

NATURAL LANGUAGE PROCESSING AS A SOURCE OF LINGUISTIC KNOWLEDGE

L.L. Iomdin

Institute of Information Transmission Problems

Russian Academy of Sciences, Moscow

iomdin@cl.iitp.ru

Keywords: natural language processing, machine translation, lexical and grammatical ambiguity resolution, theoretical syntax, theoretical and applied lexicography

Abstract¹

The paper discusses a number of specific problems of natural text parsing that emerge during the operation of a highly developed rule-based machine translation system, ETAP-3. Emphasis is laid on two classes of problems: 1) adequacy of linguistic description of the working languages of the MT system and 2) means of resolving lexical and syntactic ambiguity of the source text. It is claimed that no parser, however sophisticated or advanced, can be made entirely free of lacunae and gaps. The reason is that many of the linguistic facts, including those critical for parser operation, have never come into view of researchers simply because they have not had at their disposal mass material of unexpected or incorrect parsing. It is exactly such material that is amply provided by a highly developed NLP system. If handled properly, this feedback helps the researcher to find the gaps of scientific descriptions and eliminate them. Consequently, linguistic experimentation with NLP systems becomes a rightful and very promising scientific method. In a way, linguistic applications start to stimulate theoretical research, thus inverting the situation that has existed ever since NLP came to life.

Introductory remarks

The paper deals with a number of instructive cases that opened up in the course of experimental operation of the Russian-to-English automatic translation module of a high-level multifunctional NLP system, ETAP-3, developed by a Moscow research team [1-3]. The system, largely based on the Meaning \Leftrightarrow Text theory by Igor Melčuk [4], makes use of dependency syntax: the syntactic structure of

any sentence is represented as a dependency tree whose nodes correspond to all words of the sentence and whose arcs are labeled with names of one of several dozens syntactic relations. The method of syntactic representation will be essential in the following account.

All cases were evolving in a fairly similar way: the MT module was offered Russian sentences for translation, for which it yielded unsatisfactory English equivalents. Normally, the sentences came from current Russian Internet news sites. The system's operation was subsequently subjected to severe scrutiny, which enabled the experimenters to locate the errors and correct them if at all possible. In the reverse case, when any of the errors proved incorrigible, the experimenters could make a step towards establishing natural limits to machine translation performance. Below, I will consider some representative situations in which unsatisfactory performance of the NLP system has led to important theoretical findings.

A Lesson in Theoretical Syntax

It can readily be assumed that in a language so well investigated as Russian core syntactic structures should bear no surprise gaps that can tell on parsers. Relative clauses introduced by the word *kotoryj* 'which/that/who' definitely belong to the core syntax of Russian and have appeared to be fully described in the grammar. As these relative clauses are somewhat cumbersome², their grammatical description is rather detailed. Respectively, the clauses seemed to have been adequately represented in the syntactic module of the NLP system. This is why a serious error in the

¹ The work was in part supported by the Russian Foundation of Fundamental Research, grant No 02-06-80106.

² The cumbersome character of relative clauses is manifested in the fact that the word *kotoryj* may be either a direct dependent of the relative clause head (*chelovek, kotoryj smeetsja* 'the man who laughs') or else occupy a position in the dependency tree which is quite far from the head (*chelovek, s zhenoj odnogo is starshix brat'ev kotorogo ja xorosho znakom* 'the man whose relative - the wife of one of his elder brothers - I know well' (lit. 'the man with the wife of one of elder brothers of whom I am well acquainted).

translation of a rather simple sentence (1) came as a complete surprise to the experimenter:

(1) *Vlasti Afganistana izdali rasporyazhenie, soglasno kotoromu vooruzhennym licam predpisano pokinut' Kabul*

(lit. Authorities of Afghanistan have issued an order, according to which it is prescribed to armed persons to abandon Kabul)

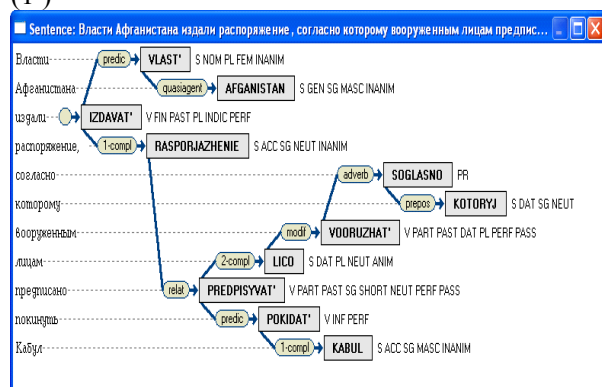
‘The Afghan authorities have issued an order which prescribes that armed persons must abandon Kabul’

