

Semantics of light verbs in lexicographical presentation

Light verbs have always been a difficult lexicographical material disregarding the language described. The main problem consists in polysemy and idiomacity of their meanings. Most of them function both as full lexical units and as semilexical ones. Dictionary entries dedicated to them are always the longest ones and often far from being exhaustive.

We can often observe similarity of basic meanings of such verbs but significant discrepancies between their derived, grammatical meanings even in closely related languages, and the latter is sometimes difficult to grasp both by dictionaries and language learners.

In the article, we discuss methods of possibly accurate description of derived meanings of light verbs that can function as semantic qualifiers in systematic description of complex predicates and ways of extracting them from monolingual dictionaries.

The term “light verb” was introduced by Otto Jespersen in 1965 and referred to verbal part of composite nominal predicates like *take a shower*, *have a smoke* [1, 117]. There is no agreement in the literature at the moment concerning the definition and the semantics of light verbs. These are normally opposed to complex predicates in the Western linguistics and we can come across terms like semi-notional (grammatical) vs full notional (lexical) in the (post-)Soviet linguistic literature (also *nepolnoznamematelnyye* vs *polnoznamematelnyye*) [2].

While analyzing languages like English, Urdu, Hindi, Mandarin Chinese and tracing historical correspondents back to Sanskrit, Miriam Butt comes up

with some interesting cross-linguistic parallels regarding the role and semantics of light verbs. She claims these are a “handful of verbs (somewhere between 5 and 20)” that are “semantically bleached versions of main verbs... Some common examples crosslinguistically are the verbs for ‘come’, ‘go’, ‘take’, ‘give’, ‘hit’, ‘throw’, ‘give’, ‘rise’, ‘fall’ and ‘do/make’. One can think of this set of verbs as *passepartouts*: their lexical semantic specifications are so general that they can be used in multitude of contexts, that is, they “fit” many constellations” [3, 18]. As for their semantics, Butt mentions that “light verbs modulate or structure a given event predication and do so in a manner similar to that of modifiers with respect to semantic notions such as benefaction, suddenness ... also tend to add further information about the aktionsart of the complex predication. In particular, there is often a telic/boundedness or a causation component” [ibid.].

The highlighted by us here meanings differentiated by Butt give a very general idea of what the semantics of light verbs is and do not attach particular meanings to particular verbs. The thesis that “the more subtle semantic notions such as benefaction, force etc., are derived from the collection of entailments usually associated with the lexical semantics of the main verb” [3, 18] is not clear to the end as it doubles the thesis that these meanings are transferred by light verbs. On the other hand, this idea is well in tune with the basic differentiation between compensators and modifiers by [2], where the former ones only play the role of syntactic fitting, while the latter ones complicate the meaning of the whole. Also, Butt’s research helps develop the traditional lexicographic approach regarding conventionality of constructions with light verbs. It is quite intuitive if we look at dictionary explanations of verbs like **take**, e.g. **take** 22. ‘to do an examination or a test, etc., in order to obtain a qualification’ or **take** 26. ‘to make sth by photography; to photograph sth/sb’ [5, 1216] explain the meanings not of **take** alone but constructions it is used in: *take a test* and *take a picture* correspondingly. The meaning rendered by

take alone is too abstract and it is more practical rather to exemplify it than represent in a pure state in a dictionary for human readers.

Presently, the need that arises in connection with automatic NLP, meaning representation and information extraction on different levels (from lexemes and texts) allows us disregarding such a restriction and demands a more formalized way of recording meanings of words. Deeper analysis of light verbs' meanings, their compatibility with main verbs (a cross-linguistic research could be even more informative) might bring more light about the systemic character of "light" structures.

In our approach we follow several assumptions, among them – correspondence of light verbs' meanings to basic ontological categories and semantic compositionality.

The idea of semantic decomposition of predicates is far from new, e.g. following [6] Butt argues for the need of a notion of subevents at the syntax/semantics interface:

'build the house' ($e = e_1 \rightarrow \langle e_2, e_3 \rangle$)

where e_1 = the causing, intentional impulse

e_2 = the process of house-building

e_3 = the state of the house having been built.

