6th International Conference on Meaning-Text Theory

Proceedings
of the
6th International Conference on Meaning-Text Theory
Prague, August 30-31, 2013

Valentina Apresjan, Boris Iomdin,
Ekaterina Ageeva (eds.)

CONTENTS

Syntactic Subject, Once Again

Igor Mel'čuk

Observatoire de Linguistique Sens-Texte, University of Montreal
Université de Montréal, CP 6128 Centre-ville, H3C3J7 Canada

igor.melcuk@umontreal.ca

To the fond memory of Sasha Kibrik (1939.03.26 — 2012.10.31)

Keywords

Dependency syntax, syntactic subject, nominative *vs*. ergative construction, ergative case, ergative language.

Abstract

The paper proposes a definition of the notion of syntactic subject [= SyntSubj] and discusses it on basis of various languages (Russian, Lezgian, Tongan, Mandarin, Hindi, Archi, Georgian, Basque, Acehnese, and Amele). The SyntSubj is the dependent member of the subjectival surface-syntactic relation, and it is defined by seven properties describing the members of this relation (omissibility, linear position, morphological interaction with the Main Verb [= MV], etc.); all other properties of the SyntSubj are not definitorial, but simply characterizing. The SyntSubj is the most privileged element of the clause in a given language, its privileges being language-specific. The SyntSubj in three major types of language is considered: in languages 1) with no agreement of the MV, 2) with monoactantial agreement of the MV, and 3) with pluriactantial agreement of the MV. Three senses of the adjective ergative are distinguished: ergative language (which has no transitive verbs), ergative construction (in which the SyntSubj is marked not by the nominative), and ergative case (which is used exclusively to mark the SyntSubj or the Agentive complement).

1 The Problem Stated

The notion of Grammatical Subject is a popular topic in linguistics: it suffices to indicate, for instance, such studies as Keenan 1976, Van Valin 1981, Kozinskij 1983, Kibrik 1997, 2001, Testelec 2001: 317–359, Falk 2006, Zimmerling 2012, etc., as well as the collections Li, ed. 1976, Aikhenvald *et al*., eds. 2001, Bhaskarao & Subbarao, eds. 2004, and Suihkonen *et al*.,

eds. 2012. This notion and that of Syntactic Object, known as Grammatical Relations, continue to generate controversy. There is no definition of Syntactic Subject [= SyntSubj], accepted by the whole — or at least by a majority — of linguistic community. The goal of this paper is to propose a rigorous definition for this notion and discuss, in sufficient detail, several complex cases involving the SyntSubj.

The notion of SyntSubj presupposes the notions of syntactic structure of sentences (both surface and deep), of actants, of diathesis and grammatical voice, of transitivity, of ergativity, of agreement and government, of zero lexemes, and still other things. As a consequence, I am forced to limit myself to approximate and sketchy characterization of many relevant phenomena.

2 Conceptual Preliminaries

2.1 *~~Grammatical~~ Relations ⇒ Syntactic Relations*

Speaking of Subject, linguists often mention Grammatical Relations. However, There is no such thing as ***Grammatical** Relations in language: the relations between lexical units in a sentence include semantic, syntactic, and morphological relations. The relations under discussion are, in fact, **syntactic**; therefore, the only term allowed from now on is *syntactic relations*. Moreover, these relations are dependencies. The present discussion is thus based on the following two postulates:

> 1. In any language, an utterance is represented at the syntactic level by its syntactic structure.
>
> 2. The syntactic structure must be a dependency structure, since only this type of structure represents syntactic relations directly and explicitly.

As soon as we agree on these postulates, it becomes obvious that syntactic relations are cross-linguistically universal—and that, in the strongest sense possible: syntactic relations are necessary in any multilexemic utterance of any language, and they always form, in the utterance, a connected structure (= all words of an utterance are syntactically linked between themselves). From this it does not, of course, follow that any **particular** syntactic relation—in our case, the **subjectival** syntactic relation—is universal; that is what has to be shown.

Our discussion of the SyntSubj is based on a **dependency representation** of the syntactic structure of sentences (Mel'čuk 1988, 2004 and 2009). A syntactic dependency relation [= SyntRel] **r** represents a family of syntactic constructions—a set of syntactically similar phrases. The expression of the form "L_1–**r**→L_2" describes all phrases (of language **L**) that can be produced out of two lexemes L_1 and L_2, if L_2 depends on L_1 via SyntRel **r**.

2.2 Syntactic Subject is the Dependent Member of the Subjectival SyntRel

Since the classic paper Keenan 1976, the SyntSubj has been understood as a cluster concept defined inductively. In Keenan's view, the notion of SyntSubj is based on 1) some intuitively

clear cases in the simplest sentences possible—canonical SyntSubjs, and 2) a list of cross-linguistically universal syntactically relevant properties of clause elements (omissibility/non-omissibility, particular linear position, imposing/receiving grammemes, participation in syntactic processes, etc.). Different Synt-elements are compared to canonical SyntSubjs according to these properties; those Synt-elements that are similar enough to the canonical SyntSubjs are also recognized as SyntSubjs. Keenan supplied a detailed checklist of syntactically relevant properties—some 30 plus; this list, developed and supplemented, is extensively used (see, for instance, Iordanskaja & Mel'čuk 2009).

I follow Keenan's approach, defining SyntSubj as **the most privileged clause element** in language **L**. It is the most privileged Synt-element in that it has more of Keenan's properties than any other Synt-element of the clause. However, the 40 years that have passed since Keenan 1976 make it possible to introduce some refinements—namely, the following four guiding principles:

- One has to distinguish between **definitorial** *vs*. **characterizing** properties of the SyntSubj.

- Definitorial properties of the SyntSubj in **L** are established based on **the description of the corresponding SyntRel** in **L**.

- Definitorial properties of SyntSubjs are **language-specific**.

- **"Violations" of SyntSubjs' definitorial properties** caused by clearly statable factors can be allowed, i.e. ignored.

Consider, in **L**, the syntactic configuration "MV–**r**→L", where MV is the Main Verb, i.e., the finite verb ≈ Synt-predicate, and L is a lexeme checked for Synt-subjecthood; **r** is a SyntRel being checked for the status of the subjectival SyntRel.