The translation of this sentence was not only ungrammatical but utterly enigmatic:

(1a) *Authorities of Afghanistan have issued an order it is prescribed to according to which armed persons to abandon Kabul*

The enigma could soon be accounted for: the parser produced for (1) the following dependency syntactic structure (SyntS):

(1)³



It can be easily seen that translation (1a) reproduces structure (1') with practically no changes. What is wrong with structure (1'), though? As a matter of fact, we will never find the answer in classical Russian grammar (or in the ETAP-3 syntax that conscientiously represents this grammar). No restrictions that could be made responsible for the ungrammaticality of (1') can be found there. Our first conjecture was that the “arboreal” distance between the head of the relative clause (the short passive participle *predpisano* ‘prescribed’) and the word *kotoryj* might be too big and that perhaps a constraint must be imposed on the number of nodes that can be located between the head and *kotoryj*. However, the two words are only four “arboreal” steps apart while SyntS of the longer relative clause given in Footnote 2 above contains as many as six arboreal steps between the head and the relative pronoun and sounds entirely natural. After thorough research we

³ SyntS are reproduced here and below as screenshots of the internal graphic editor of the ETAP-3 system. Cyrillic letters have been however replaced by Latin characters.

were able to suggest another hypothesis to explain the situation. We presumed that the character of the elements that separate the head and the relative pronoun in SyntS (1') is more important than their number. The string of words connecting the head of the relative clause in (1') and the relative pronoun only contains three words: the noun *licam* ‘(to) persons’, the preposition *soglasno* ‘according to’ and the participle *vooruzhennym* ‘armed’. Nouns and prepositions are too typical elements of relative clauses to suspect them of being the cause for ungrammaticality.

I must admit that I did not see such a cause in the participle, either. As a matter of fact, there are no semantic barriers for interpreting sentence (1) as (1'): if we replace the relative clause as an independent sentence, we will come up with a sufficiently sensible text: *Vlasti Afganistana izdali rasporyazhenie. Vooruzhennym soglasno etomu rasporyazheniju licam predpisano pokinut' Kabul*. ‘The Afghan authorities have issued an order. Persons that were armed in accordance with this order are required to abandon Kabul.’

However, a very comprehensive search of modern Russian corpora has shown that indeed no participles may appear in the syntactic string of words that separate the head node of the relative clause from the pronoun *kotoryj*⁴. Even in shortest contexts, where the participle is directly dependent on the head and directly dominates the relative pronoun, the constraint is absolute and valid for both passive and active participles, cf.

(2a) **Malysh, pozvavshix kotorogo djadja poshel kupat'sja, po-prezhnemu vozilsja v peske* (‘the kid whose uncle who called him went bathing, was still playing in the sand’).

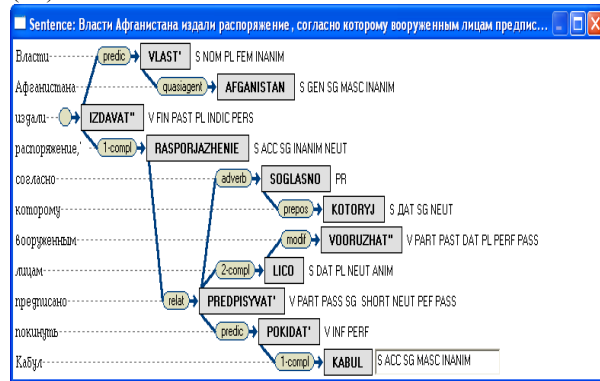
(2b) **Eto uchenyj, sdellanoe kotorym otkrytie proizvelo perevorot v nauke*. ‘This is a scientist a discovery made by whom made a revolution in science’.

