

Tracking grammaticalization across Romance: Evidence from the subjunctive

Recent work on linguistic change has attempted to characterize language families in terms of the positioning of their daughters along the cline of grammaticalization. In Romance, there is widespread agreement that French is the most innovative vis-à-vis its Latin source (e.g. Carlier et al. 2012; Harris 1984; Posner 1996), though the relative positioning of its sisters is less clear. We contribute to this debate with usage data on a key grammatical diagnostic: the choice between indicative and subjunctive in embedded complement clauses. Most scholars view mood selection as semantically motivated (with real, asserted, etc. predicates expressed in indicative, while irrealis complements select subjunctive). However, typological studies adopting a diachronic perspective suggest that subjunctives grammaticalize into concomitants of subordination (Bybee, Perkins & Pagliuca 1994). Such grammaticalization would entail lower rates of subjunctive morphology and “vacuous” variability (i.e. alternation that is *not* semantically motivated). In this paper, we address these alternative perspectives, focusing on two major parameters of grammaticalization: semantic bleaching (desemanticization) and “obligatorification” (Lehmann 1995).

The data on which our cross-linguistic comparison is based come from four corpora of French (FR; Poplack 1989), Italian (IT; Cresti & Moneglia 2005), Spanish (SP; Martín Butragueño & Lastra 2011; 2012; in prep) and Portuguese (PTG; Gonçalves 2003) spontaneous speech. Systematic extraction of every clause embedded under a matrix verb that governed a subjunctive at least once resulted in a dataset of nearly 5000 tokens, which were coded according to the mood selected, as well as a number of potential explanatory factors.

Results show robust variability, not only across languages, but also within them, since the same matrix verb can co-occur with both subjunctive and indicative in the same context. This is exemplified with the governor ‘*believe*’ in (1) – (4):

- (1) a. Je **crois** pas que ce **soit**_[SUBJ] la fin du monde. (FR.060.195)
‘I don’t think that it would be the end of the world.’
b. Je **crois** pas que l’âge **a**_[IND] tant à faire que ça. (FR.003.189)
‘I don’t think that age has that much to do with it.’
- (2) a. Eu **acredito** que **vá**_[SUBJ] **sair**. (PTG.143.356)
‘I believe that it will come out.’
b. Eu **acredito** que ele **devia**_[IND] **ter** em torno de setenta anos de idade. (PTG.99.148)
‘I believe that he must be around seventy years old.’
- (3) a. **Credo** che tutti lo **sappiate**_[SUBJ]. (IT.438.218)
‘I believe that everyone knows it.’
b. **Credo** che tutto **ritorna**_[IND]. (IT.511.264)
‘I believe that everything comes back.’
- (4) a. No **creo** que **haya**_[SUBJ] nadie aquí que no pague la renta. (SP.073.668)
‘I don’t think there’s anybody here who doesn’t pay rent’
b. No **creo** que **hay**_[IND] que firmar. (SP.086.555)
‘I don’t think you have to sign’.

Moreover, contrary to received wisdom, semantic considerations play a minor role, if any, in variant choice. With the arguable exception of SP, subjunctive selection is constrained neither by semantic classes (e.g. volitive, epistemic) of governors harmonic with its oft-ascribed meanings, nor by contextual elements (clause type, polarity, presence of other indicators of non-factual

modality) consistent with such meanings. On this basis, we place IT, FR and PTG farther along the *desemanticization* path than SP. But in all four languages, lexical bias is the major predictor of subjunctive selection. Tellingly, however, the “same” governor (whether determined etymologically or as translation equivalents) does not display consistent associations cross-linguistically. Thus ‘*fear*’ co-occurs with subjunctive 33% of the time in PTG, 64% of the time in FR and 100% of the time in IT and SP.

Therefore, to gauge productivity of the subjunctive, we put forward three measures: 1) the contribution of the governor (as instantiated by independent measures of its associated rate, the proportion it represents of the governor pool, and how much subjunctive morphology it accounts for), 2) the dispersion of subjunctive morphology across embedded verb types and 3) the extent to which the subjunctive is associated with particular structural contexts.

Results show that in each language, a handful of governors accounts for at least half of the governor pool and a large proportion of all subjunctive morphology. In addition, the cohort of embedded verbs carrying subjunctive morphology, though theoretically unrestricted, is also extremely limited. With near categorical subjunctive selection under the smallest number of governors and embedded verbs, FR is indisputably the least productive. But the usage facts militate against productivity in *all four* languages, with potential governors either highly or rarely associated with subjunctive, and the only variability occurring *among* governors and not within. Importantly, these associations are *community*-specific and not a function of meaning.

The relative positioning of SP with respect to the other languages on the cline of *desemanticization* should not obscure the fact that by these measures, it too displays *obligatorification*: indeed, it is the most sensitive to clause structure (negative, interrogative) and governor tense (e.g. conditional). While often taken to reflect a semantic contribution, these effects are also undeniably structural. Such associations with elements of the linguistic context indicate limitations on semantic motivations for subjunctive selection. Thus, although situated at different points on the cline of grammaticalization, all four languages are quite distant from the presumed source, and are all well embarked on the overriding processes of lexicalization and obligatorification.

- Carrier, A., De Mulder, W. & Lamiroy, B. 2012. Introduction: The pace of grammaticalization in a typological perspective. *Folia Linguistica* 46: 3 287–302.
- Cresti, E. & Moneglia, M. (eds.) (2005). *C-ORAL-ROM. Integrated reference corpora for Spoken Romance Languages*. Amsterdam, Philadelphia: John Benjamins.
- Gonçalves, S.C.L. (2003). Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista. <http://www.iboruna.ibilce.unesp.br/>.
- Harris, M. 1984. On the causes of word order change. *Lingua* 63: 175–204.
- Lehmann, C. 1995. *Thoughts on Grammaticalization*, 2nd revised edition. Munich: Lincom.
- Lindschouw, J. 2010. Grammaticalization and language comparison in the Romance mood system. In M. G. Becker & E.-M. Remberger (eds.), *Modality and mood in Romance: Modal interpretation, mood selection, and mood alternation*, 181–207. Berlin: de Gruyter.
- Martín Butragueño, P. & Lastra, Y. (eds.), 2011; 2012; in preparation. *Corpus sociolingüístico de la ciudad de México*. México: El Colegio de México.
<http://lef.colmex.mx/Sociolingüística/CSCM/Corpus.htm>
- Poplack, S. 1989. The care and handling of a mega-corpus. In R. Fasold & D. Schiffrin (eds.), *Language change and variation*. Amsterdam: Benjamins. 411–451.
- Posner, R. 1996. *The Romance languages*. Cambridge: Cambridge University Press.