

Grammatical change and hindcast model statistics – A comparison between Medieval French and Brazilian Portuguese

Eduardo Correa Soares

CLILLAC-ARP, Département d'Études Anglophones, École Doctorale Sciences du Langage, Université Paris Diderot – Paris VII, France

<https://doi.org/10.36505/ExLing-2016/07/0039/000298>

Abstract

This paper presents a methodology to analyse the ongoing linguistic change in Brazilian Portuguese[BP] as regards the pro-drop parameter. I propose to apply a hindcast statistical regression model to a sample of data from Medieval French[MF], whose outcome is the obligatory subject use in Modern French and to compare to a sample from BP. The results suggest that the change in such languages contrasts and is related to different reasons. While the change in MF appears to have uniformly gone toward non-pro-drop parameter, the BP change seems a by-product of semantic preference of null subjects to corefer to non-animated and non-specific antecedents.

Key words: Hindcast statistical model, grammatical change, pro-drop parameter, Brazilian Portuguese, Medieval French.

Introduction

This paper proposes a new methodology to address the grammatical change regarding the pro-drop parameter in Brazilian Portuguese[BP]. I propose that statistical hindcast regression model comparing Medieval French[MF] and BP may verify whether some assumptions about BP are akin to what came out in MF. This model is applied to two samples of data from MF and BP. The results show seemingly diverging patterns of linguistic change.

BP is taken to be a language on the way to become non-pro-drop (Tarallo 1983, Galves 1987, 1992, 1998, Duarte 1993, 1995, *inter alia*). In many standard pro-drop contexts in other Romance languages (for instance, European Portuguese, Spanish and Italian), an overt pronoun is indeed obligatory in nowadays colloquial spoken BP (see Duarte 1995, Barbosa, Duarte & Kato 2005, *inter alia*), such as in the example in (1) below.

- (1) então a gente lê pra ele₁ sentado ali... *(ele₁) gosta...
So the people read.pres.3s for him seated there he like.pres.3s
“So there we read for him₁ when seated down and he₁ likes that.”
(NURC-RJ, inquiry_011, data_set: “70s”)

Such contexts and data lead many works to suggest that BP is changing due to the simplification of agreement marking system, the so called Taraldsen's generalization (Roberts 1993, 2014, Kato 1999, 2000, *inter alia*). In this vein, it has been proposed that BP is following the same path by which French has passed from the MF to Modern French (notable exceptions to this claim are Kaiser 2009 and Roberts 2014). In MF, overt and null pronouns have been in apparent free variation, as in (2) below.

- (2) Aucassins₁ s' en est tornés / (...) Vers le palais ₋₁ est alés / *i*₁ en monta les degrés
 / une canbre ₋₁ est entrés / si ₋₁ comença # a plorer
 "Aucassin₁ departed/ to the palace he₁ went / he₁ went upstairs / (into) a bedroom
 he₁ entered / this way he₁ began to weep"
 (SRCMF, aucassin, data_set: "XII_century")

In the next section, I put forth a hindcast model statistics, by applying inferential logistic regression to data from MF and from BP.

Methodology

I propose to use a hindcast model to compare the change regarding the pro-drop parameter in BP and French. This methodology consists of (i) analysing a set of data from a specific period of time whose outcome is already known; (ii) statistically describing what has taken place and testing for some parameters and (iii) predicting possible similarities and differences from another set of data by changing or adding one or more parameters.

I have analysed MF change (Adams 1987) whose outcome has been the non-pro-drop status of modern French. I have compared this hindcast analysis of MF data to BP data in order to evaluate the status of the current so-called "on-going" change in BP. I have taken 9 texts from the historical corpus of MF *SRCMF*¹, 6 interviews of BP *NURC-RJ* corpus (3 carried out in the 70s and 3 in the 90s)² and 3 movie subtitles produced after 2010 from the *OPUS corpora* project³. These texts were automatically annotated. The sample was gathered by a concordance toolkit. The MF subcorpus was thus constituted of 1500 sentences (a half of them without subject), distributed into 3 subsets of data according to the year of the text (*group1*, the IXth and Xth centuries; *group2*, the XIth and XIIth centuries; and *group3*, from the XIIIth century on). The BP corpus was equally formed by 1500 sentences (50% of subjectless sentences) and split into 3 subsets: *group1*, data from 70s; *group2*, from the 90s; and *group3*, data from 2010 on. The collected data was then analysed with a Generalized Linear Model using the software R, with the packages *lme4*, *languageR* and *stats*.

Results

Table 1 sums up the logistic regression analysis and the results obtained. In French, the so-called impoverishment of agreement marking has predominantly affected singular forms and 3rd person plural. The fixation of non-pro-drop in MF is taken to be a strong effect of such an impoverishment (Adams 1987, Roberts 1993a). If Taraldsen's generalization is correct, it is expected that 1st and 2nd person plural are significantly more null subjects than the others. But this prediction does not hold. What the data have shown is a gradual increasing in the number of overt subjects regardless the verbal inflection, and no significant difference along the time and among the person markings. In BP, however, the number of null subjects is stable across the discourse persons and the periods in the last 50 years, except for 1st and 3rd person singular. In a further statistical regression, I have analysed the features animacy and specificity (previously suggested in the literature about BP by Cyrino *et al.* 2000). In MF, these features were not significant in any statistical regression. In BP, however both non-animated and non-specific are relevant in the increasing number of 3rd person singular null subjects (P-value: 0.00615 and 0.00771 respectively).