I have to state that the constraint that, to the best of my knowledge, has never before been mentioned in the literature, is of purely syntactic nature and appears unmotivated from the semantic viewpoint: the meaning of ungrammatical sentences (2) is quite

⁴ Note that the string of words is understood in the “arboreal”, rather than the linear, sense. The participle may easily appear in the linear string if it is not involved in the chain of syntactic links that connect the head and the relative pronoun. An example is a relative clause which is a slight modification of the one given in Footnote 2: *chelovek, s zhenoj odnogo iz uexavshix [= past participle] za granicu brat'ev kotorogo ja xorosho znakom* ‘the man whose relative - the wife of one of his elder brothers who went abroad - I know well’

transparent. It goes without saying that once the result has been obtained it can be easily introduced in the parsing rules of the NLP system. The salutary effect of the feedback will be seen at once: the updated parser will come up with a correct SyntS (1'') for sentence (1).

(1'')



This structure will be sent to the transfer phase, which will produce a passable translation like

(1b) *Authorities of Afghanistan have issued the order according to which it is prescribed to armed persons to abandon Kabul.*

It must be added that the constraints to be imposed on relative clauses must include other items beside the ban on participles. To mention but a few, elements that separate the head of the relative clause from *kotoryj* cannot include any finite forms of verbs, or any conjunctions.

The latter fact is especially instructive since it clearly demonstrates the syntactic nature of the constraint. Even though prepositions and conjunctions can be largely synonymous, sentence

(3a) *Eto byl master, podobno kotoromu ne rabotal nikto I nikogda* (lit. 'This was an expert like who nobody ever worked.')

where *podobno* 'like' is a preposition sounds impeccably, whereas sentence

(3b) **Eto byl master, kak <slavno> kotoryj ne rabotal nikto I nikogda* (lit. 'This was an expert like who nobody ever worked.')

where *kak* and *slavno* are comparative conjunctions is definitely ungrammatical.

I should like to emphasize once again the unique role that the NLP system played in the formulation of this challenging syntactic task. In fact, the applied system is becoming to act as a fully fledged research instrument of theoretical linguistics that stimulates its progress. In this case, the machine translation system proved to be a source of invaluable negative

linguistic material: I believe that it is only an artificial system that can produce such material in a spontaneous manner: no native speaker of a language or even a sophisticated linguist can do that.

A Lesson in Syntax and Lexicography

In what follows I will demonstrate an entirely different type of lesson drawn from NLP operation.

Last spring, the ETAP-3 system was tested on the ITAR-TASS newswire that happened to publish the following "news" message:

(4) *Glavnaja cel' Velikogo posta, kotoryj načalsja dlja pravoslavnyx – npravstvenno vozvysit' čeloveka*
'The main purpose of the Lent that has begun for the Orthodox is to elevate a person morally'.

The English translation produced for (4) was, surprisingly, inadequate and ungrammatical:

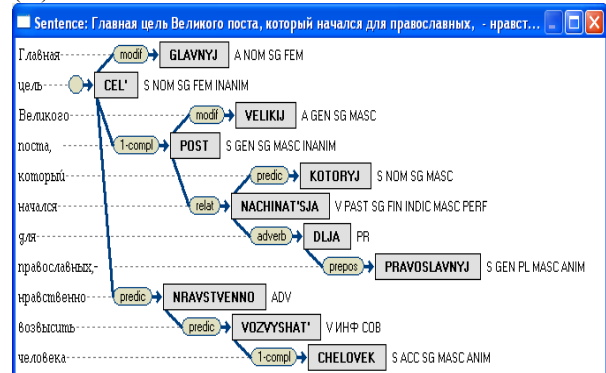
(4a) **The main purpose of the Lent that has begun for the Orthodox – it is moral to elevate a person.*

One of the sources of the translation error is quite obvious and could be detected at once. The parser treated the syntactically ambiguous right-hand part of the copulative sentence (4), namely

(5) *npravstvenno vozvysit' čeloveka*

as a full sentence with an infinitive subject and a zero copula: 'to elevate a person is moral', rather than an infinitive clause 'to morally elevate a person'. Indeed, the syntactic structure (SyntS) produced by the parser for (4) looked as follows⁵:

(4')



Interestingly, the syntactic ambiguity of text fragment (5), namely, the possibility to parse it as an infinitive clause or a full predicative sentence, is

⁵ As a matter of fact, SyntS (1'') even contains two zero copulas: *Glavnaja cel' Velikogo posta* \emptyset_{be} ... *npravstvenno* \emptyset_{be} *vozvysit' čeloveka* (lit. The main purpose of the Lent IS: to raise a person IS moral). The fact that zero copulas are not presented in (4') is of purely technical nature.

accounted for by an interplay of a number of lexical and syntactic factors and may be viewed as a lucky coincidence. In the absence of the ambiguous word *npravstvenno* (which can belong to the paradigm of an adverb ‘morally’ or an adjective ‘moral’, where it is the singular neutral short form), the right-hand part of the sentence would lose its ambiguity, it could not be interpreted sentimentally, and we would not be able to draw the lesson at issue.