1. **The SyntSubj's definitorial *vs*. characterizing properties.** Not all the properties on Keenan's checklist have the same weight. Some of them are definitorial; these are coding properties of SyntSubj, which specify the way the **subjectival** SyntRel is realized in texts—roughly, its linear placement and inflection of its both members. These properties concern only the MV, the SyntSubj, and their mutual relationships—and nothing else. **If and only if** at least some of these properties are satisfied, the element under consideration is the SyntSubj.

Other properties on modernized Keenan's list are characterizing; these properties specify the behavior of the SyntSubj with respect to other elements of the clause. They accrue to prototypical, or canonical, SyntSubjs of **L**, but not necessarily to all **L**'s SyntSubjs and not necessarily only to SyntSubjs: a language can have non-canonical SyntSubjs, and a clause element can "masquerade" as a SyntSubj—such that, without being the SyntSubj, it can feature some of its syntactic behavioral properties.

> A Synt-relation—in particular, the subjectival SyntRel—must be defined only by its definitorial (= coding)   properties, strictly separated from the characterizing properties of its dependent member—that is, the corresponding clause element (Iordanskaja & Mel'čuk 2009: 159–160).

Characterizing properties of a SyntSubj are themselves defined on syntactic structures; therefore, this syntactic element must be defined independently from its syntactic behavior in the clause.

> As a particular clause element the SyntSubj must be defined exclusively by its unique coding properties, which specify its relationship with the syntactic head of the clause—that is, the syntactic predicate (≈ MV).

Once defined, the SyntSubj of language **L** must, of course, be characterized by its syntactic behavior in larger formations: for instance, its ability to relativize, its control of deverbal adverbials and/or of reflexives, its control of deletions under coreference, etc. This can throw an interesting light on it—yet this behavior can by no means define it.

The root of disagreement with respect to the identification of SyntSubjs lies in the adopted principle for defining them: either we define the SyntSubj in **L** solely by its coding properties or also we use as well its syntactic behavior—that is, its participation in syntactic processes concerning the whole clause. For me, the choice is clear-cut: the SyntSubj in **L** must be defined exclusively by its coding properties (and then—additionally—characterized by its behavior).

2. **The definitorial properties of SyntSubjs**. The SyntSubj L is the dependent member of a particular SyntRel, which is naturally called **subjectival**: MV–**subjectival**→L$_{SyntSubj}$. SyntSubj's definitorial properties are all and only parameters that can be "read out" from this formula. These parameters deal exclusively with the elements involved in it: they specify under what conditions the **subjectival** SyntRel can be present in, or absent from, the Synt-structure of the clause and how it is implemented in its Morphological Structure. There are seven such parameters, and they are, as an inventory, cross-linguistically universal—in the sense that they are potentially applicable to all languages; however, which parameter is actually relevant in a given language is, of course, language-specific.

---

1) L's immediate dependence exclusively on the MV (L cannot depend on any other clause element).

2) L's non-omissibility from the syntactic structure of the clause.

3) L's particular linear position with respect to the MV and/or with respect to other clause elements.

4) L's morphological impact on the MV (the MV's personal-numeral/class agreement):
$$L–\textbf{agreement}→MV$$

5) The MV's morphological impact on L (the SyntSubj's case marking):
$$L←\textbf{government}–MV$$

6) The MV's inflection that affects morphological links between the MV and L (voice and voice-like phenomena).

7) L's pronominalization that affects morphological links between the MV and L.

Table 1: Defining Parameters of the Syntactic Subject

Comments
– The SyntSubj's definitorial parameters must be tested in the simplest clauses of **L**. in the examples throughout this paper only the simplest clauses are presented: declarative and communicatively most neutral. The MV must be taken in its least marked form: in the present tense of the indicative, in the imperfective (if **L** has aspects), in the active (if **L** has voices), without negation, etc.
– Parameter 2 is aimed at omissibility of a clause element from the syntactic structure of the clause, not from the clause itself. In a Pro-Drop language, a syntactic element can be omitted from the clause, while it is still present

in its structure (e.g., _{Sp.} *Desapareció detrás de la esquina* lit. 'Disappeared behind the corner' actually means '**He/She** disappeared…', where 'he/she' is contextually given). Consider an example from Navajo:

(1) Navajo (Foley & Van Valin 1977: 300–301)

    a. *'Ashkii 'at'ééd yi+ztał* lit. 'Boy girl kicked'.   = 'The boy kicked the girl'.

   and

     'At'ééd yi+ztał lit. '**H**e girl kicked'.   = '**He** kicked the girl'.

  vs.

    b. *'At'ééd 'ashkii bi+ztał* lit. 'Girl boy was.kicked'. = 'The girl was.kicked by the boy'.

   and

     'At'ééd bi+ztał lit. '**H**e girl was.kicked'.   = '**He** was.kicked by the girl'.

Here none of the actants is omissible from the sentence Synt-structure: its physical absence from the sentence signals its pronominalization with the subsequent Pro-Dropping. However, in a sentence such as *The bridge was destroyed* the Synt-actant expressing the Agent is not present in the Synt-structure: the sentence does not mean '… destroyed by HIM/HER/THEM'. In other words, the agent need not be recoverable from preceding discourse (and so it is not amenable to pronominalization) and need not be known or knowable to the speaker.

– Parameter 3 presupposes a preferred word order in a clause without any communicative effects.
– Parameter 6 covers actant-manipulating inflection of the MV—grammatical voice and (in)transitivization (= changes that affect the MV's syntactic valence, but not its semantic valence).
– Parameter 7 requires considering the pronominalization of L, since pronouns often behave differently from nouns (thus, English and Romance pronouns have cases, while nouns do not).

3. **Subjecthood properties are language specific.** A **general** checklist of subjecthood properties is a necessary research tool; however, for each particular **L**, a **particular** list of properties (parameters) should be established, since **L** may have no agreement on the MV, lack case government, and its word order may be too flexible to be relevant. Therefore:

> The list of definitorial and characterizing parameters of the SyntSubj in **L** is specific for **L**.

