

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

LIDILEM



Agnès Tutin
Université de Grenoble Alpes, Laboratoire LIDILEM

Aim

With a corpus study, check against corpus evidence some assumptions of the literature concerning verbal MWEs :

- 1) Are verbal idioms really variable?
- 2) By comparing the syntactic and morphological variations, are really support verb constructions (included in the class of collocations) more variable than true verbal idioms?

Definitions

Support verb constructions and full phrasemes (true idioms)

- ✓ **Support verb constructions** (or light verb constructions) (Gross, 1981):
 - a) the verb is semantically weak and has a very general meaning (*faire une promenade* vs *to take a walk*)
 - b) the predicate is supported by the noun which selects the semantic arguments.
- ✓ **Full phrasemes** : are non compositional (e.g. *faire face*, *take into account*)

Selection of verbal MWEs

- ✓ **30 verbal MWEs selected**, equally divided in the 2 categories of Support Verb Constructions and Full phrasemes.
 - Construction: **V (Det) N**
 - **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').
 - **Frequent verbal MWEs**

Method

Extracting syntactic and morphological properties from corpora

- ✓ **Corpus** : 32 million words divided in two genres: contemporary literature and newspapers in French, analysed with Connexor dependency parser (Tapanainen & Järvinen, 1997)
- ✓ **Variation properties** extracted with the help of Lexicoscope (Kraif & Diwersy 2014) :
 - **Plural form**: Does the noun occur with the plural form? For example *peur* in *avoir (DET) peur* ('have fear') is always singular.
 - **Determiner variation**: Is the determiner (or the absence of determiner) fixed or variable? 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').
 - **Relative construction**: Is the relative construction possible, for example, *le rôle qu'il a joué* ('the role that he played')?
 - **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 ...').
 - **Modified noun**: Can the noun be modified with an adjective, for example for *avoir un mal fou* ('to have great difficulty')?

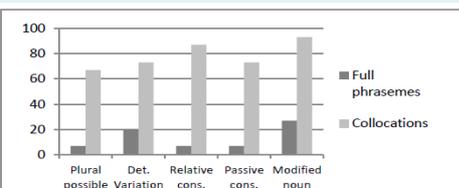


Fig. 1 Variation properties (in %) and type of verbal MWEs

Variation properties

- ✓ Automatic extraction from parsed corpora and manual checking
- ✓ Pervasive ambiguity: e.g. *avoir du mal*, ambiguous with :
 - *avoir mal* 'have pain'
 - *avoir le mal de X* 'be sick'
- ✓ Among properties, variation is higher with full phrasemes with determiner variation and modified noun
- ✓ Among collocations, plural variation is the less frequent

Results

Type of MWE and variability

- ✓ In general, high correlation between variability and type of MWE, even with collocations often used without any determiner (*avoir peur*, *rendre visite*)
- ✓ Some (rare) exceptions:
 - *avoir recours*: « fixed » collocation
 - *avoir du mal*: a « variable » idiom

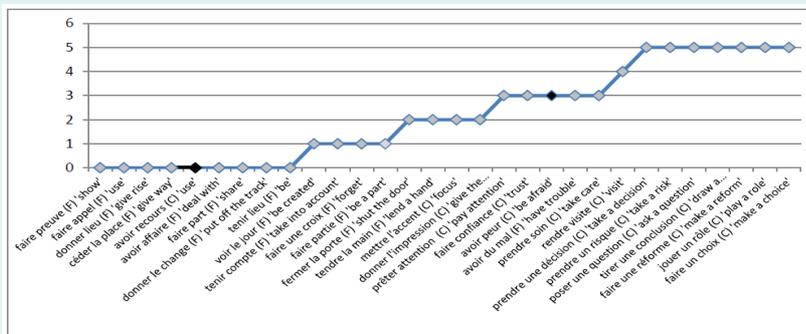


Fig. 2: Classification of verbal MWEs along a continuum according to the number of syntactic and morphological variations (F: full phrasemes. C: collocations)

Conclusion

- ✓ As claimed, but rarely checked against corpus evidence for French, **semantic compositionality and morphological and syntactic variation go hand in hand.**
- ✓ **Collocations are in general by far more variable** – but not so variable for a subset of collocations – than full phrasemes.
- ✓ However, **full phrasemes show very little variation.**
 - need to extend the study to a larger subset of verbal MWEs
- ✓ Variability of MWEs in general should perhaps not be overestimated.
- ✓ **Semantic ambiguity is pervasive among frequent verbal MWEs**, which makes difficult the automatic extraction of variation properties.