Table 1. Logistic regression analysis of Medieval French (MF) data and Brazilian Portuguese (BP) data (int = intercept term)

period/person	1_sing	2_sing	3_sing	1_pl	2_pl	3_pl
group1	int	int	int	int	int	int
group2	MF(ns) BP(.)	MF(ns) BP(ns)	MF(ns) BP(**)	int	MF(ns) BP(ns)	MF(ns) BP(ns)
group3	MF(ns) BP(*)	MF(ns) BP(ns)	MF(ns) BP(***)	int	MF(ns) BP(ns)	MF(ns) BP(ns)
Signif. codes: Pr(> z) 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 'ns' 1						

Discussion

This study suggests that the null subjects in BP are becoming scarcer in a way different from MF. Firstly, the null subject in BP is crucially likely to be 3rd person singular. This person is the less marked form in BP (Kato 1999). In MF, no significant difference concerning person, animacy or specificity was found. The MF change can also be related to other factors (e.g. the use of clitic subject pronouns). In BP such a difference seems to be a semantic-functionally motivated by-product of two factors – the semantic features animacy and specificity. This difference can be crucial to shed light on the *partial* pro-drop status of BP (Biberauer *et al.* 2010) and the *non*-pro-drop status of modern French (Adams 1987).

Notes

1. See <http://srcmf.org> and Prévost & Stein (2013) for more information.
2. This corpus is available online in <http://www.lettras.ufrj.br/nurc-rj/>.
3. See <http://opus.lingfil.uu.se/OpenSubtitles2016/> and Lison & Tiedmann (2016).

Acknowledgements

I am thankful to CAPES Foundation for providing me the financial support of this research and to my supervisors Philip Miller, Barbara Hemforth and Sergio Menuzzi, who give me *to-the-point* advice to carry out my projects.

References

- Adams, M. 1987. From Old French to the Theory of Pro-drop. *Natural Language and Linguistic Theory* 5: 1-32.
- Barbosa, P., Duarte, M. E. L. & Kato, M. A.. 2005. Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 4. v. 2: 11-52.
- Biberauer, T. Holmberg, A. Roberts I., Sheehan, M. 2010. *Parametric Variation*. CUP.
- Cyrino, S. M.L.; Duarte, M.E. L., Kato, M. A. Visible subjects and invisible clitics in Brazilian Portuguese. In: Kato, M.A. & Negrão, E.V. (eds.). 2000. 55-104.
- Duarte, M. E. L. 1993. Do pronome nulo ao pronome pleno. In Roberts, I., Kato, M.A. (eds.): 107-28.
- Duarte, M.E.L. 1995. A Perda do Princípio "Evite pronome" no Português Brasileiro. Campinas, SP, UNICAMP: Ph.D. Dissertation.
- Galves, C. 1987. A sintaxe do português brasileiro. *Ensaio de lingQística* 13: 31-50.
- Galvez, C. 1993. O enfraquecimento da concordância no Português Brasileiro. In: Roberts & Kato (eds.): 387-408.
- Galves, C. 1998. Tópicos e sujeitos, pronomes e concordância no português do Brasil. *Cadernos de Estudos Linguísticos*, 34: 19-32.
- Kaiser, G. A. 2009. Losing the null subject. A contrastive study of (Brazilian) Portuguese and (Medieval) French. In *Proceedings of the Workshop Null-subjects, expletives, and locatives in Romance*: 131–156.
- Kato, M. A. 1999. Strong pronouns, weak pronominals and the null subject parameter. *PROBUS*, 11, 1: 1-37.
- Kato, M. A. 2000 The partial pro-drop nature and the restricted VS order in Brazilian Portuguese. In: Kato, M. A. & Negrão, E. V. (eds). 2000. 223-258.
- Kato, M. A. & Negrão, E. V. (eds). 2000. *The Null Subject Parameter in Brazilian Portuguese*. Frankfurt-Madrid: Vervuert-IberoAmericana.
- Lison, P., Tiedemann, J. 2016. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*
- Prévost, S.; Stein, A. 2013. *Syntactic Reference Corpus of Medieval French (SRCMF)*. ENS de Lyon/ILR Stuttgart.
- Roberts, I. 2014. Taraldsen's Generalization and Language Change. Prepublished ms.
- Roberts, I. & Kato, M. A.(eds.) 1993. *Português Brasileiro: uma viagem diacrônica (Homenagem a Fernando Tarallo)*. Campinas, SP: Editora da UNICAMP.
- Tarallo, F. 1983. *Relativization Strategies in Brazilian Portuguese*. University of Pennsylvania: Ph.D. Dissertation.