The acceptance of subevents by Butt is syntax-driven, due to examples from Hindi like "make+give+house" or "write+take+letter" resulting in perfective meanings of correspondingly 'to have built a house', 'to have written a letter'. Interestingly, those meanings of perfectivity (telicity) are expressed by prefixes in Slavic languages and are parts of lexical units. Such parallels make us think of compositionality on the level of vocabulary and accept for this purpose formal notation of verbal semantics based on predicate calculus with a built-in set of basic predicates proposed by Olgierd Wojtasiewicz [7] ¹.

¹ Wojtasiewicz suggests using twelve basic predicates which can serve as semantic primitives and whose combinations can represent invariant meanings of natural language verbs. $P_0(x)$ and $P_i(x)$ describe states, where

Another our assumption was connected with the language material used for research: if lexicography in many cases explains meanings of constructions with light verbs instead of light verbs themselves (see examples with **take** above), they are more a “building” material for lexicon than lexicon itself; hence, they are often used for explications of full-notional verbs and at the same time comprise part of their meaning, probably the most abstract one. This way, by analyzing definitions of verbs we could discover semantic elements corresponding to semi-notional meanings of light verbs within the semantic structure of complex predicates.

Definitions of Ukrainian verbs from the comprehensive dictionary of the Ukrainian language (*Slovnyk ukrainskoyi movy*, hereafter SUM) in electronic database format, developed by Linguistic-Informational Centre of NASU, were used, cf. also [8].

Definitions of verbs are probably more difficult from the point of view of lexicography than those of nouns, as verbs do not organize such a hierarchy as nouns and only a small group can be explicated in terms of classical definitions with a more general verb and a specification of the way of action. Quite a large group is comprised by LA (lexical aspects) modified verbs, also known as *Aktionsarten*, whose meaning is a composition (conjunction) of the main verb and an LA modifier, see [9]. For the majority of other verbs, a synonymic type of definitions is used with either a one-word synonym or a description in the form of a light verb construction. Thus, on the one hand, light verbs make up *definienda* with the highest quantity of meanings; and on the other hand, they are the most frequent explanatory material.

A short analysis allows differentiating between three basic groups of predicates that can be presented as the following chain of meaning complication:

P_0 is a standard state (as seen by the speaker and preferably accepted as standard by the majority of the users of the language); *Trans* (x,y) expresses the idea of change of state x into state y; by default x precedes y in time; *Ag*(x,y) expresses the idea of agentivity (x does something and y is the result); *V*(x,y,z) is used for the estimation of a situation, where x is the evaluator, y is the situation of evaluation, z is the result (name) of the evaluation; *B*(x,y) is used for the declaration of an opinion; *Exp*(x,y) is used for physical perception declaration.

$\neg P_o(x)$	He is mad
Trans ($P_o(x), \neg P_o(x)$)	He got mad
Ag ($y, \text{Trans } (P_o(x), \neg P_o(x))$)	He drives me mad

Slavic languages allow representing each of the patterns by single verbs, e.g. Ukr. *біліти* (in two meanings: of a state and of a process), *сушити*. Semantic elements are expressed in definitions with the help of different lexemes, to name just a few: *змінюватися*, *переходити*, *ставати*, *переробляти* (change); *діяти*, *робити* (agentivity).

We have looked through the selection of verbs whose definitions include the verb **робити** ‘do, make, work’. In its semi-notional meaning that is used in dictionary explanations it clearly corresponds to predicate $Ag(x,y)$ and expresses the idea of agentivity. The automatically extracted from SUM group comprised 3228 verbs out of over 64700 units in general². The formal criterion for meaning disambiguation in this case was the presence or absence of a direct object and its semantic type. The analysis showed that only in ca. 20 cases the explanatory verb was used in its full lexical meaning as a synonym to the definiendum, e.g. *виконувати*, *здійснювати*, *споруджувати*, *ставити*, *створювати*, *заподіювати*, *чинити* for verbs like **завдавати**, **коритися**, **покорятися**, **поступатися**. Those verbs (including **працювати** ‘work’) can be considered bearers of the predicate $Ag(x,y)$ with no further embedded predicates in the role of arguments. Definitions that include adverbial phrases like *з напруженням*, *за допомогою*, *наспіх* are also few. They can be considered complications of $Ag(x,y)$ and expressed as conjunctions of it with some specified state of the agent $P_i(x)$.

