



“WORDLETS”

One of Zholkovsky’s Major Contributions to the Notion of Deep-Syntactic Structure

IGOR MEL’ČUK
(Montréal)

About half a century ago, a few fresh ideas from Alexander Zholkovsky, then a rookie linguist and a beginning researcher, revolutionized linguistics.¹ These ideas are simple, clear, and explicit, and they can be summed up as follows:

1. Linguistics—or, more precisely, natural language description—must be **meaning-oriented**. (Natural language is a system designed to express meanings.)
2. Linguistic meaning must be **described formally**, just like other aspects or other units of language. (Without a formal representation of meaning there cannot be a real linguistic science.)

1 See Aleksandr Žolkovskij, Nina Leont’eva, Jurij Martem’janov, “O principal’nom ispol’zovanii smysla pri mašinnom perevode,” *Mašinnyj perevod* 2 (1961): 17–46; Žolkovskij, “Predislovie,” “O pravilax semantičeskogo analiza,” and “Leksika celesoobraznoj dejatel’nosti,” *Mašinnyj perevod i prikladnaja lingvistika* 8 (1964): 3–16, 17–32, 67–103 (in English: Viktor Rozencvejg, ed., *Essays on Lexical Semantics* [Stockholm: Skriptor, 1974], 1: 171–82, 155–69, 197–234).

3. The main research tool of linguists must be **synonymic paraphrasing**, so that synonymy—different linguistic expressions for the same meaning—must be the focus of attention. (What a speaker does is exactly selecting one of several near-synonymous expressions available for a meaning he wants to express.)

In the early 1960s the linguistic landscape was completely overwhelmed by the Chomskyan Transformational-Generativist tidal wave. Against such a background, Žolkovsky’s postulates represented a breakthrough—they led to two extremely important conclusions:

- The semantic description of words becomes the central task of linguistics, so that the lexicon, and not the grammar, is the primary target of linguistic studies. (The grammar is, of course, an absolutely essential component of a language and deserves the full attention of linguists, but it is secondary with respect to the lexicon, since the grammar represents a set of useful generalizations over properties of individual words.)
- Linguistics should be concerned not with the generation of texts from a mysterious starting symbol, but with the translation of given meanings into the corresponding texts, and vice versa.

I was knocked off my feet by the depth of Žolkovsky’s daring ideas; I offered him collaboration, and my offer was gracefully accepted. We worked together very closely for the next ten years.² Soon Jurij Apresjan joined us, and thus the Meaning-Text linguistic approach was born.

2 Aleksandr Žolkovskij and Igor’ Melčuk, “O vozmožnom metode i instrumentax semantičeskogo sinteza,” *Naučno-texničeskaja informacija* (1965), No. 5: 23–28; Žolkovskij and Melčuk, “O sisteme semantičeskogo sinteza: I. Stroenie slovarja,” *Naučno-texničeskaja informacija* (1966), No. 11: 48–55; Žolkovskij and Melčuk, “O semantičeskom sinteze,” *Problemy kibernetiki* 19 (1967): 177–238; Žolkovskij and Melčuk [A. Jolkovsky, I. Melčuk], “Essai d’une théorie sémantique applicable au traitement de langage (modèle ‘Sens-Texte’),” *Proceedings of COLING’67, C67-1028* (1967), accessed April 10, 2017, <http://aclweb.org/anthology/C/C67/>; Melčuk and Žolkovskij [I. A. Melčuk and A. K. Žolkovskij], “Towards a Functioning ‘Meaning-Text’ Model of Language,” *Linguistics* 57 (1970): 10–47; Melčuk and Žolkovskij, *Tolkovo-kombinatornyj slovar’ sovremennogo russkogo jazyka/Explanatory Combinatorial Dictionary of Modern Russian*, Wiener Slawistischer Almanach, Sonderband 14 (Wien: Wiener Slawistischer Almanach, 1984).

It is not my intention here to characterize the value of Zholkovsky's general input to Meaning-Text linguistics. I want rather to discuss a particular point, which, I am afraid, has not been duly acknowledged: I mean Zholkovsky's "wordlets," a concept that is vital to the elaboration of the deep-syntactic structure of sentences.