In addition to universal SyntSubj coding parameters and the standard inventory of characterizing parameters, **L** may have its own SyntSubj's characterizing properties. Since these properties are language-specific, it is only possible to give examples:

– In Dyirbal, only the SyntSubj can be the semantic target of the pluralizing verbal suffix **-ḍay**, which expresses a large quantity of referent(s) of the SyntSubj:

(2) Dyirbal (Australian family; Dixon 1972: 250)

   a. *Bayi*   *yaṛa+Ø*   *ɲinan +ḍa +ɲu*
     the-NOM   man   NOM   sit.down FREQ PRES/PAST
     '**Many** men sat down'.

   b. *Balam*   *miraɲ +Ø*   *baŋgul yaṛa+ŋgu*   *gundal+ḍa +n*
     the-NOM   black.bean NOM the-INSTR man   INSTR   get.collected FREQ PRES/PAST
     '**Many** black beans got collected by the man'.

    VS.

     Bayi   *yaṛa+Ø*   *gundal +ŋa+ḍa +ɲu*   *bagum*   *miraɲ +gu*
     the-NOM   man   NOM   get.collected PASS FREQ PRES/PAST the-DAT   black.bean DAT
     '**Many** men collected black beans'.[i]

– In Malagasy, the interrogative particle VE, which marks a general question, can be linearly placed only before the SyntSubj:

(3) Malagasy (Malayo-Polynesian; *o* = /u/, *ao* = /o/)

    N +anome   *vola*   *an-dRabe **ve***   *ianao?*

PAST give   money to   Rabe   INTERR   you_{SG}

'Did you give money to Rabe?'

4. **The "violation" of subjecthood properties**. A definitorial property of SyntSubjs may be "violated"—yet if a "violation" is triggered by a clearly statable factor, it is irrelevant. Therefore:

> The situation where a definitorial property is not satisfied under precisely described conditions can be safely ignored —as if it were satisfied.

Thus, in Finnish, the SyntSubj is, generally speaking, marked by the nominative; however, if its referent is indefinite, the SyntSubj is in the PART(itive):

(4) Finnish

    Lapse+t   *leikk+i +vät*   *ulkona*
    child   PL.**NOM**   play   PAST **3PL**   outside
    'The children played outside'
   VS.
    Laps+i +a   *leikk+i +Ø*   *ulkona*
    child   PL **PART**   play   PAST **3SG**   outside
    '(Some) children played outside'.

This "violation"—that is, the SyntSubj in the partitive instead of the nominative—can be ignored, since it has an obvious semantic motivation, unrelated to the syntactic role of the SyntSubj.

Summing up: The SyntSubj is to be defined, in language **L**, by using only (some of) the seven language-universal SyntSubj's definitorial properties, with two important provisos:

- Some of the universal subjecthood parameters may be invalid in a particular **L**.

- The privileged character accrues not to the parameter itself, but to its concrete value; which value is privileged in **L** must be established by a detailed examination of the facts of **L**.

Thus, take the linear position of the SyntSubj with respect to the MV. The fact that a particular clause element occupies a fixed position with respect to the MV is in itself not a privilege. In English, the subjecthood privilege is to be linearly placed before the MV and its other actants, which follow the MV, because in the simplest clause featuring the MV and only one actant L, this L precedes MV. But in Malagasy, the Synt-subjecthood privilege is to be placed after the MV and its other actants. Similarly, the control of the MV's agreement is in itself a privilege in English, Russian or French, because only one clause element can control the personal-number agreement of the MV. But in Acehnese, where both the SyntSubj and the DirO impose agreement on the MV and the only actant of the MV can be either SyntSubj or the DirO (see **4.3**, **3b**), the control of the MV's agreement as such is not a privilege. Here, the privilege is to impose agreement by a prefix, which is obligatory and cannot be linearly separated from the verb, while the agreement suffix is not obligatory and can migrate from the verb to the last word of the verb phrase. Non-omissibility is a subjecthood privilege in English, French, etc., because only the SyntSubj is not omissible in these languages, but not in Tagalog, where any actant of the MV is omissible, including the SyntSubj: *May dumating* lit. 'There is having.arrived'. = 'Someone or something has arrived'.

2.3 Syntactic Subject and *"Ergativity"

Most cases of problematic SyntSubjs come from languages with "ergativity." But the noun *ergativity* is vague and does not correspond to a clearly defined notion. It is easier to make more precise the meaning of the adjective *ergative*: it is applicable to three different nouns—*language*, *construction* and *case*—with three different interpretations (Mel'čuk 1988: 251).

– Ergative language is a language in which a typical bi-actantial V that semantically corresponds to a transitive V in a non-ergative language has as the generic component of its meaning the semantic expression 'X <u>undergoes</u> a change, caused by an action of Y on X'; in a non-ergative language the corresponding meaning is converse: 'Y, by an action on X, <u>causes</u> that X changes'. As a result, a V in an ergative language cannot, generally speaking, have a DirO (see, however, Note 4); since a transitive verb is a V that allows a DirO, an ergative language does not have "basic" transitive Vs (it can have transitive Vs produced by diathetic modifications). As the counterpart of transitive Vs, an ergative language features agentive verbs, which require an agentive complement.

Ergative languages include, for instance, Dyirbal, Lezgian, Avar and Archi, see below. This is what could be called deep, or semantic, ergativity. (The current term is syntactically ergative languages.)

– Ergative construction is a construction "SyntSubj←**subj**–MV" where the SyntSubj is marked by a case other than the nominative, something like 'By.me am.reading a.book'. This construction is found, for instance, in Georgian, Hindi, Tongan, Chukchi, Inuktitut and Warlpiri; the presence of an ergative construction characterizes surface, or syntactic, ergativity. (The current term is morphologically ergative languages.) An ergative language, as a general rule, should not have an ergative construction, although logically it is not excluded.

– Ergative case is a case that exclusively marks either a certain type of SyntSubj—namely, a "transitive" or "active" SyntSubj—or an agentive complement; it is found, for instance, in Lezgian, Georgian, Basque and two dead languages of Asia Minor, Urartean and Hurrian. The ergative case does not imply the existence of an ergative construction, and the inverse is also true: an ergative construction can exist without ergative case. The ergative as a typical case of certain SyntSubjs and agentive complements in certain languages is opposed to the nominative, which is also typical of certain SyntSubjs in other languages. The nominative is defined not by its syntactic functions, but by the fact that it is **the case of nomination**: the least marked grammatical case of nouns, appearing, in the first place, when a noun is used to designate an entity (Mel'čuk 1988: 208).