Contrary to examples from Urdu and Hindi presented by Butt [3, 2,4–5] and some English uses like *make sb feel good*, the Ukrainian correspondent **робити** does not take a verbal object, although it takes an adjectival one, with unquestionably predicative nature. Those are the most numerous examples

² For comparison: *бути* (1350), *ставати* (2189), *переходити* (133).

(‘*робити байдужим, близьким, важким, видним*’, etc.) and the main predicate $Ag(x,y)$ of the general pattern embeds predicate of change $Trans(x,y)$.

Most nouns used as direct objects of **робити** in definitions have predicative semantics (*отвір, зміни, дірки, натяжки, обмін, перерву, виправлення, вчинки, перелік, ескіз, малюнок, жест, зауваження, копії, мітки, поділ, зупинку, висновки*, etc.), cf. [10, 12-13]. The common pattern for verbs defined this way also embeds $Trans(x,y)$ with *робити зміни* as its closest correspondent, but in general they are even more composite within $Trans(x,y)$ and need further analysis for finer representation.

Verbs whose definitions include the element **робитися** ‘become’ correspond to the pattern $Trans(P_i(x), P_j(x))$ without specification of the agent. Those comprise a formal subgroup of 380 units. Literally, the postfix **-ся** implies that the object of change itself is the agent. However, no will is evident from its side, so we normally interpret the predicate as unspecified from the point of view of agentivity. The morphologically derivative character of verbs with **-ся** let us consider them semantically derived as well, where unspecified agent is left out.

Wojtasiewicz does not present any rules that guide the composition of basic predicates to make up a meaning of a whole verb. However, some patterns can be seen quite clearly, e.g. agentivity usually embraces transitivity and not vice versa. The assumption of lexical semantic compositionality lets us consider light verbs as higher order predicates with main verbs in the role of their predicative arguments, cf. [11].

Observations of the material let us see some interesting parallels of the meanings of light verbs with basic ontological categories such as change, influence, causativity, belonging, different phases of existence, cf. [12].

Further work connected with light verbs in definitions might benefit from morphological and syntactic annotation of glosses, which would allow a more accurate analysis of meanings according to concrete syntactic patterns.

Literature

1. Jespersen, Otto. A Modern English Grammar on Historical Principles, Part VI, Morphology. London: George Allen and Unwin Ltd., 1965.
2. Золотова, Г. А., Н. К. Онипенко, М. Ю. Сидорова. Коммуникативная грамматика русского языка, Москва, РАН, 1998.
3. Butt, Miriam. The Light Verb Jungle <http://ling.uni-konstanz.de/pages/home/butt/harvard-work.pdf>
4. Словник української мови/ Electronic database file, internal document of the National Ukrainian Academy of Sciences, version of 2004.
5. Oxford Advanced Learner's Dictionary of Current English/ Ed. by A S Hornby, Oxford University Press, 1995.
6. Ramchand, Gillian. First Phase Syntax / Unpublished Ms., Oxford University, 2003. <http://users.ox.ac.uk/~ggram/research/>.
7. Wojtasiewicz, Olgierd A. Sformalizowana semantyczna interpretacja czasowników // Studia semiotyczne, Vol. 6, Wrocław, Ossolineum, 1975.
8. Сухарина, Н. М. Граматична та лексична семантика українського дієслова в лексикографічній системі/ Автореферат дисертації на здобуття наукового ступеня кандидата філологічних наук, Київ, 2003.
9. Kotsyba, N. Slavic Lexical Aspects in the Light of Modern Linguistic Theories// Proceedings of the International conference "New Trends in Mathematics and Informatics" in Sofia, Bulgaria, 6-8 July 2007, forthcoming.
10. Steedman, M. Plans, Affordances, and Combinatory Grammar/ Draft available from <http://www.iccs.informatics.ed.ac.uk/~steedman/papers.html>, 2002.
11. Karolak, S. Podstawowe struktury składniowe języka polskiego, Warszawa, 2002.
12. Jadacki, Jacek Juliusz. Spór o granice istnienia. Warszawa, wyd."Semper", 1998.