

Comparing morphological and syntactic variations of support verb constructions and verbal full phrasemes in French: a corpus based study

Agnès Tutin, LIDILEM,
Université Grenoble Alpes, agnes.tutin@univ-grenoble-alpes.fr
Domaine Universitaire - 621 avenue Centrale
38400 Saint-Martin-d'Hères

1 Introduction

This paper deals with syntactic and morphological variations of verbal MWEs in French. Our objective was to check against corpus evidence some assumptions of the literature concerning MWEs, and more precisely verbal MWEs, which are often said to be quite variable (e.g. Nunberg et al. 1994; Moon 1998). We wanted to check to what extent this claim was proved to be true in a corpus study for 30 frequent verbal MWEs in French, by comparing the syntactic and morphological variations between non compositional MWEs and verbal collocations, particularly support verb constructions (hereafter SVCs), which are very frequent in French. Being semi-compositional, SVCs should be more flexible across variations than full phrasemes.

2 Different types of verbal MWEs and selection of verbal MWEs for our experiment

Following Mel'čuk (2012), we distinguish “Full phrasemes”, similar to Baldwin and Kim's (2010) idiomatic combinations which are non-compositional expressions, e.g. *take into account*, from collocations, which are compositional but are difficult to predict e.g. *pay attention*, or *heavy smoker*. Support verb constructions (Gross, 1981) (hereafter SVCs) or light verb constructions are a kind of collocations for which: a) the verb is semantically weak and has a very general meaning (*to take a walk*, *to have a shower*). b) the predicate is supported by the noun which selects the semantic arguments. For our experiment, we selected 30 verbal MWEs, equally divided in the 2 categories of SVCs and Full phrasemes. The verbal MWEs have the same construction, V (Det) N, and include very frequent verbs such as *faire* (‘make’) or *avoir* (‘have’), e.g. full phrasemes such as *avoir affaire* (lit. ‘have deal’) and collocations such as *avoir peur* (lit. ‘have fear’). We also ensured that they were sufficiently frequent in the treebanks we used, which include 32 million words divided in two genres: contemporary literature and newspapers¹. These texts were syntactically annotated with the help of the dependency parser Connexor (Tapanainen & Järvinen, 1997).

3 Extracting syntactic and morphological properties from texts

The properties were extracted using a corpus-tool called the Lexicoscope (Kraif & Diwersy 2014), in order to observe syntactic and morphological variations of verbal MWEs. Several properties have been selected in order to assess the variability of our verbal expressions, by looking at similar work in the literature (Moirón, 2005; Laporte *et al.*, 2008; Grégoire, 2010; Weller & Heid, 2010). Regular variations concerning verbal MWEs were not taken into account, e.g. insertion of an adverb between the verb and (Det) N. The 5 following properties were chosen for every MWE of the form V (Det) N:

- a) **Plural form:** Does the noun occur with the plural form? For example *peur* in *avoir (DET) peur* (‘have fear’) is always singular. Conversely, *question* is often plural in *poser (DET) question* (‘ask (DET) question’).
- b) **Determiner variation:** Is the determiner (or the absence of determiner) fixed? For example, *avoir peur* can occur without a determiner or with a determiner in contexts such as *avoir une peur folle* (‘have a great fear’). Conversely, for *céder la place* (‘give the way’) the only option is *la*.

¹ The treebanks are available on the Lexicoscope website : <http://phrasotext.u-grenoble3.fr/lexicoscope/>

- c) **Relative construction:** Is the relative construction possible, for example, *le rôle qu'il a joué* ('the role that he played')?
- d) **Passive voice and reduced passive:** Can the expression occur in the passive voice (*le rôle est joué*, 'the role is played') or in reduced passive, without the auxiliary (*l'attention prêtée à ...* 'attention paid to ...').
- e) **Modified noun:** Can the noun be modified with an adjective, for example for *avoir un mal fou* ('to have great difficulty')?

4 Results : variation and type of verbal MWE

In our experiment, the variation properties were automatically extracted and then manually checked against concordances in order to check semantic consistency, because most MWEs could be confused with either the literal meaning or other senses of the MWE (e.g. *avoir du mal* ('to have trouble') vs *avoir mal* ('to feel pain') vs *avoir le mal de* ('to be sick')). Figure 1 presents the variation of morphological and syntactic properties according to the type of verbal MWEs. It shows very clearly the strong relationship between the semantic type of the MWE

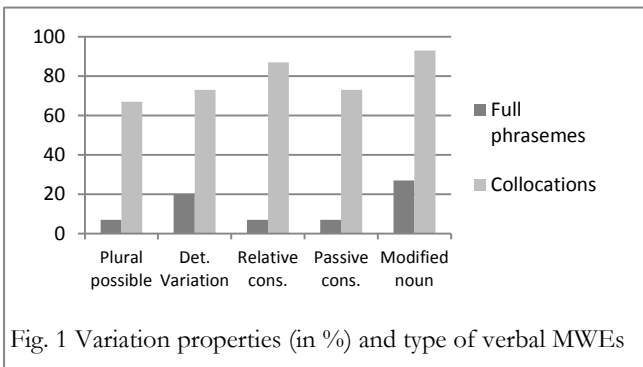


Fig. 1 Variation properties (in %) and type of verbal MWEs

and variability. Few opaque verbal MWEs under examination were likely to vary. Amongst properties, the strongest ones are plural variation, relative and passive constructions which were not encountered with any full phraseme, while plural variation was only possible for about 60% of the SVCs. As expected, SVCs were more likely to be variable, especially for syntactic alternations and the noun modifier.