One of the specific features of the Meaning-Text approach is the distinction between two levels of syntactic structure: deep-syntactic structure [DSyntS], geared to the meaning of the sentence to be produced, and surface-syntactic structure [SSyntS], targeting the sentence form—that is, word order, phrasing, and syntactic inflections.³

One of the main *raison d'être* of the deep-syntactic structure of sentences is its general—that is, linguistically universal—character. DSyntS was conceived and implemented with an eye toward a universal representation of specific syntactic constructions of particular languages. This means that all highly variegated expressive means used by actual languages to mark syntactic relations—namely, structural words, word order, prosody, and inflectional morphology (agreement and government)—should be ignored in the DSyntS and replaced with a generalized formalism. Here is an illustration.

Consider several surface-syntactic constructions from different languages:

Table 1 Different Syntactic Constructions Expressing the Same Syntactic Roles (Continued)

• Direct Object:	Eng.	[<i>John</i>] <i>broke</i> —[<i>a/the</i>] → <i>cup</i>
	Rus.	[<i>Ivan</i>] <i>razbil</i> → <i>čášk+u_{ACC}</i>
	Geor.	[<i>Ivanem</i>] <i>gat'exa</i> → <i>č'ika+Ø_{NOM}</i>
	Hebrew	[<i>Yohanan</i>] <i>šabar</i> → <i>et ha_{DEF}+kos</i>
	Turk.	[<i>Yahya</i>] <i>fincan+i_{ACC}</i> ← <i>kırdı</i> (Turk. <i>c</i> = [ǰ])
• Oblique Object:	Eng.	[<i>an</i>] <i>approach_N</i> → <i>to the city</i>
	Fr.	<i>s'approcher</i> → <i>de ['of'] la ville</i>

3 See Igor Melčuk, *Semantics: From Meaning to Text*, 3 vols. (Amsterdam/Philadelphia: John Benjamins, 2012–15), 2: 18–93.

	Rus.	<i>interesovat'sja</i> → <i>ximi+ej</i> _{INSTR} 'be interested in chemistry'
		<i>interes</i> → <i>k</i> ['to'] <i>ximi+i</i> _{DAT} lit. 'interest to chemistry'
• Adnominal Complement:	Eng.	<i>John's</i> ← <i>murder</i>
	Fr.	<i>le meurtre</i> → <i>de</i> ['of'] <i>Jean</i>
	Rus.	<i>ubijstvo</i> → <i>Ivan+a</i> _{GEN}
	Hebr.	<i>sifrei</i> _{CONSTR} → <i>limud</i> lit. 'books [of] larning' = 'textbooks' [in this construction, the noun <i>sfarim</i> 'books' takes the form <i>sifrei</i>]
	Bret.	<i>muntr</i> → <i>Yann</i> lit. 'murder Yann'
	Turk.	<i>Yahya+nin</i> _{GEN} ← <i>cinaet+i</i> _{3,SG} lit. 'of-John murder-his'
• Adjectival Modifier:	Eng.	<i>Armenian</i> ← <i>massacre</i>
	Rus.	<i>armjansk+aja</i> _{FEM, SG, NOM} ← <i>reznja</i>
	Fr.	<i>massacre</i> → <i>arménien</i> +∅ _{MASC} +∅ _{SG}
• Prepositional Complement:	Eng.	<i>for</i> → <i>John</i>
	Rus.	<i>dlja</i> → <i>Ivan+a</i> _{GEN}

This picture is fairly involved: different structural words, different linear positions, different government and agreement markers are used to express identical or very similar syntactic relations. It is further complicated by the fact that there is no one-to-one correspondence between syntactic constructions in different languages; for instance:

—What is a Direct Object in one language can be an Indirect/Oblique Object in another:

Rus. *blagodarit'* -**dir-obj** → *kogo*_{ACC} 'thank somebody' ~

Ger. *danken*-**indir-obj** → *jemandem*_{DAT} lit. 'thank to somebody'

Rus. *ždat'* –**dir-obj** → *kogo*_{ACC} lit. ‘wait somebody’ ~
 Eng. *wait* –**obl-obj** → *for somebody*

—What is an adjectival modifier in one language can be rendered by an adnominal complement in another:

Eng. *Armenian* ← **adj-modif** –*genocide* ~

Rus. *genocid* –**obj-adnom** → *armjan* + \emptyset _{PL, GEN}; and so on.