3 Syntactic Subject: An Attempt at a Universal Definition

Definition 1: Syntactic Subject

> The SyntSubj is the **most privileged Synt-actant** of the syntactic predicate (≈ Main Verb) in **L**; what are syntactic privileges in **L** has to be indicated by a specific list of SyntSubj privileges elaborated for **L**.

Comments

– Definition 1 entails the existence of SyntSubj in any **L**, because a language necessarily has the most privileged actant of the MV. It is logically possible for two actants to share the same privileges, but practically, the actants of an MV must be distinguished one way or another, so that one of them stands out.

– Definition 1 does not entail the existence of SyntSubj in any clause of an **L**: subjectless sentences are quite common (e.g., sentences without a finite MV: *What a beautiful day!*, *Ouch!*, *Never in my life*, etc.; or full-blown clauses with a finite MV, but without a SyntSubj—in an **L** that allows for such a state of affairs, such as Lezgian).

– Definition 1 is in full agreement with the hierarchy of clause element types stated in Keenan & Comrie 1977: SyntSubj > DirO > IndirO > Obl(ique)O . This hierarchy is based on the diminishing accessibility of noun phrases for relativization; later it was shown that it also covers many other syntactic operations.

Since Definition 1 does not mention particular properties of any particular **L**, it makes the SyntSubj cross-linguistically universal. However, in a different sense, the SyntSubj is language-specific in so far as syntactic privileges are different in different languages: thus, in many Indo-European languages the main privilege of a clausal element is to impose agreement on the Main Verb, while in Malagasy it is to occupy the clause-final position.

The general notion of SyntSubj can be well illustrated with Russian data, because in Russian it is straightforward.

In Russian, the **subjectival** SyntRel and, consequently, the SyntSubj (boxed in the examples), is defined by the following properties.

1.   The SyntSubj L_2 depends only on the head L_1 of the clause (boldfaced), be it a finite verb or any other element (an infinitive, an interjection, a $V_{\text{IMPER.2SG}}$ form, etc.).

(5) Russian

    a. ⟨ *Ivan* ⟩ ***spit*** ⟨***spal***⟩ 'Ivan is sleeping ⟨was sleeping⟩'.

    b. *A* ⟨ *Ivan* ⟩ *– nu* ***orat´*** *i vyskočil iz komnaty* lit. 'And Ivan—NU to.yell and ran.out of.the room'. = 'And Ivan yelled and ran out of the room'.

    c. ⟨ *Ivan* ⟩ ***bac*** *Petru po morde i vyskočil iz komnaty*
    lit. 'Ivan smack! to.Peter on [his] mug [= 'smacked Peter's mug'] and ran out of the room'.

    d. ***Pridi*** ⟨ *Ivan* ⟩ *vo-vremja, vsë bylo by v porjadke*
    lit. 'Come$_{\text{IMPER.2.SG}}$ Ivan on.time [= Had Ivan come on time], everything would.have been in order'.

2. In Russian, the SyntSubj L_2 is non-omissible from the Synt-structure of the clause whose head is a finite V, since the form of this V is controlled by the SyntSubj (= the MV agrees with the SyntSubj). The sentences in (6) include zero subjects—lexemes having empty signifiers and perceptible only due to their syntactics (a dummy $\varnothing_{\text{(neu, 3sg)}}$, similar to *Eng.* IT and *Fr.* IL; the indefinite personal $\varnothing_{\text{(3pl)}}^{\text{«PEOPLE»}}$, similar to *Fr.* ON and *Ger.* MAN; and the impersonal $\varnothing_{\text{(neu, 3sg)}}^{\text{«ELEMENTS»}}$).

(6) Russian

    a. *Menja*$_{\text{ACC}}$ $\boxed{\varnothing_{\text{(neu, 3sg)}}}$ ***tošn***+***it***$_{\text{3SG}}$ lit. 'It nauseates me'. ≈ 'I feel nauseated'.

    b. *Mne*$_{\text{DAT}}$ $\boxed{\varnothing_{\text{(neu, 3sg)}}}$ ***byl***+***o***$_{\text{NEU, 3SG}}$ *prijatno* lit. 'It was pleasant to.me'. ≈ 'I felt good'.

    c. *Mne*$_{\text{DAT}}$ $\boxed{\varnothing_{\text{(neu, 3sg)}}}$ ***povezl***+***o***$_{\text{NEU, 3SG}}$ lit. 'It favored to.me'. ≈ 'I was lucky'.

d. *Menja*$_{ACC}$ $\boxed{\varnothing_{(3pl)}^{\text{«PEOPLE»}}}$ *xorošo* **prinjal**+i_{PL} lit. '«They» received me well'. ≈ 'I was well received'.

e. *Most*$_{ACC}$ $\boxed{\varnothing_{(3pl)}^{\text{«PEOPLE»}}}$ **snesl**+i_{PL} lit. '«They» demolished the.bridge'.

vs.

Most$_{ACC}$ $\boxed{\varnothing_{(neu, 3sg)}^{\text{«ELEMENTS»}}}$ **snesl**+$o_{NEU.3.SG}$ lit. '«It» destroyed the.bridge' [e.g., a flood or a hurricane].

3. In a declarative sentence, normally the SyntSubj L_2 linearly precedes its governor L_1, although in many cases L_2 may follow L_1 (as determined by a number of particular factors, mainly communicative ones; a list of these is, of course, necessary).

4. The Synt-head of the clause L_1 agrees in person, number and gender with the SyntSubj L_2 and with no other actant. This holds, of course, only if L_1 is a finite verb capable of agreement: thus, in examples 0b–d), the boldfaced L_1 is invariant and does not show agreement.

Agreement of the MV with the SyntSubj

Speaking of MV agreement, three possible complications should be kept in mind.

• I cannot give here a rigorous definition of agreement (see, e.g., Mel'čuk 2006: 58*ff*); an intuitive understanding seems to be sufficient. But the following point must be emphasized:

> "The LU **A** agrees with the LU **B**" does not mean that **A** faithfully copies some features of **B**; this only means that **B** controls the morphological form of **A** in a particular way.