If we sort the MWEs along a continuum from the least variable to the most variable ones (see Fig. 2), we can observe that only two MWEs (with a black diamond) really have an unexpected behavior: the collocation *avoir recours* ('use') does not have any variation, while the full phraseme *avoir du mal* ('have trouble') seems as flexible as a collocation, in spite of its semantic non-compositionality (because *mal* does not have its usual meaning).

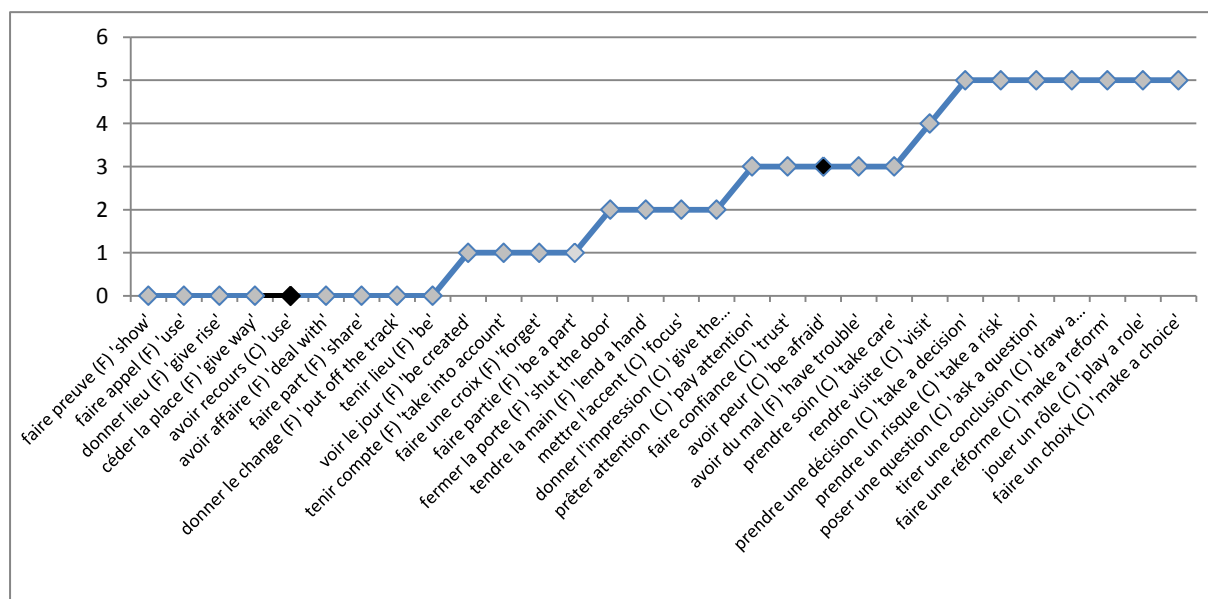


Fig. 2: Classification of verbal MWEs along a continuum according to the number of syntactic and morphological variations (F: full phrasemes. C: collocations)

5 Conclusion and Future work

This pilot study on verbal MWEs in French confirms some assumptions on the topic of MWEs: as a general trend, semantic properties and syntactic properties go hand in hand, as observed by Wulff (2008) in English. The less compositional, the less variable are MWEs. SVCs thus show a high level of variability, but are nevertheless far from being “freely variable”. However, in general, verbal full phrasemes do not seem highly variable, which contradicts some assumptions (e.g. Moon 1998). However, the variability of MWEs in general should perhaps not be overestimated. A surprising fact observed in this experiment is the pervasiveness of semantic ambiguity among frequent verbal MWEs, probably less frequent with longer MWEs such as *prendre le taureau par les cornes* (lit ‘take the bull by the horns’, ‘to face problems’). Obviously, this study needs to be extended with a larger number of expressions in order to confirm these results for French, and to better understand the linguistic nature of MWEs.

References

- Baldwin, T., Kim, S. N. 2010. Multiword Expressions. *Handbook of natural language processing*, 2, 267-292.
- Grégoire, N. 2010. DuELME: a Dutch electronic lexicon of multiword expressions. *Language Resources and Evaluation*, 44(1-2), 23-39.
- Gross, M. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages*, (63), 7-52.
- Kraif, O., Diwersy, S. 2014. Exploring combinatorial profiles using lexicograms on a parsed corpus. *Nouvelles perspectives en sémantique lexicale et en organisation du discours*, eds. P. Blumenthal, I. Novakova & D. Siepmann, 381–394. Frankfurt: Peter Lang.
- Laporte, É., Ranchhod, E., Yannacopoulou, A. 2008. Syntactic variation of support verb constructions. *Linguisticae Investigationes*, 31(2), 173-185.
- Mel’čuk, I. 2012. Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3(1), 31-56.
- Moirón, B. V. 2005. Linguistically enriched corpora for establishing variation in support verb constructions. In *Proceedings of the 6th International Workshop on Linguistically Interpreted Corpora (LINC-2005)*, Jeju Island, Republic of Korea.
- Moon, R. 1998. *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Oxford University Press.
- Nunberg, G., Sag, I. A., Wasow, T. 1994. Idioms. *Language*, 491-538.
- Tapanainen, P., Järvinen, T. 1997. A non-projective dependency parser. *Proceedings of the 5th Conference on Applied Natural Language Processing*, Association for Computational Linguistics, Stroudsburg, 64-71.
- Weller, M., Heid, U. 2010. Extraction of German multiword expressions from parsed corpora using context features. *Proceedings of LREC 2010*, 3195–3201.
- Wulff, S. 2010. *Rethinking idiomaticity: A usage-based approach*. London: A&C Black.