It is exactly in situations of this type that the DSyntS enters the fray. All of the above SSynt-constructions in all languages can be safely represented by just one DSynt-construction: namely, deep-syntactic actant **II**, that is, L_1 – **II** → L_2 .⁴ To do so, we have to supply every lexical unit L of every language with what is known as Government Pattern [GP], which is part of L's lexical entry. In the headword L's GP all expressive means of the surface realization of L's actants are listed, along with the conditions for their choice, if there are alternatives. Thus:

BREAK _(V) :	II ⇔ – dir-obj → N
APPROACH _(V) :	II ⇔ – obl-obj → to N
MURDER _(N) :	II ⇔ – possess → N's
Rus. K 'to' :	II ⇔ – prepositional → N _{DAT}
Rus. VERNYJ 'faithful':	II ⇔ – obj-obl → N _{DAT}
Rus. UBIJSTVO 'murder' _(N) :	II ⇔ – obj-adnom → N _{GEN} ⇔ – modif → A _(poss) (N)

The same treatment is reserved for all possible actant-expressing constructions: they can be systematically described by seven deep-syntactic actantial relations plus GPs in the lexical entries for their governors. Then all modifiers (restrictive and qualifying), all extra-structural elements (like addresses and parentheticals), and all quasi-coordinated and coordinated strings can also be generalized so as to be subsumed under a few deep-syntactic relations. To sum up: most of the known syntactic constructions of natural languages can be safely reduced to thirteen deep-syntactic relations [DSyntRels], shown below.

4 On semantic and syntactic actants of lexical units, see Igor Melčuk, “Actants in Semantics and Syntax,” *Linguistics* 42 (2004), no. 1: 1–66; no. 2: 247–91; Melčuk, *Semantics*, 3: 4–107.

Table 2 Deep-Syntactic Relations

Subordinate DSyntRels										Coordinate DSyntRels		
Strong subordinate DSyntRels							Weak subordinate DSyntRels					
Actantial DSyntRels						Attributive DSyntRels						
I	II	III	IV	V	VI	II _{dir-sp}	ATTR _{restr}	ATTR _{descr}	ADDRESS	APPEND	PSEUDO-COORD	COORD
1	2	3	4	5	6	7	8	9	10	11	12	13

The DSyntRels are presented here (Table 2) in the order of decreasing “subordinateness”: from the strongest hypotaxis to the loosest parataxis. DSyntRel 7, **II**_{dir-sp}, stands for Direct Speech functioning as DSynt-actant **II** of a speech verb; DSyntRels 8 and 9 cover, respectively, restrictive and descriptive(= qualifying) modifiers; **ADDRESS** links a direct address to the head of the clause, while **APPEND** represents appenditive constructions, such as sentence adverbials, prolepses, and interjections; and the **PSEUDO-COORD** DSyntRel describes constructions that express elaboration (*John stays in Canada, in Quebec, in Montreal, at his parents’*) and the like.

So far, so good. However, as a general rule, languages have particular syntactic constructions that express meanings of the lexical type—that is, meanings that are normally expressed by lexical units. The stock example is the Russian “approximate” numeral construction:

1.
 - a. *desjat’ kilometrov* ‘10 kilometers’: NUM + N
versus
 - b. *kilometrov desjat’* ‘maybe 10 kilometers’: N + NUM

The construction in (1b) expresses the uncertainty of the Speaker as to the number and is best translated by adding “maybe” to the numeral phrase. This meaning is expressed by purely syntactic means—word order: the noun, which otherwise follows the numeral, is positioned before it. This construction is represented easily in the surface-syntactic structure by a special surface-syntactic relation, N–**approximate-quantitative** → NUM (which contrasts with the **quantitative** surface-syntactic relation, used for the NUM + N construction). However, this technique is not available at the deep-syntactic level. If each meaningful syntactic

construction of this type is represented by a particular DSyntRel, the universal character of the DSyntS is lost. And such constructions are quite numerous across languages and highly unpredictable! So, what should then be the technique for deep-syntactic representation of such constructions?

Today, we would approach the problem as follows: (1) the Deep-SyntS has two, and only two, available expressive means—deep-lexical units and deep-syntactic relations;⁵ and (2) we cannot introduce as many new specific DSyntRels as we would need. Therefore, the “deviant” surface-syntactic constructions should be represented by artificial lexical units: fictitious lexemes. However, this is how we reason today! Almost fifty years ago this idea had never occurred to anybody—that is, anybody except Zholkovsky. In 1971, he spoke quite explicitly of “some meanings that, within the proposed deep-syntactic framework, appear as lexical units—words, or, better, ‘wordlets’ of our metalanguage, while at the surface-syntactic level they are implemented by morphological markers or syntactic constructions.”⁶ And he illustrated his solution with several examples, one of which I will reproduce here⁷ (Zholkovsky’s “wordlets” are shown by double angle quotes: «. . .»).

