

Zakharia Pourtskhvanidze

Institute of Empirical Linguistics, Goethe University Frankfurt (Germany)

VP Hypostasis Translation Problems

ABSTRACT

The translation of idiomatic expressions represents a particular challenge in the context of machine translation. The meaning of these expressions goes beyond their word meaning and corresponds to the meta-language knowledge about the culture and/or current discourse of the language community. A subgroup of the VP hypostasis represent idiomatic content which is used in a certain pattern of use and can only be interpreted in connection with a corresponding discourse. This article discusses an example from English, which was used in the European Parliament and translated into five languages. The analysis shows different translation strategies and illustrates the relevance of the discourse. The method used is the corpus linguistic computation of corresponding discourse.

Keywords: *Multi-Word-Token, Europarl Corpus, Corpus linguistics, Statistical machine Translation, CQP, VP hypostasis.*

When writing the term VP hypostasis (Germ. Satzypostase), we describe the variant conception of reification by using the verbal syntagm as a first part of the noun phrase ([VP+N]NP) or as the self-contained noun ([[VP]N]NP). "Satzypostase" can also be called "Sentence Hypostasis" or "Clausal Hypostasis". However, I would like to avoid the discussion about the uniform definition of the "Sentence" or "Clause" in

this article and therefore focus on VP hypostasis. I just want to point out that this is about the phrases that manifest their own "Sentencehood" (as the property of being a sentence) through a finite verb. Among Indo-European studies, the term "Satzypostase" denotes the verbalization of full-featured finite sentence construction with a neologism status. Most of the examples are reconstructions from Sanskrit. In

Sanskrit, one of the most important function of VP hypostasis is the formation of new words, so the lexemes formed in this way are classified under neologisms.

- (Skr.) *yadaśanīya* - “the edible” < *yad aśanīyam (asti)* “what is edible”

In Georgian, the VP Hypostasis can only be called "neologisms" to a limited degree. These constructions in Georgian are rather ad hoc formations, which have not yet been usualized in linguistic usage, let alone that they do not occur in the lexicons of the standard Georgian language.

- (Geo) *ras-ṭqvis-xalxi* “what-people-say”

In [*ras-ṭqvis-xalxi-s*]_{GEN} *šiši*]_N]_{NP}
“the fear of what what-do-people-say” < *ras ṭqvis xalxi* “what people say”

In the present article VP hypostasis is defined as the reification of a verbal syntagma. I consider the subsequent integration of VP hypostasis into a superordinate structure (typically a noun phrase) as one possible application of VP hypostasis, but not the only one. The main argument for establishing VP

hypostasis as a special class of compounds is the fact that, unlike the rest of the compounds, they are based on syntactically independent sentences. In the foreground is the observation that VP hypostasis does not simply change the POS characteristics of a lexical function, but the syntactic status of the whole phrase.

The cultural backgrounds of the language play a key role in multilingual education. The linguistic expressions that represent the cultural tradition of a language community require the knowledge of corresponding extra-linguistic content in order to ensure the interpretability of a concrete expression. In most cases, these are so-called multi-word expressions, which are characterized by a speech usage pattern. Such patterns only function in connection with certain cultural discourses, which they trigger simultaneously. When multi-word expressions with cultural references (such as idioms or proverbs) are translated, the actual meanings are often lost and the communicative sense of the utterance is incomplete or not understood by the

listener. In the following article I will discuss the special forms of phrase compounds, which have two basic structural prerequisites: (1) phrase structure and (2) obligatory presence of a predicate. The term VP hypostasis refers to the syntactic role such a structure plays in the noun phrase: attributive modification of the head noun within a noun phrase:

[[VP hypostasis_{Att}][N]_{NP}]

For the empirical data I use the Eurparl Corpus. It is a corpus collection that was compiled for the establishment of machine translation in the European

Parliament¹. The corpus comprises of about 30 million words for each of the 11 official languages of the European Union: Danish (da), German (de), Greek (el), English (en), Spanish (es), Finnish (fi), French (fr), Italian (it), Dutch (nl), Portuguese (pt), and Swedish (sv). To extract the VP hypostasis I use a Regular Expression `[word="(.*\-){4,}.*"]` in different modifications. This concerns especially the number "4". It stands for 4 hyphens connecting 5 elements within a token and can vary i.e. become more or less.