So, the first source of the erroneous parsing is the ambiguity of the second part (5) of the copulative sentence (4). We now have to explain why this ambiguity does not extrapolate to the whole sentence, i.e. why (5) does not allow for a sentential interpretation and can only be treated as an infinitive clause. Is it not the case that copulative sentences, either in Russian or in English, cannot contain full predicative clauses as their integral part? Of course they can. Moreover, we can easily construct an immaculate copulative sentence of this kind if we just replace the noun *cel’* ‘purpose’ with another noun, *ideja* ‘idea’:

(6) *Glavnaja ideja Velikogo posta, kotoryj načalsja dlja pravoslavnyx – npravstvenno vozvysit’ čeloveka.*

In contrast to unequivocal sentence (4), (6) is syntactically ambiguous and can be interpreted as having a predicative clause or an infinitive clause in the right-hand part. Since, in contradistinction to Russian, English does not have an ambiguity similar to that exemplified by (5) (because it does not use a zero copula.), the two readings of (6) will be translated differently:

(6a) *The main idea of the Lent that has begun for the Orthodox is that it is moral to elevate a person;*

(6b) *The main idea of the Lent that has begun for the Orthodox is to elevate a person morally.*

Sentence (6a) contains a predicative clause and (6b) an infinitive clause. Unlike (4a), both translations (6a) and (6b) are acceptable and both are grammatically correct.

It can thus be seen that SyntS like (4’), namely, SyntS of copulative sentences whose left-hand part is a nominal phrase and whose right-hand part (following the copula) is a predicative clause are not configurationally wrong.⁶ Their acceptability depends on lexical instantiation of the head noun of the left-

⁶ In my opinion, the NP is most likely the p r e d i c a t e of such a copulative sentence rather than its subject, while the predicative clause part is its subject. Detailed validation of this opinion lies beyond the scope of this work; suffice it to say here that the NP part of the sentence can be in the instrumental case, as in example (8) below, which is extremely untypical for a subject.

hand NP, and a thorough investigation will enable us to determine which lexical instantiation is admissible in (4’) and which is not.

In particular, if the NP part of the copulative sentence is formed by predicate nouns like *cel’* ‘purpose, aim, goal, target’, *naznačenie* ‘destination’, *prednaznačenie*, *missija* ‘mission’, *namerenie* ‘intention’, *stremlenie* ‘ambition, aspiration’, *prizvanie* ‘vocation’, *želanie* ‘wish’, *strast’* ‘passion’, *zadača* ‘task’, *problema* ‘problem’, the SyntS of type (4’) remains faulty.

In order to be correct, the second part of the copulative sentence with such an NP must be an infinitive clause, as in (7) or (9), a nominal phrase (as a rule, it must be formed by another predicate noun, as in (8)), or a subordinate clause with the conjunction *čtoby* ‘so that’, as in (10), but never a predicative clause — even if such a clause contains a modal verb which might compensate for the infinitive or ČTOBY-clause. Cf.

(7) *Naša cel’ - ustanovit’ istinu* ‘Our purpose is to establish the truth’;

(8) *Naša cel’ - ustanovlenie istiny* ‘Our purpose is the establishment of the truth’

but not **Naša cel’ – istina dolžna byt’ ustanovlena* (lit. Our purpose is: the truth must be established’ *ustanovlenie istiny* <*ustanovlenie istiny*>),

(9) *Osnovnoju moej cel’ju bylo vse-taki otvleč’ Axillesa ot razgovora o Xarone* ‘Still, my basic purpose was to distract Achilles from the talk about Charon (Arkady and Boris Strugatsky),

but not **Osnovnoju moej cel’ju bylo: Axilles dolžen otvleč’sja ot razgovora o Xarone* ‘My basic purpose was: Achilles must be distracted from the talk about Charon;

(10) *Glavnaja cel’ - čtob sosedi ne videli* ‘the main purpose is that the neighbours should not notice’ (Vasily Aksenov),

but not **Glavnaja cel’ - sosedi ne uvidjat* ‘the main purpose is: the neighbours will not notice’.