Thus, the Russian MV agrees with a prepositional phrase PO + NP ≈ 'NP each …' in 3SG, NEUTER: *Prixodil+o* [NEU.3.SG] ***po pjat΄ posetitelej v čas*** lit. 'Came **each** five visitors in hour'. = 'Each hour five visitors came'.

• The MV often agrees with a zero dummy SyntSubj, as, for instance, in Zoščenko's sentence [*Nadkus sdelan, i*] *pal΄cem* **smjato** lit. '[A bite is done, and] with.finger [it is] crumpled'. When the MV has the "unmarked/neutral/default form" (e.g., 3SG) in the absence of an overt SyntSubj, this can mean that there is a zero-lexeme SyntSubj $\varnothing_{(3, sg)}$, which imposes this agreement (Mel'čuk 2006: Ch. 9). The failure to have recourse to a zero SyntSubj leads to bizarre results, such as treating a normal DirO as a "derived subject"; see below, **4.2.1**.

5. In Russian, the SyntSubj L_2 is marked by the nominative, except for two cases:

   • if L_2 is not a nominal and cannot have cases;

   • if an over-riding factor intervenes—for example, if L_2 subordinates a numeral (7d), or if L_1 is negated (7e–f).

The nominative is—in any language that has grammatical cases—the case of nomination. It is therefore privileged, and the SyntSubj is generally expected to be marked by the nominative.

(7) Russian
   a. $\boxed{\textit{Idti}}$ **bylo** *trudno* 'To.walk was difficult'.

b. *Čego on* $\boxed{xočet}$, **bylo** *nejasno* 'What he wanted was unclear'.

c. $\boxed{Čto}$ *on bolen*, **bylo** *očevidno* 'That he [is] sick was obvious'.

d. $\boxed{Ix_{GEN}}$ **bylo** *pjatero* 'They were five'.

e. $\boxed{Pis'ma_{PL.NOM}}$ *ne* **prišli** 'The.letters did not arrive'. ~

$\boxed{Pisem_{PL.GEN}}$ *ne* **prišlo** 'No letters arrived'.

f. $\boxed{Ivan_{SG.NOM}}$ *ne* **byl** *na beregu* 'Ivan wasn't on the beach'. ~

$\boxed{Ivana_{SG.GEN}}$ *ne* **bylo** *na beregu* 'There was no Ivan on the beach'.

6. In Russian, the SyntSubj gets demoted by passivization, and its syntactic position goes to the former DirO; cf.:

(8) **a**. $\boxed{Ivan_{NOM\text{-}SyntSubj}}$ *pokupaet kvartiru*$_{ACC\text{-}DirO}$ 'Ivan is buying the apartment'. ~

$\boxed{Kvartira_{NOM\text{-}SyntSubj}}$ *pokupaetsja Ivanom*$_{INSTR\text{-}AgCo}$ 'The apartment is being bought by Ivan'.

b. $\boxed{Ivan_{NOM\text{-}SyntSubj}}$ *kupil kvartiru*$_{ACC\text{-}DirO}$ 'Ivan bought the apartment'. ~

$\boxed{Kvartira_{NOM\text{-}SyntSubj}}$ *byla kuplena Ivanom*$_{INSTR\text{-}AgCo}$ 'The apartment was bought by Ivan'.

7. Pronominalization does not affect the Russian SyntSubj's properties in any special way. All other SyntSubj properties on Keenan's checklist concern not so much the syntactically defined clause elements as some semantic or communicative entities. for instance:

– The control of the coreferential Actor in a phrase ČTOBY + V$_{INF}$ 'in.order.to V' belongs to the semantic Actor rather than to the SyntSubj (*pace* Kozinskij 1983: 18–19); the use of such a phrase depends on the coreference not with the SyntSubj, but with the semantic Actor:

(9) Russian

Mnogie $\boxed{sotrudniki}$ *byli uvoleny* ⟨**lišilis' raboty*⟩, *čtoby sokratit' štaty*
'Many employees were fired ⟨*lost [their] jobs⟩ in.order.to reduce [the] staff'.

The choice of the ČTOBY + V$_{INF}$ construction happens during the SemS ™ DSyntS transition, and it is only natural that the conditions for this choice are semantic (i.e., unrelated specifically to SyntSubj).

– Nichols *et al.* 1980: 376–377 demonstrate that the control of deverbal adverbials in Russian, traditionally ascribed to the SyntSubj, can depend on its Thematicity (= Topicality):

(10) Russian

a. The SyntSubj is thematic:
Pereexav v Moskvu, $\boxed{Ivan_{\textbf{THEME}}}$ *ustroilsja na ètot post*
'Having moved to Moscow, Ivan obtained this position'.
vs.
b. The SyntSubj is rhematic:

**Pereexav v Moskvu, na ètot post ustroilsja* $\boxed{Ivan_{\textbf{RHEM.FOCUS}}}$
'Having moved to Moscow, it is Ivan who obtained this position'.

The authors note (pp. 383–384) that the control of deverbal adverbials with psychological predicates by a dative IndirObj (*Uznav ob ètom,* **mne**$_{DAT}$ *zaxotelos' poznakomit'sja s nim* lit. ≈

'Having learned of this, the desire came to me to meet him') does not constitute an argument in favor of the IndirO's subjecthood: its control capacity—to the extent that such sentences are accepted by speakers—is explained by its semantic and pragmatic roles (it denotes the Experiencer and is Thematic).

– The control of the coreference with the understood "subject" of an infinitive is not an exclusive syntactic property of SyntSubj, either. For instance, in (11), such control belongs to an obvious oblique object *dlja Ivana* 'for Ivan', which is coreferential with the "subject" of the infinitive (it is Ivan who will be going to London):

(11) *Dlja Ivana važno poexat´ v London* 'For Ivan [it is] important to go to London'.

This property accrues to a semantic role (the Experiencer, in this sentence—Ivan, for whom it is important) rather than to a syntactic entity. For a detailed review of characterizing, or behavioral, properties of the Russian SyntSubj, see Testelec 2001: 317–359.