Language	Days	Chapters	Speaker Turns	Sentences	Words
Danish (da)	492	4,120	90,017	1,032,764	27,153,424
German (de)	492	4,119	90,135	1,023,115	27,302,541
Greek (el)	398	3,712	66,928	746,834	27,772,533
English (en)	488	4,055	88,908	1,011,476	28,521,967
Spanish (es)	492	4,125	90,305	1,029,155	30,007,569
French (fr)	492	4,125	90,335	1,023,523	32,550,260
Finnish (fi)	442	3,627	81,370	941,890	18,841,346
Italian (it)	492	4,117	90,030	979,543	28,786,724
Dutch (nl)	492	4,122	90,112	1,042,482	28,763,729
Portuguese (pt)	492	4,125	90,329	1,014,128	29,213,348
Swedish (sv)	492	3,627	81,246	947,493	23,535,265

Table 1: Size of the released EUROPARL corpus (version 2; Koehn, 2004:390).

¹ Available online at
<http://www.statmt.org/europarl/>.

The purpose of this article is not to discuss the system of Statistical Machine Translation used in the Europarl Corpus. It is important to note that the CQP syntax allows the use of regular expressions and finds the hyphen construction. The classification of the hyphen constructions as VP hypostasis is the result of the individual analysis.

Let us consider a VP hypostasis from the speech of the Irish MEP Arvil Doyle in the European Parliament in 2002.

„It is very hard to talk about these issues, be objective and keep the focus on climate change if it is suggested that sustainable development is not included. It is a sort of 'hit-me-with-a-

(1) 'hit-me-with-a-child-in-my-arms' argument.

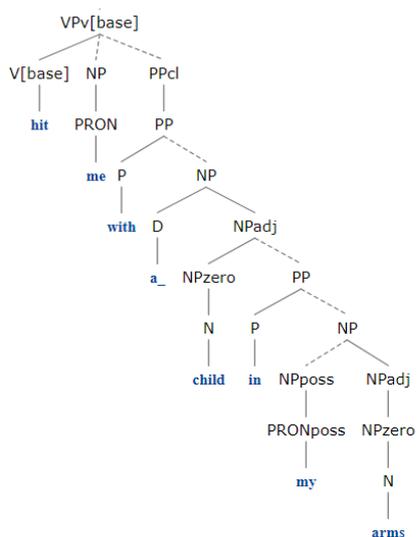
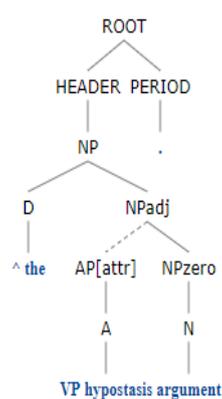


Table 1.1 Out-of-context analysis

child-in-my-arms' argument, i. e. how can you justify what you are saying if you do not include sustainable development? The Americans and others used the serious issue of sustainable development to divert focus from climate change and from the Kyoto Protocol requirements in general. It was a fascinating exercise.“

A structure that we find interesting is in the sentence „It is a sort of 'hit-me-with-a-child-in-my-arms' argument [...]“. First, I will design two different structural analyses of the phrase: (1) independent of context and (2) embedded in context. (Table 1. (1.1/1.2))

C-structure



F-structure

PRED	'argument'
SPEC	DET PRED 'the' 5 6 DET-TYPE def
NTYPE	4 NSYN common
CHECK	3 _LEX-SOURCE morphology
ADJUNCT	1 { 2 PRED 'hit-me-with-a-child-in-my-arms' DEGREE positive, ATYPE attributive }
STMT-TYPE	0 header, PERS 3, NUM sg

Table 1.2 Analysis within the context

Table 1. Structural analysis of the phrase

When comparing the two analyses in (1.1) and (1.2) we find two contrasting structures syntactically: VP (1.1) vs. NP (1.2). Both analyses have their own basis of correctness, with 1.1 playing a subordinate role from the strict empirical perspective. The phrase *'hit-me-with-a-child-in-my-arms'* argument represents a

noun phrase in which the head noun is lexically realized and the member subordinate to the head noun represents a hypostatized verb phrase.

If the intention is to translate this multi-word token into other languages, then the aim is easily achieved based on the construction in (1.1).

Original	<i>hit me with a child in my arms.</i>
German	<i>? schlag mich mit einem Kind in meinen Armen!</i>
Russian	<i>? ударь меня с ребенком на руках!</i>
Georgian	<i>? დამარტყი ჩვილით ხელში!</i>

Although the nuances of the translation may vary, all translations create one and the same picture of marked, narrowly defined situations of life in which the meaning of translated sentences can make sense.

Eorparl Corpus, where the original example was taken. The translation takes into account the general context of the speech of the parliamentarian Arvil Doyle and places the literal meaning in the background.