On the other hand, the copulative construction with the sentential second part will be perfect if its first NP part is formed by a sufficiently broad class of nouns, such as *ideja* ‘idea’, *mysl’* ‘thought’, *smysl* ‘meaning’, *pašos* ‘pathos, spirit’, *posylka* ‘premise’, *tezis* ‘thesis’, *položenie* ‘thesis, proposition’, *princip* ‘principle’, *postulat* ‘postulate’, *utverždenie* ‘assertion, claim’, *vyskazyvanie* ‘utterance, statement’, *lozung* ‘slogan’, *deviz* ‘motto’, *rezul’tat* ‘result’, *itog* ‘result, total’, *urok* ‘lesson’ etc. Cf. sentences like

(11) *Osnovnaja ideja konkursa - pust' pobedit sil'nejšij* ‘The main idea of the competition is: let the strongest win’;

(12) *Pervaja mysl' - vy zabyli, o čem vy razgovarivali* ‘The first thought was – you forgot what you talked about’ (Alexander Solzhenitsyn);

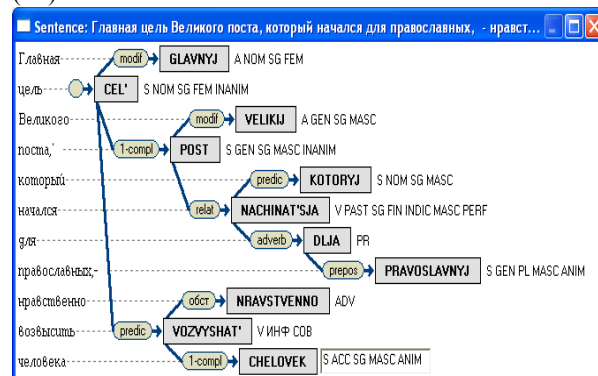
(13) *Važnyj urok stixotvorenija Pasternaka - byt' znamenitym nekrasivo* ‘An important lesson of Pasternak’s poem is: it is not seemly to be famous’, etc.

Mutatis mutandis, these limitations also hold for English. In particular, all English nouns used as equivalents to Russian nouns of the first type (*purpose, aim* etc.) do not allow for the copulative sentence at issue to have a predicative clause, whereas the nouns of the second type (*idea, thought, slogan* etc.) make such a sentence immaculate.

What is it that differs the nouns of these two lexical classes? I believe that the differences between these classes cannot be explained away by semantics alone. In fact, semantic representations of both lexical classes have much in common. In particular, all nouns of the two classes are predicates that have a valency of content. In all examples above, this valency is instantiated by the second part of the copulative sentence, either a non-predicative clause, as in (7) to (10), or a predicative sentential clause, as in (11) to (13). Furthermore, in all cases the text fragment that implements this valency of content characterizes a particular situation.

In my opinion, the difference in the behavior of the two classes is of syntactic nature. We can see that all nouns of the “purpose” class have a specific syntactic feature, **predinf**. Let it be reminded that this syntactic feature, fairly popular in the Meaning ⇔ Theory⁷, is assigned to nouns (or adjectives, see below) that “govern” an infinitive through a real or zero copula, constituting such Russian and English constructions as *Dozvonit'sja na Kavkaz stalo celoj problemoj* [=predinf] ‘It has become a serious problem to get somebody on the phone in the Caucasus’, *Our primary goal* [=predto] *was to find the right operator*; *It was my greatest ambition* [=predto] *to please my father*, etc. Note that the infinitive clause instantiates the (semantic) valency of content of the respective nouns *problema* ‘problem’, *goal* and *ambition* in all these examples even though there is no direct syntactic domination of the

infinitive on the part of the predicate noun. Since the noun *cel'* also possesses the “predinf” feature, the parse of sentence (4) as (4'')



where the second part of the copulative sentence is interpreted as an infinitive clause, is perfectly correct.

Conversely, a correct SyntS like (4'), i.e. a copulative sentence with a predicative clause in the right-hand part, requires that the noun serving as the head for the NP in the left-hand part should possess a **different syntactic feature** that can be defined as the ability to govern a predicative clause through a real or zero copula. By analogy with “predinf” this feature may conveniently be called “**predsent**”. No such feature has existed so far in the MTT or in applied NLP systems based thereon. Accordingly, no constraints could be imposed on the makeup of copulative sentences of the type discussed here, which resulted in the generation by the parsing of SyntS (1') for sentence (1).