Thus, in Russian, the SyntSubj can be defined clearly and robustly since it is specified by the positive values of all definitorial parameters of SyntSubjs: it depends only on the MV; it is non-omissible; in a declarative sentence, it precedes the MV (if communicative factors do not require inversion, which constitutes an explicable "violation"); it is the only actant of the MV that controls the MV's agreement; it is marked by the nominative case; its role is targeted by the passive; and its pronominalization does not affect its status in any way.[ii] However, the theoretical debate over SyntSubjs (and DirOs) started not with Russian, but with other languages, where this notion is not so straightforward.

4 Establishing the Syntactic Subject in a Language

The most "material," easily observable properties of the SyntSubj is agreement on the MV and the case marking of the SyntSubj itself. Based on the agreement properties of the MV, three major types of language must be examined: the MV either does not agree with its actants at all (= no agreement on the MV): **4.1**; the MV agrees just with one actant (= monoactantial agree-ment on the MV): **4.2**; or else the MV agrees with more than one actant (= polyactantial agree-ment on the MV): **4.3**.

4.1 No Agreement on the Main Verb

Language type 1: If in language **L** the MV does not agree with any of its actants, then we have two situations: **L** either has nominal cases, or it does not.

Subtype 1a. In **L** the MV does not agree with its actants, but the actants are case-marked for their syntactic role.

Subtype 1b. In **L** the MV does not agree with its actants and the actants are not case-marked for their role.

In a Subtype **1a** language, the SyntSubj is the actant L marked by one of four grammatical cases:

1) by the nominative (= the least marked case, that of nomination);

2) by a special case called the subjective (= the case used to mark all and only SyntSubjs, including the only actant of an intransitive verb; the best known subjective is found in Japanese, the case in **-ga**);

3) by another special case, the absolutive (the case used to mark intransitive SyntSubjs and DirOs; we find it, for instance, in Tongan, see below);

4) the SyntSubj can be in a different case, but only exceptionally—with some lexically marked verbs and under special conditions.

A good example of the **1a** language subtype is Lezgian.

The Lezgian MV does not agree with its actants (no person-number or class inflection); there is no voice or any voice-like category. The actants of a verb are distinguished solely by case markings: the only actant of a monoactantial MV is in the nominative, as in (12a), while with a biactantial MV the actant that expresses the Agent is in the special ergative case in **-di**, and the other one, which expresses the Patient, is in the nominative, see (12c):

(12) Lezgian (Daghestanian; Mel'čuk 1988: 207–249)

a. $\boxed{Gada+\varnothing/jar+\varnothing}$ χta $+na$ '[The] boy/s returned'.
    boy   SG/PL   NOM   return   AOR

b. *$\chi ta+na$ 'There.was.returning'.
    return   AOR

c. $Buba+\varnothing+di$ $\boxed{gada+\varnothing/jar+\varnothing}$ gat^ha+na 'Father beat.up [the] boy/s'. =
    father   SG   ERG   boy   SG/PL   NOM   beat   AOR lit. 'By.Father [the] boy/s got.a.beating'.

d. $\boxed{Gada+\varnothing/jar+\varnothing}$ gat^ha+na '[The] boy/s got.a.beating'.
    boy   SG/PL   NOM   beat   AOR

e. *$Buba+\varnothing+di$ gat^ha+na lit. 'By.Father [somebody] got.a.beating'.
    father   SG   ERG   beat   AOR

f. $Buba+\varnothing+divaj$ $\boxed{gada+\varnothing/jar+\varnothing}$ gat^ha+na 'Because.of.father [somebody] beat.up [the] boy/s'.
    father SG   ADEL   boy   SG/PL   NOM beat   AOR

g. $Buba+\varnothing+di$ $\check{c}^h ukur+izva$ 'Father is running'. =
    father   SG   ERG   run   PRES   lit. 'By.Father there.is.running'.

h. $\check{C}^h ukur+izva$ 'There.is.running'.
    run   PRES

i. $Gišin+da$ 'There.is.hunger'.
    hungry   PRES

The actant in the ergative is always omissible, as in (12c) *vs*. (12d), even if it is the only actant explicitly present in the clause, as in (12g) *vs*. (12h). In addition, the name of the Actor can be marked not by the ergative, but by the adelative, and then it is an obvious circumstantial of Cause, cf. (12f). The actant in the nominative is, on the contrary, not omissible, cf. (12b) and (12e). Crucially, (12d) is an absolutely normal, context-independent type of sentence. If both actants are present with a transitive MV, the N$_{NOM}$ is positioned closer to the MV.

Now, some sentences such as (12g–h) might give the impression that the nominative actant is absent, yet it is not the case: the verb Č^{h}UKUR+UN '[to] run' is, in point of fact, a contraction of the phrase Č^{h}UKUR AV+UN 'running do', so that the noun Č^{h}UKUR 'running', not used as such in Modern Lezgian outside of this phrase, plays the role of SyntSubj. Sentences of the type of (12g–h) can be produced exclusively with such "contracted" verbs (which are rather numerous in Lezgian). Genuine subjectless sentences are possible only with semantically

specific—e.g., meteorological or physiological state—verbs: *Meqʻida* '[It] is. cold', *Mičʻida* '[It] is.dark', etc., cf. (12 i). The semantically corresponding Indo-European sentences have—if not an explicit dummy SyntSubj such as Eng. IT, Fr. IL, Ger. ES—a zero lexeme SyntSubj $Ø_{3.SG}$, which imposes the 3SG/NEU form on the verb: Rus. *Xolodn+$o_{3.SG.NEU}$* '[It is] cold' or Sp. *Hac+$e_{3.SG}$ frío* lit. '[It] does cold'. But Lezgian meteorological and similar sentences have no dummy SyntSubj, since the verb knows no number-person or noun class agreement. I conclude that the SyntSubj in Lezgian is the actant marked by the nominative; it has four out of seven SyntSubj's privileges:

1) the exclusive dependence on the MV;

2) non-omissibility;

3) the preferred linear position immediately before the MV;

4) nominative marking.