The Somali «MORE» Construction

Somali does not have a comparative degree of the adjective—it does not even have adjectives: the qualificative verb does the adjective’s job, very much as in Mandarin or Vietnamese. The meaning ‘X is more.P than.Y’ is expressed by the syntactic construction “X and Y, Y more is.P”:

2. *Nin + ka iyo naag+ta nin + ka baa ka wanaagsan*
 man the and woman the man the SentPart CompPrev is. good
 lit. ‘The.man and the.woman, the.man is more good’ =
 ‘The man is better than the woman.’

[BAA is a sentence particle, an element necessary in any full-fledged clause, which, without it, would be a nominal phrase; this particle fulfills a communicative

5 Inflectional characteristics do not exist in all languages, so we cannot use them in a general discussion of deep-syntactic structures.

6 Aleksandr Žolkovskij, *Syntaksis somali* (Moscow: Nauka, 1971), 9 (translation is mine—IM).

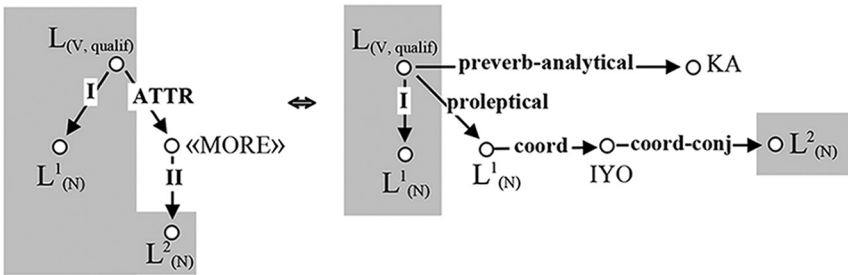
7 *Ibid.*, 145.

role: it marks the Rheme of the clause. KA is a comparative preverb: it marks the use of the verb in a comparative construction.]

Using modern notations, it is possible to represent this Somali syntactic construction as follows, «MORE» being the corresponding “wordlet”—that is, a fictitious lexeme.

Deep-syntactic structure

Surface-syntactic structure



This is a deep-syntactic rule (DSyntS \Leftrightarrow SSyntS). Shadowing shows the context of the rule: elements that are not affected by the rule itself, but whose presence in the DSyntS is necessary for the rule to apply. The **proleptical** SSynt-relation subordinates the prolepsis, a nominal phrase loosely linked to the rest of the clause.

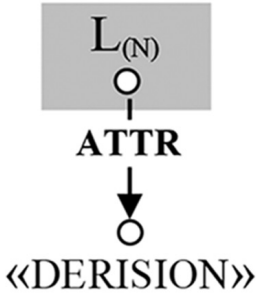
The English «DERISION» Construction

Now let me illustrate the use of fictitious lexemes in English. Meaningful syntactic constructions are not widespread in this language, but there are some, and here is an example.

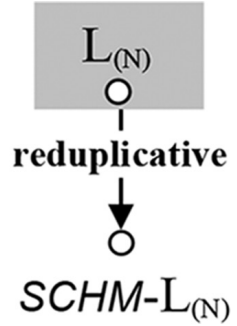
- 3. *Politics, schmoltitics!* | *Theory, schmeory.* | *Books, schmooks.* | *Baby, schmaby.*

The so-called *schm*-reduplication of a noun L expresses the Speaker’s derision and skepticism about L: ‘SCHM-L \approx ‘I dismiss L as being ludicrous and worthless.’ In the DSyntS of a sentence with the lexeme *schm*-L, this meaning can be represented by the fictitious lexeme «DERISION»:

Deep-syntactic structure



Surface-syntactic structure



SCHM- is the name of the corresponding derivational means (adding to L the prefix /šm/ and deleting the initial prevocalic cluster, if any).

The Russian «MAYBE» and «ORDINARY» Constructions

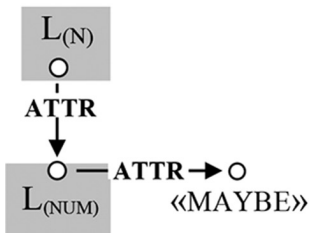
Two other examples of fictitious lexemes come from Russian.

4. Russian [= (1), see the discussion above]

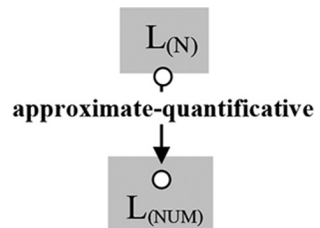
- a. *desjat' kilometrov* '10 kilometers'
- versus
- b. *kilometrov desjat'* 'maybe 10 kilometers'

The Russian approximate construction has to be represented at the DSynt-level by the fictitious lexeme «MAYBE».