By assigning the “predsent” feature to nouns of the “idea” class and placing them in the relevant entries of the dictionary, we will be able to formulate the respective constraints and introduce them into parsing rules. As a result, the construction of correct SyntS for sentences like (11) to (13) will become easy. On the other hand, since no “purpose” class nouns should be assigned this feature, the emergence of SyntS like (4') will be prevented. We will also be able to explain away the ungrammaticality of sentences like

(4b) **Glavnaja cel' Velikogo posta, kotoryj načalsja dlja pravoslavnyx – vozvyšenie čeloveka javljaetsja nravstvennym*,

which is a backward translation of the wrong sentence (1a), as well as the unacceptability of phrases like **Naznačenie etoj pasty – ona xorošo očiščaet zuby* ‘the purpose of this paste is: it cleans the teeth well’, which occur quite often in uneducated oral and written speech. Last but not least, if we ensure that (1) is parsed as (4''), rather than (4'), we

⁷ The “predinf” feature as well as other features of the “pred” series were first introduced in [5] and have been widely used since in research and applications. The English counterpart of “predinf” as introduced in ETAP-3 is “predto” if the infinitive to be used as subject must be preceded by the *to* particle; I do not distinguish between the two features in this paper.

will be able to achieve a more adequate translation in the MT system:

(4c) *The main purpose of the Lent which has begun for the Orthodox is to elevate a person morally.*

Unlike the “predinf” feature that is assigned not only to nouns but also to a great number of adjectives, such as *nравственный* ‘moral’ (it is due to this feature that sentence (2) could be interpreted sententially), the “predsent” feature in Russian is probably confined to nouns, even though, technically speaking, a few exceptions can be observed in the spoken language for adjectives, as in (14), and predicative adverbs, as in (15-16); cf.

(14) *Xorošo ja vovremja kruto vyvernul rul* ‘It’s good I had time to turn the wheel sharply’;

(15) *Žal’, tebja tam ne bylo* ‘Pity you weren’t there’

(16) *Ladno by on prosto ušel* ‘It would be good if he just left’.

As a matter of fact, (14) and (15) are the results of omission of the conjunction *čto* that introduces the subordinate clause and can thus be viewed as transformations of *Xorošo, čto ja vovremja kruto vyvernul rul* ‘It’s good that I had time to turn the wheel sharply’ and *Žal’, čto tebja tam ne bylo* ‘It’s a pity that you weren’t there’. However, such omissions are extremely untypical for Russian unless the subordinate clause is directly dominated by a verb, as in *Ja znaju, čto ty menja ne ljubiš* → *Ja znaju, ty menja ne ljubiš* ‘I know (that) you don’t love me’. Besides, structures like (14) are subject to a number of further constraints. For instance, they are only possible with a zero copula (**Xorošo bylo ja vovremja kruto vyvernul rul* ‘It was good that I had time to turn the wheel sharply’ is totally excluded), cannot be negated or questioned, possess a unique prosodic pattern and have a very peculiar semantics (they convey an idea of rescue or overcoming a difficulty, so that a sentence like *Xorošo u tebja bogataja biblioteka* ‘Good you have a rich collection of books’ is either infelicitous or else implies that the collection helped solve some urgent problem.

In English, the “predsent” feature is easier assigned to adjectives, cf. *It is interesting they didn’t say anything about this, Funny you mention that, It was lucky she had that money on hand* etc.

Let it be emphasized that the two features, “predinf” and “predsent”, are not mutually exclusive. While all nouns of the “purpose” class only have the former and not the latter feature, many nouns of the “idea” class must be assigned the “predinf” feature in addition to “predsent”. These are Russian words like *ideja, princip, lozung, deviz* and their English equivalents *idea, principle, motto* etc., as in

Zamančivaja ideja – dobit’sja upravljaemogo termojadernogo sinteza ‘It’s an alluring idea to achieve controlled thermonuclear fusion’

I would like to note that copulative sentences with the predicative clause are examples of improper, mediated syntactic control of clauses by lexical units. Indeed, the ability of the “idea” class nouns to attach a predicative clause via a copula relates to the ability of nouns to directly govern a predicative clause (as in (17) *Tezis “Bytie opredeljaet soznanie” byl predložen Marksom* ‘The thesis “Being determines consciousness” was proposed by Marx’)

in the same way as the ability of words with the “predinf” feature relates to direct control of infinitives (compare sentences *His plan [= predinf] was to move to the capital* and *He announced his plan to move to the capital*). However, the correlation between a word having the “predsent” feature and this word’s ability to directly dominate a predicative clause is much weaker than the correlation between the presence of the “predto” feature and direct control of an infinitive. While the latter correlation is doubtless (about half the words that have the “predinf” feature also subcategorize for an infinitive), words that subcategorize for a non-interrogative⁸ predicative clause are few and they all probably require that the subordinated clause be graphically marked, as in (17). In any case, the phenomena of direct and indirect domination of predicative clauses must be described independently both in the grammar and the dictionary.

