

Empirical evidence shows the selection of auxiliary *ter* explains the delayed generalization of the Portuguese compound pluperfect

William Balla-Johnson (University of North Georgia)

The Romance compound tenses have garnered much attention from linguistics over the past century. However, the pluperfect has often been overlooked in favor of the present perfect in particular, but also the future and the conditional. The pluperfect deserves attention as an object of study in its own right since it has been observed to not only follow, but to lead changes that subsequently occur in the present perfect (Balla-Johnson 2020), as well as develop independently from other tenses. This paper uses large scale empirical data to describe the historical development of the Portuguese pluperfect and concludes that the generalization of auxiliary *ter* is the primary reason the Latin synthetic pluperfect form has survived with unchanged semantics longer in Portuguese (and Galician) than anywhere else in Romance.

The Portuguese pluperfect tense is fairly unique in the Romance world. In Portuguese, the Latin synthetic pluperfect is maintained, though phonologically evolved (Wanner 2014): Lat. *cantaveram* > Port. *cantara* 'I had sung'. However, this form has been reserved for elaborate literary usage since the 19th century (Squartini 1999), while in all other registers the pluperfect is formed as a compound tense generally with the auxiliary *ter* < Lat. *tenere* 'to hold/possess' (Harre 1991) in the imperfect with a past participle, ex. *eu tinha cantado* 'I had sung'. The auxiliary *haver* 'to have' is also available, though much less frequently selected, and in earlier Portuguese the auxiliary *ser* 'to be' was used for the pluperfect of unaccusative verbs (Brocardo 2010). Historically, there was competition between the synthetic and the compound forms of the Portuguese pluperfect. However, Squartini (1999) finds that the compound and the synthetic pluperfect are not exactly in complementary distribution in 19th century literary Portuguese. While the compound pluperfect can have both perfect and perfective readings in the past, the synthetic pluperfect has lost the ability to refer to perfect events in the past, observed by its inability to collocate with reference time temporal adverbs. Additionally, there was another compound form available, here referred to as the *past anterior* consisting of an auxiliary verb in the preterit and a past participle (ex. *tive cantado* 'I had sung'). This form was common in the 13th and 14th centuries, but declined sharply in use thereafter.

Therefore, the historical Portuguese pluperfect was complex with variation evident between the synthetic pluperfect, the compound pluperfect, and the past anterior, as well as auxiliary variation within the compound pluperfect and the past anterior. This paper uses empirical methods and corpus data from the Corpus do Português (Davies & Ferreira 2006) to comprehensively describe the historical development of the Portuguese pluperfect between the 13th and the 18th centuries and to answer the following question:

- Why did the synthetic pluperfect descended from the Latin forms in -VERAM (ex. CANTAVERAM 'I had sung') survive with its original interpretation longer in Portuguese and Galician (ex. *cantara* 'I had sung') than in any other area of Romance?

First, in order to describe the historical development of the Portuguese pluperfect, several queries were run in the corpus. The adjusted tokens per million words was calculated for the synthetic pluperfect, the compound pluperfect, and the past anterior based on proportions observed in random samples of 100 tokens per verb form, per century. The results are shown below in figure 1. Likewise, the adjusted tokens per million words was calculated for each of the auxiliaries used in the compound pluperfect (figure 2) and the past anterior (figure 3). These

show that: (1) the past anterior all but disappeared by the 17th century, (2) the synthetic pluperfect was declining but still the most frequent form in the 18th century, and (3) *ter* became the overwhelmingly preferred auxiliary for the compound pluperfect by the 16th century.

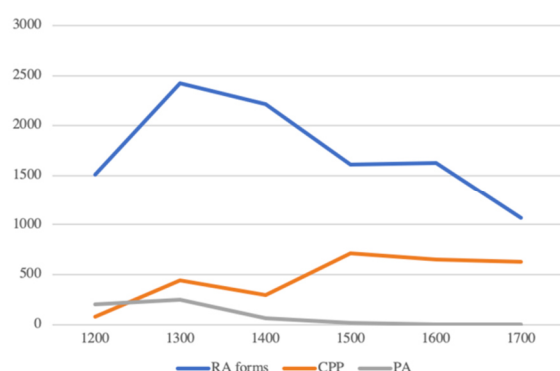


Figure 2 – Adjusted tokens per million words for Portuguese pluperfect forms

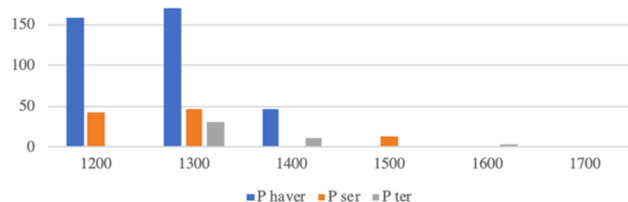


Figure 4 - Adjusted tokens per million words of auxiliaries for the Portuguese past anterior

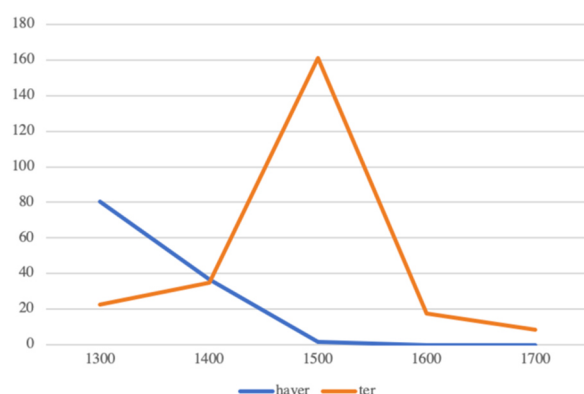


Figure 3 - Estimated compound pluperfect tokens per million showing agreement (non-masculine singular)

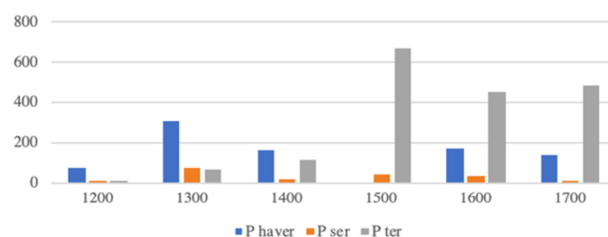


Figure 1 - Adjusted tokens per million words of auxiliaries for the Portuguese compound pluperfect

The expansion of *ter* was then further investigated. Quantitative data from the Corpus do Português was collected to show that the generalization of *ter* was well established in both the present perfect and the compound pluperfect by the 16th century. It was then empirically shown (Figure 4) that there was more participle-complement agreement in the constructions with *ter* than with *haver*, indicating a greater proportion of resultative, rather than eventive, verbal meanings. This led to a specialized meaning of the compound pluperfect (one indicating some amount of resultativity) which delayed a complete semantic overlap with the synthetic pluperfect, at least until participle-complement agreement was lost. That overlap, present at an earlier stage in French and Spanish, resulted in a change in meaning or the complete loss of the synthetic pluperfect, a stage not yet attained in Portuguese.

The generalization of *ter* as the perfect and pluperfect auxiliary in Portuguese (and Galician) delayed the semantic grammaticalization of the compound pluperfect. Therefore, the compound pluperfect and the synthetic pluperfect did not enter into a situation of free variation until much later than in other areas of Romance. It was only after the compound pluperfect could be used with the same semantics as the synthetic pluperfect that the latter could fall into disuse, a situation which did not attain until the 19th or 20th century.