

On the usefulness of using parallel texts in diachronic investigations: insights from a parallel corpus of Spanish medieval Bible translations

**Andrés Enrique-Arias
Universitat de les Illes Balears**

This paper discusses some of the advantages of the *Biblia Medieval* (BM) corpus as opposed to other data sources in the historical study of morphosyntactic variation and change. Containing over 5 million words and 16,000 digital images, BM is a freely accessible online tool that enables linguists to consult and compare side-by-side the existing medieval Spanish versions of the Bible and their Latin or Hebrew originals (www.bibliamedieval.es/corpus).

Most historical investigations of Spanish nowadays use one or both of the two large diachronic databases that are currently available: the *CORDE* (300 million words) or the *Corpus del español* (100 million words). However, despite the many advantages afforded by these conventional corpora the historical study of morphosyntactic variation and change with these databases faces certain limitations in areas in which, as I will try to demonstrate, a parallel corpus like BM can afford clear advantages.

One such area is *perspective*, understood as how the information is accessed in a corpus. In conventional corpora users enter queries for the forms that are relevant for the research question they are investigating. This means it is necessary to know beforehand, via historical grammars, dictionaries or previous studies, what are the possible expression units for the structure that is being investigated. Two disadvantages emerge from this. One is that, no matter how well we do our research of reference materials, there is always a risk that some relevant form will be overlooked because it has never been studied. Another problem is that, as research with conventional corpora gravitates towards searching for explicit markers, it is impossible to identify all the occurrences of linguistic phenomena that may be expressed in a great variety of ways, or even zero-marked. Conventional historical corpora also have limitations in the area of *comparability*. It is rather difficult to identify and to define the contexts of occurrence of linguistic variants, since these are normally conditioned by a complex set of syntactic, semantic and discursive factors. In order to minimize this problem, it would be necessary to locate and to examine a large number of occurrences of the same linguistic structure in versions that were produced at different time periods. Ideally, these occurrences should proceed from texts that have been influenced by the same textual conventions. Conventional corpora, however, are not well-suited for locating such occurrences.

These problems are alleviated when the linguist has access to a parallel corpus such as BM. A parallel corpus consists of texts that are translated equivalents of a single original. Therefore they have the same underlying content and have been influenced by similar textual conventions. The perspective afforded by parallel corpus methodology is more open as it is possible to look up a form in the source languages or in one of the parallel versions and then observe the forms that are used in the same context and with the same function in the parallel versions. In turn, we can search for these forms, which will yield more forms and more contexts than can be used for further queries. This perspective, from particular textually embedded contents to form, facilitates the observation of elements that otherwise would have been overlooked. Likewise, the

parallel texts in BM offer direct comparability of concrete examples, as translation equivalents are likely to be inserted in the same or very similar, syntactic, semantic and pragmatic contexts of occurrence. Finally, in BM it is possible to analyze stylistic variation in a more controlled manner by examining how the same translator selects different linguistic options depending on the subgenres represented in the Bible (i.e. narrative, lyrical poetry, wisdom literature, prophecies, etc).

In order to illustrate the possibilities that BM offers, this paper gives an overview of specific case studies of Spanish morphosyntactic variation: discourse markers, possessive structures and clitic related phenomena. In sum, this study demonstrates how a parallel corpus of medieval biblical texts may enrich our theoretical understanding of change and variation phenomena in Old Spanish while overcoming some of the limitations of conventional corpora.