To conclude this study, I would like to note that cases of improper control on the part of nouns are not confined to copulative identifying sentences. For instance, improper control of infinitives can often be observed in a variety of sentence types with specifications, as in

(18) *Glavnaja cel’ u oboix obščestv odna – vozrodit’ moguščestvo zemnoj civilizacii* ‘The main goal of both societies is the same – to resuscitate the power of the terrestrial civilization’ (Arkady and Boris Strugatsky);

(19) *U nee byla kakaja-to strast’ - lovit’ vsex na slove, uličat’ v protivorečii, pridirat’sja k fraze* ‘She had a kind of a passion – take everyone on their word, catch them in contradiction, cavil at the phrase (Anton Tchekhov).

Another type of specifying sentences revealing improper control of the infinitive is represented by the sentences with an introductory *vot* ‘this’:

⁸ A lexical unit’s direct or indirect control of an interrogative sentence (*The question why he did not intervene is a very deep mystery*) is a different linguistic phenomenon that has been thoroughly described in the MTT syntax.

(20) *Byt' ili ne byt' – vot v čem vopros* 'To be or not to be, this is the question'.

The same types of specifying sentences reveal cases of improper control of predicative clauses:

(21) *U Bernarda Šou byla zamečatel'naja mysl': svoboda označae otvetstvennost', poetomu mnogie ee tak bojatsja* 'Bernard Shaw had a splendid thought: Liberty means responsibility. That is why most men dread it'.

(22) *Zdorovyj platit za bol'nogo - vot glavnyj princip medicinskogo straxovanija.* 'The healthy man pays for the sick man – this is the main principle of medical insurance'

Interestingly, in such cases the nouns that indirectly control a predicative clause show a still more liberal behavior. So, neither of the two nouns '*cel*' 'purpose' and '*strast*' 'passion' possesses the "predsent" feature if this is to be defined from prototypical copulative sentences. However, both words acquire the respective controlling properties in sentences with specifications:

(23) *U vsech naučno-populjarnyx žurnalov odna strast' - oni propagandirujut nauku i soedinjajut ee s žizn'ju* 'All popular-science journals have one passion: they popularise the science and combine it with life';

(24) *My delaem gazetę dlja prodaži - vot naša cel'* 'We make the newspaper for sale – this is our purpose'.

It can easily be seen that the study of complex syntactic constructions with nontrivial subcategorization properties of lexical units that I had to undertake because of unsatisfactory performance of an MT system is far from being complete. Many of the things require further research and need a detailed explanation. However, the role of negative material produced by the machine is hard to overestimate.

References

1. Apresjan Ju.D, Boguslavsky Ju.D., Iomdin L.L. *et al.* Lingvisticheskoe obespechenie system ETAP-2. (Linguistics of the ETAP-2 NLP system). Moscow, Nauka, 1989. [In Russian].
2. Apresjan Ju.D, Boguslavsky Ju.D., Iomdin L.L. *et al.* Lingvisticheskij processor dlja slozhnyx informacionnyx system. (A linguistic processor for advanced information systems). Moscow, Nauka, 1992. [In Russian].
3. Iomdin L.L., Sizov V.G., Tsinman L.L. Utilisation des poids empiriques dans l'analyse syntaxique: une application en Traduction Automatique // META, vol. 47. 2002. № 3. P. 351-358.
4. I.A. Melčuk . Opyt teorii lingvističeskix modelej "Smysl ⇔ Text". [The theory of linguistic models "Meaning Text"] Moscow, Nauka, 1974. [In Russian].
5. Iomdin L.L., Melčuk I.A., Pertsov N.V. Fragment modeli ruskogo poverxnostnogo sintaksisa. I. Predikativnye sintagmy. [A fragment of the Russian surface syntax model. I. Predicative syntagms. Naučno-texničeskaja informacija. Serija 2. 1975, No. 7, 30-43. (In Russian).