Lezgian does not have an ergative construction—its SyntSubj is always in the nominative; however, it does have an ergative case, which marks only the agentive complement. And most importantly, Lezgian is an ergative language: a Lezgian verb semantically corresponding to a transitive verb of a language with the nominative construction (most Indo-European, Altaic, Semitic, Bantu, etc.) or of a language with the ergative construction (Hindi, Georgian Basque, Chukchi, etc.) has the basic diathesis that is inverse with respect to this transitive verb. Thus, The English verb 'X beats_up Y' corresponds in Lezgian to a verb meaning 'Y gets a beating from.X'; 'X sees Y' is in Lezgian 'Y is.visible to.X'; etc.

Next, let's examine Tongan. As in Lezgian, the Tongan MV has no number-personal or noun class agreement; but in Tongan the linear placement of actants does not give a clue as to their syntactic role, since it is relatively flexible. Tongan has cases, expressed analytically, among which I will indicate four: the nominative (unmarked, i.e., having a zero marker **Ø**), the absolutive marked by **'a**,[iii] the ergative with the marker **'e** and the dative with the marker **ki**.

(13) Tongan (Malayo-Polynesian; Tchekhoff 1979, Otsuka 2000, 2010)

a. *'Oku   'alu* | *'a   Sione* | 'John is leaving'.
    PRES   leave   ABS John

b. *'Oku   'alu*   'He/She [mentioned in the preceding text] is leaving'.
    PRES   leave

In Tongan, the SyntSubj is not omissible: in (13b), it is present in the Synt-structure, although it is elided from the sentence by a Pro-Drop rule.

c. *'Oku   sio+Ø* | *'a   Sione* | 'John sees'. = 'John is not blind'.
    PRES   see   NEUTR   ABS   John

d. *'Oku   sio+Ø* | *'a   Sione* | *ki   Mele*   'John sees Mary'.
    PRES   see   NEUTR   ABS   DAT Mary

e. *'Oku   sio+'i*   *'a   Mele* | *'e   Sione* | 'John stares at Mary'.
    PRES   see   TRANS   ABS   Mary   ERG John

f. *'Oku   sio+'i*   *'a   Sione*   'He/She [mentioned in the preceding text] stares at John'.
    PRES   see   TRANS   ABS   John

g. *'Oku   sio+'i*   ⎡ *'e   Sione* ⎤   'John stares at him/her [mentioned in the preceding text]'.
    PRES   see   TRANS   ERG John

For the grammemes NEUTR(al) and TRANS(itivizer), see immediately below.

The case marking does not allow us to decide which of the two actants of a two-actantial MV in (13e) is more privileged. Their omissibility is the same (cf. (13f–g)), and, as Lezgian, Tongan has no voice-like (= actant-shuffling) verbal alternations. Yet there are two phenomena that are helpful: cliticization and transitivization with the suffix **-'i** (13e–g)**.**

<u>Cliticization</u>: personal pronominal clitics, which are nearly the only signs allowed between the tense marker and the MV, correspond to the single actant of a $V_{(intrans)}$ and to the ergative-marked actant of a $V_{(trans)}$; the clitics replacing the N_{ABS} and the N_{ERG} are homophonous:

(14) **a.** *'Oku*   ⎡ *ne   /ou* ⎤   *'alu*   'He is leaving'. /'I am leaving'.
    PRES   he-ABS/I-ABS   leave

    b. *'Oku*   ⎡ *ne   /ou* ⎤   *sio+Ø*   'He sees'. /'I see'.
        PRES   he-ABS/I-ABS   see   NEUTR

    c. *'Oku*   ⎡ *ne   /ou* ⎤   *sio+'i*   *'a   Sioné*   'He stares at John'. /'I stare at John'.
        PRES   he-ERG/I-ERG   see   TRANS   ABS John

    d. *'Oku*   ⎡ *ne   /ou* ⎤   *sio+'i*   'He stares at him. /I stare at him'.
        PRES   he-ERG/I-ERG   see   TRANS

    e. **'Oku   ne   /ou*   *sio+'i*   ⎡ *'e   Sione* ⎤   'John sees him [TM *ne*] / me [TM *ou*]'.
        PRES   he-ABS/I-ABS   see   TRANS   ERG John

Clitics correspond either to the N_{ABS} with a $V_{(intrans)}$, as in (14a), or to the N_{ERG} with a $V_{(trans)}$, as in (14b–d), but not to the N_{ABS} with a $V_{(trans)}$, as in (14e); one can conclude that an N_{ABS} with a $V_{(intrans)}$ and an N_{ERG} with a $V_{(trans)}$ are SyntSubjs, as shown by the boxes in (14).

<u>Transitivization</u>: the suffix **-'i**, attached to a semantically bi-actantial $V_{(intrans)}$, turns it into a $V_{(trans)}$, without affecting its semantic valence; V+**'i** requires that its second semantic actant be explicitly expressed as a DirO—i.e., as an N_{ABS}. (NEUTR(al) and TRANS(itivizer) are grammemes of the inflectional category of transitivization, see Note 5) Cf. (14b–d) and (15b), which also identify N_{ERG} as the SyntSubj:

(15) **a.** *'Oku*   *'uma+Ø*   ⎡ *'a   Sione* ⎤   *mo Mele*   lit. 'John kisses with/at Mary'. =
        PRES   kiss **NEUTRAL**   ABS   John   and Mary   'John kisses Mary'.

    b. *'Oku*   *'uma+'i*   *'a   'Mele*   ⎡ *'e   Sione* ⎤   'John kisses Mary'.
        PRES   kiss   **TRANS**   ABS   Mary   ERG John

The SyntSubj's privileges in Tongan then are as follows:

1) it depends only on the MV;

2) it is non-omissible;

3) its case is affected by transitivization;

4) it is the only clause element expressible by a preverbal pronominal clitic.

Thus, like Lezgian, Tongan does have an ergative case, but unlike Lezgian, it does have an ergative construction and is a non-ergative language.

In a **1b** subtype language, which has no "syntactic-oriented" morphology at all, the SyntSubj can be privileged only by its linear position. It is the actant L of the transitive MV that occupies a special linear position in the sentence—the same that occupies the only actant of an intransitive MV. Vietnamese is a good example; here, the SyntSubj immediately precedes the MV:

(16) Vietnamese (Trương 1970)

    a. *Tôi/Giáp* *đã về*   lit. 'I/Giap PAST return'.   = 'I/Giap returned'.

    b. *Tôi/Giáp* *đã đọc* *quyển sách* lit. 'I/Giap PAST read book'.   = 'I/Giap read [the] book'.

