Segments, Syllables, and the Phonological Word

We define all of the *static* element of this system through recursive types in ML.
Recursive types are basically a combination of **and** & **or**'s.

The Dynamics

Rewriting of Syllabic Constituents, Syllables, and Phonological Words

ML provides extensive supports for composing functions.

```
(* rewrites that changes
a constituent of the syllable *)

datatype rewrite = Onsetism of (onset -> onset)
                 | Nucleusism of (nucleus -> nucleus)
                 | Codism of (coda -> coda)

(* syllabism that rewrites a syllable to another *)

datatype context = NoContext
                 | Predicate of (syllable -> bool)
                 | WordInit
                 | WordFinal

datatype syllabism
= Syllabism of (syllable -> syllable) * context
```

Pearls of Sound Changes from Latin to Romance

In the remainder of this presentation, I am going to demonstrate 6 sets of sound changes to demonstrate how this system works. The majority of the words in examples comes from *Romance Languages: A Historical Introduction*[6], *From Latin to Spanish*[7], and *A History of the Spanish Language*[8].

[6] Ti Alkire and Carol Rosen. *Romance Languages: A Historical Introduction*. Cambridge University Press, 2010.

[7] Paul M. Lloyd. *From Latin to Spanish*. American Philosophical Society, 1987.

[8] Ralph Penny. *A History of the Spanish Language*. Cambridge University Press, 2002.

Latin → Proto-Romance: Romance Vowel Shifts

Arguably the most fundamental change from Late Latin to Proto-Romance is the transformation of its vowel system.
The transformation has the following components:

1. Loss of Vowel Quantity
2. The Great Vowel Merger
3. Merger in Atonic Vowels

Another important sound change, the loss of hiatus, unfortunately we are not going to cover in this presentation.

Loss of Vowel Quantity

	Front		Cent.		Back	
High	ĭ	ī			ŭ	ū
Mid	ĕ	ē			ŏ	ō
Low			ă	ā		

	Front	Central	Back
High	i		u
High-Mid	ɪ		ʊ
Mid	e		o
Low-Mid	ɛ		ɔ
Low		a	

LATINA	Español
VĪTA	vida
VICĪNA	vecina
FARĪNA	harina
LŪNA	luna
DŪRA	dura
MŪRU	muro
HŌRA	hora
CŌRTE	corte
DĒBET	debe
TĒRNU	terno

The Great Merger

	Front	Central	Back
High	i		u
High-Mid	ɪ		ʊ
Mid	e		o
Low-Mid	ɛ		ɔ
Low		a	

	Front	Central	Back
High	i		u
Mid	e		o
Low-Mid	ɛ		ɔ
Low		a	

LATINA	PrRom	Español
GULA	[ʊ]	gola
CURRIT	[ʊ]	corre
MŬSCA	[ʊ]	mosca
BIBIT	[ɪ]	bebe
LITTERA	[ɪ]	letra
VĬCE	[ɪ]	vez

Atonic Merger

ɛ → e

ɔ → o

We can see that the Latin five vowel system is basically restored after these changes (after stressed [ɛ] and [ɔ] diphthongizes in Castilian, it is indeed fully restored.)

LAT.	Es.
HĪBERNU	iviern**o**
CIRCĀRE	c**e**rcar
VĒNĀTU	vena**d**o

Romance Vowel Shifts: Rule Ordering

Loss of Quantity < Great Merger < Atonic Merger

Latin → Proto-Romance: Fundamental Consonantal Shifts

$m \to \emptyset$

$j \sim ʝ \to dʒ \sim ɟ$

$w \to β$

$t^j \to ts$

$k^j \to tʃ$

$d^j, g^j \to dʒ \sim ɟ$

Proto-Romance → Western Romance: Intervocalic Lenition I

LAT	Es.
CABALLU	caballo [β]
DEBERE	deber [β]
HABERE	haber [β]
CRUDU	crudo [ð]
PEDE	pie [∅]
AUGUSTU	agosto [ɣ]
LIGARE	ligar [ɣ]
PAGANU	pagano [ɣ]

LAT.	Es.
SAPORE	sabor [β]
CAPUT	cabo [β]
COPERTU	cubierto [β]
VITA	vida [ð]
FATA	hada [ð]
CATENA	cadena [ð]
AMICA	amiga [ɣ]
SECURU	seguro [ɣ]
FOCU	fuego [ɣ]

LAT.	Es.
SERPENTE	serpiente
ALPES	aples
RUMPERE	rumper
ORTICA	ortiga
MENTA	menta
ARCU	arco
FALCONE	halcón

Proto-Romance → Western Romance: Degemination

LAT.	Es.
OSSU	hueso
SUMMA	suma
APPELLAT	apela
LITTERA	letra
SICCU	seco

Westerm Romance → Old Spanish: Debuccalization of [ɸ]

LAT.	Es.
FILU	hilo
FERIRE	herir
FERRO	hierro
FALCONE	halcón
FOCU	fuego
FORA	fuera
FONTE	fuente
FRONTE	frente
FLORE	flor
FLACCU	flaco

Old Spanish → Modern Spanish: The Spanish Sibilant Rearrangement

The sibilants in Alfonsino Spanish:

	dental affricate	alveolo-apical	palatal
voiceless	ç [t͡s] (→ [s̪])	-ss- [s̪]	x [ʃ]
voiced	z [d͡z] (→ [z̪])	-s- [z̪]	j, ge, gi [ʒ] (← [d͡ʒ])

Bibliography

- Smith, Raoul N. "Automatic Simulation of Historical Change". In: *COLING 1969* (1969). URL: https://aclanthology.org/C69-0901.pdf.
- Burton-Hunter, Sarah K. "Romance Etymology: A Computerized Model". In: *Computers and the Humanities* 10.4 (1976), pp. 217–220. URL: https://www.jstor.org/stable/30199805.
- Eastlack, Charles L. "A Program to Simulate Systematic Sound Change in Ibero-Romance". In: *Computers and the Humanities* 11.2 (1977), pp. 81–88. URL: https://www.jstor.org/stable/30199864.
- Lloyd, Paul M. *From Latin to Spanish*. American Philosophical Society, 1987.
- Penny, Ralph. *A History of the Spanish Language*. Cambridge University Press, 2002.
- Hartman, Steven Lee. *Phono (Version 4.0): Software for Modeling Regular Historical Sound Change*. 2003. URL: https://langnhist.weebly.com/files/theme/ver40.pdf (visited