The other translation examples I have available to us are taken from the

It is a sort of 'hit-me-with-a-child-in-my-arms' argument, i. e. how can you justify what you are saying if you do not includes tainable development.

German	French	Spanish	Italian	Dutch
Es ist sehr schwierig, über diese Fragen zu reden, [...]	C' est un argument assez primaire, comme s' il était impossible	Es un argumento torticero decir que cómo se puede justificar	Domande del tipo: 'Come potete giustificare ciò che andate	Het is een oneigenlijke manier van discussiëren, omdat je je

meint, es sei ein Totschlag-Argument, wenn die nachhaltige Entwicklung nicht dazu gehöre, anders ausgedrückt: Wie kannst du, was du da sagst, rechtfertigen, wenn du nicht die nachhaltige Entwicklung berücksichtigst?	de justifier ces propos si l' on ne parle du développement durable!	lo que se dice si no se incluye el desarrollo sostenible.	dicendo senza alcun riferimento allo sviluppo sostenibile? equivalgono a dire 'Colpiscimi quando tengo un bambino in braccio?	argumenten niet mag ondersteunen met een verwijzing naar duurzame ontwikkeling.
Totschlag-Argument	un argument assez primaire	un argomento torticero	Colpiscimi quando tengo un bambino in braccio?	argumenten
Re-translation into English				
„killer argument" / killer phras / knockout argument	Quite a primary argument	A torrid argument	Hit me when I'm holding a baby?	[...] because you should not support your arguments with a reference to sustainable development.

Table 2. The translations of the phrase "hit-me-with-a-child-in-my-arms ' argument" in 5 other languages

As it is clearly shown, the translation from Italian is the closest to

the English original. The Dutch translation can be considered the furthest.

Neither literal semantic interpretation nor separate structural analysis proves adequate. I regard the phrase with the VP hypostasis as a kind of "catchword" that indicates a particular discourse. The associative connections within this discourse establish semantic clusters that give the phrase 'hit-me-with-a-child-in-my-arms' a communicative meaning. One possibility is the multidimensional scaling of words. Here, the units of scaling are positioned in a multidimensional space in such a way that the distances or sewing correspond to

their actual distinction or similarity. The analysis shown in figure (1) represents 10 clusters. This multi-dimensional scaling of words in the web corpus results in a cohort with the VP hypostasis 'hit-me-with-a-child-in-my-arms' together with the following keywords:

- (1) 'hit-me-with-a-child-in-my-arms' ~ argument
- (2) sustainable ~ development
- (3) climate ~ change

The other clusters concern the European Parliament, its bodies and the political biography of MEP Doyle.

