Morphological typology in a contrastive setting: causative constructions in Spanish and Italian and their equivalents in Czech (a parallel corpus-based study)

PAVEL ŠTICHAUER & PETR ČERMÁK

Charles University in Prague, Faculty of Arts, Institute of Romance Studies
pavel.stichauer@ff.cuni.cz
petr.cermak@ff.cuni.cz

Causative constructions (or, more generally, causativity or causative alternation) are undoubtedly of major interest in a wide range of theoretical frameworks. The central question of whether analytical (i.e. syntactically expressed) constructions (of the type *Bill made Phil sing*) and morphologically realized constructions (through affixation, for example) can be accounted for within a unified syntax-based model (cf., e.g., Li 2005; Sadler - Spencer 1998: 226-231) might also be answered from the perspective of an empirically oriented - that is, a corpus-based - contrastive study of typologically different languages. This is precisely what we wish to carry out in the study being submitted.

Spanish and Italian (along with other Romance languages) are known to have a totally productive and unrestricted causative construction with the light verb *hacer / fare* + infinitive (*hacer pensar / far pensare* 'to make think'), but Czech, a West Slavic language which will be the target language of our comparison, lacks this analytical construction displaying other structural possibilities of expressing causativity. One of these possibilities is, as we might expect, a morphological realization through affixation. In fact, one of the traditional view of how to translate a causative construction into Czech is by using a specific causative prefix *roz-* (*plakat* 'weep' / *roz-plakat* 'to make weep'). In a wide range of languages, this kind of morphological causativization is quite common and completely regular; this situation, as Sadler and Spencer claim (1998: 227), has led "*many researchers [to] regard the morphological causative as an instance of an argument-structure alternation, rather than lexemic derivation proper. This is particularly attractive when the causative is completely productive and lacking in lexical idiosyncrasies.*"

In Czech, however, the prefix *roz-* is a very constrained affix which is limited in use not only by its productivity, but especially by the restrictions imposed on the verb selection: only a very limited subset of verbs allows for the *roz-* prefixation. Moreover, the attested prefixed verbs display also some unpredictable idiosyncrasies. It is therefore evident that an attempt at drawing a contrastive typology cannot possibly result in such a "clean" binary picture. In Czech, there is thus more variation as far as the range of structural possibilities is concerned. The overall picture of what exactly these possibilities are and whether they can be quantified is precisely the purpose of the work to be presented. From the methodological point of view, this research is strictly corpus-based: all the data we will present have been obtained from an Italian-Czech / Spanish-Czech parallel corpus InterCorp (for the theoretical description of parallel corpora in general, cf. Aijmer 2008) which is part of the Czech National Corpus (www.korpus.cz).

We intend to present the results in two major steps. First, on the basis of the qualitative analysis of the data we will draw a typology of Czech equivalents of Spanish / Italian causative constructions. Second, we will provide a quantitative picture of the distribution of the individual types.

As far as the first step is concerned, we have established a typology of equivalents (or, to be entirely correct, "translation solutions") that one might come across when sampling a large corpus. We have defined six major types: 1) the *roz-* prefixation; 2) the lexically

autonomous and lexemic equivalent, i.e. a verb which is not a complex verb and which conveys causativity as a lexically specified feature of the meaning; 3) the analytical / syntactic construction structurally equivalent to the Spanish / Italian construction (e.g. *dát pocítit* lit. 'to give feel'); 4) other analytical constructions, especially idiomatic multi-word expressions; 5) the bi-clausal construction, i.e. a sentence with a clause specifically expressing the causativity; 6) the change of syntactic functions (e.g., the subject-object alternation where the object - identical to the subject of the base verb - introduced by the causative construction is syntactically realized or maintained as a subject in the targer language).

As for the second step, we will present (along with the corpus size) the frequency lists of the most frequent verbs which are found in the causative constructions. The frequency spectra of Spanish and Italian verbs correspond perfectly to the typical word frequency distribution being there a limited group of very frequent verbs, on the one hand, and a large number of *hapaxes*, on the other hand (cf., e.g., Baroni 2009). Since we wish to proceed to a quantification of the most frequent types defined above, we have set up a frequency limit on which verbs we will deal with. It has turned out that the verbs whose token frequency is less than 10 tokens cannot be taken into consideration because of the simple fact that in this case the number of "translation solutions" is very low and these solutions tend to be distorted in different ways (for example, by the unbalanced textual typology, or by some "idiolectal effects" due to one concrete translator).

Finally, on the basis of the data elaborated in this concrete way, we will offer a quantitative picture of the frequency of the individual types. Anticipating partly the results we will put forward, it turns out that in Czech out of the six types defined above only two can be safely said to dominate (it is the type 2 and 6), while the allegedly dominating *roz-* prefixation is by far the least represented equivalent. We will discuss the theoretical implications of these findings.

We believe that this strictly empirical study might also shed some light on the theoretical issues briefly mentioned above. We will argue that from the point of view of morphological typology it is very important to investigate not only the "pure" types (syntactically *versus* morphologically realized causativity), but especially the "mixed" types, i.e. languages which display a wider range of structural possibilities. Such an investigation might be thus considered as a modest contribution to a modern morphological typology.

References

AIJMER, K. (2008). Parallel and comparable corpora. In Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*, vol. 1. Berlin: Mouton de Gruyter, pp. 275-292.

BARONI, M. (2009). Distributions in text. In Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*, vol. 2. Berlin: Mouton de Gruyter, pp. 803-822.

LI, Y. (2005). *X°. A Theory of the Morphology-Syntax Interface*. Cambridge (MA): The MIT Press.

SADLER, L. - SPENCER, A. (1998). Morphology and Argument Structure. In Andrew Spencer and Arnold M. Zwicky (eds.), *The Handbook of Morphology*. Oxford: Blackwell, pp. 206-236.