A corpus analysis of deverbalization: Spanish temporal hacer 'make'

Borja Herce

1. Introduction and background

Spanish time constructions involving the verb *hacer* (to make) appear in two different constructions: a so-called (Real Academia Española 2009) clausal construction (1), where *hacer* appears to be the main verb in the sentence and an adverbial construction (2), where *hacer* introduces a syntactically optional constituent:

(1) Hace mucho que fui a la escuela. (Clausal construction)
   *It has been a long time since I went*  

(2) Fui hacer mucho. (Adverbial construction)
   *I went a long time ago*

Deverbal expressions like Spanish *hace*, English *ago* etc. tend to be quite idiosyncratic cross-linguistically (Culicover 1999: 71-74; Kurzon 2008; Plank 2011). The Spanish constructions have been the object of both diachronic (Pérez Toral 1992) and synchronic (Rasmussen 1981; Rigau 1999; Howe 2011; Brucart 2015) studies but their current status remains controversial. Some regard (1) and (2) as derivationally related and others as independent from one another, some consider *hacer* in these constructions adpositional and others verbal. These differences are not only the result of different theoretical standpoints, but also of many disagreements concerning the actual grammatical possibilities of the two constructions. Without aiming at exhaustivity, disagreements include the possibility for pronominalization of the object NP, the TAM morphological possibilities of *hacer* or the possibility to have time adjuncts in the adverbial construction. This illustrates the necessity for linguists to move away from individual introspective grammaticality judgments and towards quantitative empirical data.

2. Objectives and methodology

My aim in the current investigation has been to obtain a clearer picture of the synchronic grammatical possibilities of these constructions, as well as of the diachronic developments leading to the current situation. This I have done by analyzing synchronic and diachronic computerized corpora (CREA and CORDE) as well as by systematically gathering acceptability judgments from native speakers on some of the less clear cases. This hopefully illustrates how quantitative empirical evidence can help decide between different formal analyses and inform about the underlying grammatical competence at different points in time.

Earlier quantitative, corpus-based analyses of these constructions, like Howe (2011), have not yielded the expected results. In my opinion, they are adversely affected by a lack of a consistent separation of the data from the clausal and adverbial constructions (which demonstrably have very different properties regarding word order, negation, time adjunction, TAM morphology etc.) as well as by the impossibility to enquire about the less frequent uses, some of which will be absent entirely from a corpus without that implying they are impossible. This is why different data and methodologies should be used to solve these riddles. This research is based on the analysis of 1500 tokens of these time constructions in different periods and on the acceptability ratings for various sentences provided by 35 native speakers for the less clear properties.

3. Results and implications

As I advanced earlier, synchronic usage data from CREA paint a very different picture for clausal and adverbial constructions. Whereas in the former TAM morphology or time adjunction appear to occur as in other verbs, in the latter they are all severely restricted. For example, I found that only 5.3% of the temporal instances of present tense *hace* appear in clausal constructions compared to a 55% of the temporal tokens of imperfect *hacia*, which illustrates the greater morphological flexibility of the clausal construction.

Native speaker intuitions, in addition, inform us that adverbial constructions displaying time adjunction, non-present verb forms, negation etc. are degraded, these sentences having been rated with only around 2 out of 5 on a Likert-scale. Moreover, differences between younger and older speakers seem to suggest a grammatical change in progress in the adverbial construction. The wildly different properties displayed by the two constructions argue in favour of a synchronic independence of the two constructions and against some previous formal analyses (Rigau 1999; Fábregas to appear) which favour a derivational relation.
Diachronic quantitative evidence, in turn, shows that a major change took place in the adverbial construction between the 17th and the 19th centuries. That period witnessed an exponential increase in the raw frequency of the present tense collocates of this construction, a development which we know (e.g. Bybee 2010) frequently accompanies grammaticalization:

![Graphic 1]

Before that, the two constructions appear to show exactly the same quantitative patterns regarding the various parameters analyzed (i.e. TAM morphology, time adjunction, word order, negation and semantics) whereas in later periods a progressively wider gap is found between them. We witness a desententialization of the adverbial construction: near-disappearance of time adjuncts, fixation of word order, near-loss of TAM morphology etc.

In conclusion, all the data gathered, both synchronic and diachronic, seem to tell a consistent story whereby what was likely originally a single construction diverged progressively into two: a clausal construction involving a full verb hacer and a non-sentential, adverbial construction with a near-prepositional hace. Quantitative, replicable evidence can and should, therefore, be used to understand and describe more deeply the internal dynamics of change. Moreover, quantitative data external to the observer are a must if we want to understand the actual possibilities of less frequent or less clear-cut grammatical properties.

4. Bibliography


