The use of inflected infinitive in a spoken corpus

Fernanda CANEVER (USP)

As opposed to the most dominant trends in linguistics, which have focused on language as a more or less fixed system, a usage-based model such as Langacker's Cognitive Grammar takes our linguistic systems to be grounded in usage events (Langacker 2000, Barlow & Kemmer 2000). That is, once our linguistic knowledge is formed on the basis of instances of concrete use, this system or "inventory" of cognitive routines is constantly reshaped through actual usage in a kind of feedback loop between instances and their representations (Barlow & Kemmer 2000).

Through actual language use, novel structures may become part of the system. That is, the concrete use of language structures in the daily life of a speech community is crucial for the emergence of new patterns in the system.

As for the storage of such novel structures, Langacker explained that when a non-conventional structure gets into the system, it may be reinforced by frequent use or disappear due to non-use. What is central to this process is the cognitive ability of habit formation, which Langacker refers to as entrenchment: the more frequent an element is, the more entrenched it becomes. Not only does the frequency of occurrence of an element (token frequency) but also the frequency of a pattern (type frequency) affect speakers' linguistic knowledge and play a role in the characterization of a structure as being conventional. The convention is always changing (Croft 2000:132), as well as the notion of grammaticality, which according to Bybee (2010) is gradient. Bybee (2006a, 2006b) claims that increase in token and type frequencies may influence speakers in such a way that the most frequent variant of a particular linguistic phenomenon gradually ends up sounding more correct.

Because in the usage-based approach frequency is correlated to the degree of entrenchment of constructions in speakers' grammars, quantification of variation in corpora is essential to investigate how entrenched constructions are in the grammars of the speakers and to better understand how speakers' ideas regarding what is conventionalized change over time. As usage has been more and more acknowledged in current linguistic research and frequency studies have become crucial in usage-based theories and analyses, advances in computational and corpus linguistics have facilitated studies with real data. This means that we are now able to manipulate large corpora and large amounts of data in an automatic fashion, and this is the realm this study belongs to. To validate the hypothesis that use constantly shapes speakers' grammars, this usage-based study intends to explore the question of how innovations may be incorporated in our linguistic systems by investigating the inflection of the infinitive in spoken corpora of Brazilian Portuguese.

According to Maurer (1968), since its emergence in Portuguese, the inflection of the infinitive gradually spread to different constructions and in many cases has become the conventionalized variants. In numerous constructions, on the other hand, the inflection was considered optional and dependent on the free choice of the speakers, who could prefer to emphasize the grammatical person or not. Regarding the variation at the time he did his frequency study, Maurer stated that the uninflected variants were more frequent and sounded more natural.

Recently, though, Canever (2011) investigated the frequency of occurrence of some of the so-called optional inflected infinitive constructions mentioned by Maurer (1968) in a corpus of standard written language, more specifically a corpus of academic written Brazilian Portuguese. The results reveal a high frequency of occurrence of the inflected variants, suggesting that the inflected forms are highly entrenched in the grammars of the investigated speakers.

Using the same methodology used in Canever (2011), that is, the software R for the data extraction and manipulation, this study investigates the frequency of occurrence of the inflected infinitive in a corpus of academic spoken Portuguese formed by transcriptions of classes at the University of São Paulo.

In this talk I will (i) discuss issues concerning the data manipulation with R, (ii) discuss issues related to the spoken corpora used, and (iii) present the partial results of the study, which, as predicted, suggest the inflection is also frequent in the spoken variety.