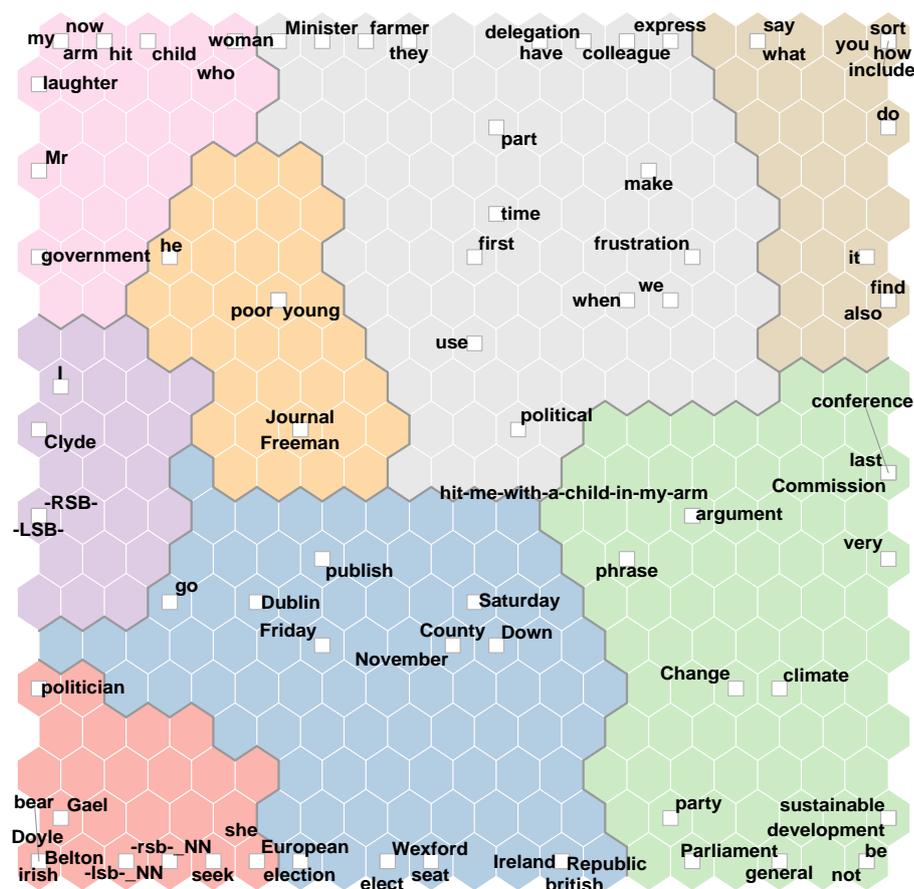


Figure 2. Self-Organizing Map of Words

The analysis of the construction *hit-me-with-a-child-in-my-arms* presupposes the determination of superordinate discourses that give the construction the discourse-specific communicative meaning. In general, the meaning of *hit-me-with-a-child-in-my-arms* is metaphorical speaking. It can be used to intensify the power of an argument in political debate. A direct use of the phrase (Table 1.1) in an appropriate life situation can be interpreted as a provocation and/or as a call to a miserable act, which presents the offender in a bad light.

In the example of the Europarl Corpus, the discussion concerns climate change debate. The metaphorical transfer of semantics creates a certain communicative sense of the phrase, in which the defenseless state (*a-child-in-*

my-arms) is presented as a protective shield against possible bad behaviour (*hit-me*). This meaning seems adequate when the noun "argument" is in: "*hit-me-with-a-child-in-my-arms*" argument. The narrow meaning, related to a concrete debate, points to the importance of an X argument that allows the only counter-argument, and that is, one that makes its user look unworthy. MEP Arvil Doyle criticises this type of debate, which uses the *hit-me-with-a-child-in-my-arms arguments* to argue such a debate.

The VP hypostasis, which transpose metaphorical meaning, are an additional challenge for both translation and foreign language teaching. The calculation of superordinate discourses, which give the VP hypostasis the discourse-specific communicative meaning, can provide a basis for this.

REFERENCES

- Bubenkofer, Noah. 2008. Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse. In: Spitzmüller, Jürgen; Warnke, Ingo H. Methoden der Diskurslinguistik: sprachwissenschaftliche Zugänge zur transtextuellen Ebene. Berlin / New York: de Gruyter, 407-434.
- Bubenkofer, Noah. 2009. - Bubenkofer, Noah, Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin & New York: De Gruyter. S.30. 2009.
- Koehn, 2004 - Koehn, Philipp Europarl 2004 A Parallel Corpus for Statistical Machine Translation. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 388-395.
- Papineni, Kishore, Ward and Zhu, 2002 - Papineni, Kishore Salim Roukos, Todd Ward, and Wei-Jing Zhu, A Method for Automatic Evaluation of Machine Translation. IBM T. J. Watson Research Center. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311-318. 2002.
- Pourtskhvanidze, 2019 - Pourtskhvanidze, Zakharia, Comparative Analyses of Phrasal Compounds. In *Kartvelology, Kartosphere, Kartophony. Language, Culture and Identity*. Ed. Z. Pourtskhvanidze, S. Schmitt, I. Vashakidze. Cuvillier. Göttingen, 31-52. 2019.
- Pourtskhvanidze, 2020 - Pourtskhvanidze Zakharia VP Hypostasis (Satzhypostase) in Georgian. (*Will be published*) In: *Leeds Working Papers in Linguistics and Phonetics*. Leeds, UK. 2020.
- Sabban, 2010 - Sabban Annette Zur Übersetzung von Idiomen im Wörterbuch und im Text: die Rolle von Kontextsensitivität und semantischer Variabilität. *transkom* 3 [2] (2010): 192-208. S. 192. 2010.

Tiedemann, Jorg , 2012 - Tiedemann, Jorg Parallel Data, Tools and Interfaces in OPUS.

Department of Linguistics and Philology. (Online http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf)

Ziegler, 2013 - Ziegler, S. Der Raub des Sonnenrades – der Kutsa-Mythos und eine Interpretation von paruṣá- gáv- in RV 6,56,3. In: Münchener Studien zur Sprachwissenschaft. ISSN 0077-1910. p. 82. 2013.

Corpora:

<https://corpora.linguistik.uni-erlangen.de/demos/CQP/Europarl/frames-cqp.html>