Deep-syntactic structure



Surface-syntactic structure

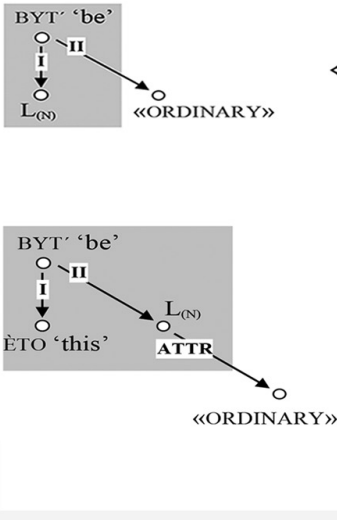


5. Russian

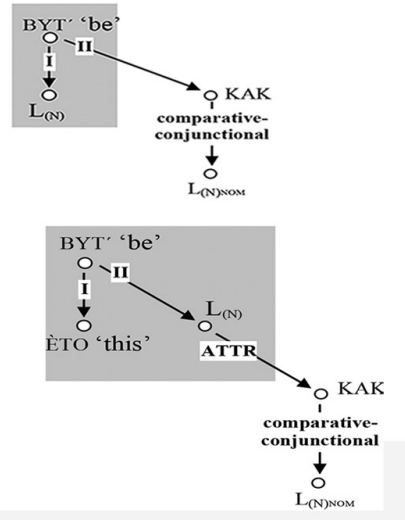
- a. *Kniga kak kniga* lit. ‘Book as book’ = ‘Quite an ordinary book’ or ‘The book is quite ordinary.’
- b. *Kniga byla kak kniga* lit. ‘Book was as book’ = ‘The book was quite ordinary.’
- c. *Èto (byla) kniga kak kniga* lit. ‘This is/was quite an ordinary book.’

We see here a syntactic phraseme N KAK N ‘quite an ordinary N’, which can be represented in the DSyntS by the fictitious lexeme «ORDINARY». Since this fictitious lexeme can appear in two different contexts—as a complete independent clause or as an attribute of the copula BYT’ ‘be’, we need two DSynt-rules to describe the corresponding constructions, the first one for the cases like (5b) and the second one for those like (5c); (5a), which either has no finite verb or contains a zero form of BYT’, is syntactically ambiguous and corresponds to either of these rules.

Deep-syntactic structure



Surface-syntactic structure



Here is a list of fictitious lexemes gleaned from various languages.⁸ They feature quite abstract meanings that one could qualify as “grammatical”; that

8 Melčuk, *Semantics*, 2: 37–42.

is why Zholkovsky spoke of them as “belonging to a border zone between the lexicon and the grammar.”⁹

«AFFECT»	«CONDITION»	«IF _{IRR} »	«ORDINARY»
«AFTER»	«DERISION»	«INCLUDE»	«SAY»
«ALTHOUGH»	«FOR» (<i>buy her a</i>	«INSTRUMENT»	«SHOULD»
«BE»	<i>dress</i>)	«MATERIAL»	«START_BRUS-
«BE_ABLE»	«FROM» (<i>one</i>	«MAYBE»	QUELY»
«BE_FROM»	<i>from these</i>)	«MORE»	«TITLE» (<i>Professor</i>
«BECOME»	«GOAL»	«MOVE _{DIR} »	<i>Drouin</i>)
«BELONG»	«HAVE»	«NAME»	«WHILE»
«CAUSE _N »	«HAVE_TO»	«NUMBER [of]»	«WILL_BE»
	«IF»		«WITH»

The introduction of fictitious lexemes (= “wordlets”) is one of the most important contributions of Zholkovsky’s to the formalism of deep-syntactic structures in the Meaning-Text approach. But it is by no means the only one in this domain. No less important and fruitful has been his proposal for distinguishing deep and surface parts of speech. That, however, is another kettle of fish. I would like to save this topic for Zholkovsky’s next anniversary.¹⁰

9 Žolkovskij, *Syntaksis somali*, 10 (translation is mine—IM).

10 Acknowledgments: This text was read by David Beck, Joshua Holden, Lidija Iordanskaja, and Jasmina Milićević. My most heartfelt thanks to them for the remarks and suggestions that helped me improve the presentation.