Vietnamese has no voice, so the only privileges of its SyntSubj are 1) the dependence on the MV and 2) the preverbal linear position. (I do not know about definitorial properties of the SyntSubj specific to Vietnamese.) However, to prevent possible misunderstandings, let me indicate that the preverbal noun in Vietnamese can also be a prolepsis that expresses the Theme of the sentence, cf. the noun GIÁP in (16c):

    c. *Giáp, nó* *đã đọc* *quyển sách* lit. 'Giap, he PAST read book'.

The same state of affairs is characteristic of many other so-called amorphous/isolating languages, which lack inflectional morphology. For instance, in Mandarin Chinese the preverbal noun is necessarily either a SyntSubj, or a prolepsis expressing the Theme; thus we have:

(17) Mandarin (Li & Thompson 1994: 234–242; $z = $ /c/)

    a. *Zei* *kai -le* *men le* 'Thieves opened the door'.
         thief open PERF door CRS [= particle signaling a Currently Relevant State of affairs]

    b. *Men* *kai -le* 'The door opened'.
         door open PERF/CRS

    c. *Men,* || *zei* *kai-le* 'Door, thieves opened [it]'.
         door thief open PERF/CRS

    d. *Men,* || *kai-le* 'The door, [someone] opened [it]'.
         door open PERF/CRS

In (17a–b) we see two different lexemes of the vocable KAI, just like the English verb OPEN: a transitive and an intransitive one (such verbs are known as *labile*). (17c–d) show MEN 'door' in the syntactic role of a prolepsis (it is marked by a pause and a rising contour); in (17c) the DirO of the verb KAI and in (17d) both the SyntSubj and the DirO are not expressed on the surface.

As one can see from (17d), in Mandarin, the SyntSubj is in principle omissible; here is a couple of another clear examples:

    e. – *Zuo sheme?* lit. 'Do what?' – *Chi zhe* lit. ≈ 'Eating be'.
This exchange is possible in any circumstances when I put my question to somebody about himself or about any other people or animals ('What is/are he/you/they doing?' – 'He/I/They is/am/are eating'.)

f. *Diu-le yi kuai biao* lit. '[Somebody] lost a watch'. = 'A watch was lost'.

4.2 Monoactantial Agreement of the Main Verb

Language type 2: If in **L** any MV agrees with only one of its actants, then this actant is the SyntSubj.
This must be true for the basic (= least marked) forms of the MV, for instance, the imperfective stem; with the perfective stem, the transitive MV may agree with the DirO.

A typical example of a type 2 language is Hindi.

(18) Hindi (*ai* = /ɛ/)
    a. **Intransitive Verbs**

Maĩ	*ā* +Ø +*ā* [⇒ *āyā*]	*hũ̄*	'I [a man] have come'.
I-NOM [male]	come PERF.PART MASC.SG	be-PRES.1.SG	

Maĩ	*ā* +Ø +*ī*	*hũ̄*	'I [a woman] have come'.
I-NOM [female]	come PERF.PART FEM.SG	be-PRES.1.SG	

Ve	*ā* +Ø +*e*	*haĩ*	'They [men] have come'.
they-NOM [males]	come PERF.PART MASC.PL	be-PRES.3.PL	

Ve	*ā* +Ø +*ī*	*haĩ*	'They [women] have come'.
they-NOM [females]	come PERF.PART FEM.PL	be-PRES.3.PL	

    b. **Transitive Verbs: Imperfective Stem**

Maĩ	*čiṭṭʰī* +Ø /*yã*		*likʰ*+Ø	*rah* +Ø	+*ā*	*hũ̄*
I-NOM [male]	letter(fem) SG/PL.NOM		write CONV	remain PERF.PART	MASC.SG	be-PRES.1.SG

lit. 'I [a man] letter/s writing am'. = 'I am writing a letter/letters'.

Ham	*čiṭṭʰī* +Ø /*yã*		*likʰ*+Ø	*rah* +Ø	+*e*	*haĩ*
we-NOM [males]	letter(fem) SG/PL.NOM		write CONV	remain PERF.PART	MASC.PL	be-PRES.1.PL

lit. 'We [men] letter/s writing are'. = 'We are writing a letter/letters'.

vs.

Tū	*čiṭṭʰī* +Ø/*yã*		*likʰ*+Ø	*rah* +Ø	+*ī*	*hai*
you_{SG}-NOM [female]	letter(fem) SG/PL.NOM		write CONV	remain PERF.PART	FEM.SG	be-PRES.2.SG

lit. 'You [a woman] letter/s writing are'. = 'You are writing a letter/letters'.

Tum	*čiṭṭʰī* +Ø/*yã*		*likʰ*+Ø	*rah* +Ø	+*ī*	*ho*
you_{PL}-NOM [female]	letter(fem) SG/PL.NOM		write CONV	remain PERF.PART	FEM.PL	be-PRES.2.PL

lit. 'You [women] letter/s writing are'. = 'You are writing a letter/letters'.

    c. **Transitive Verbs: Perfective Stem**

Maĩ+ne	*čiṭṭʰī* +Ø/*yã*		*likʰ* +Ø	+*ī*	*hai*	/*haĩ*
I INSTR [male]	letter(fem) SG/PL.NOM		write PERF.PART	FEM.SG/PL	be-PRES.3.SG/PL	

lit. 'By.me [a man] letter/s written is/are'. = 'I have written a letter/letters'.

vs.

Tū+ne	*čiṭṭʰī* +Ø/*yã*		*likʰ* +Ø	+*ī*	*hai*	/*haĩ*
you_{SG} INSTR [female]	letter(fem) SG/PL.NOM		write PERF.PART FEM.SG/PL		be-PRES.3.SG/PL	

lit. 'By.you [a woman] letter/s written is/are'. = 'You [a woman] have written a letter/letters'.

In (18c), a perfective transitive MV controls an ergative construction—with the SyntSubj in the instrumental; the noun ČIṬṬʰĪ(YÃ) 'letter(s)' is a DirO, since the passive—as shown in

(18d) —promotes this noun to the SyntSubj, demoting the former SyntSubj to an Ag(entive) Co(mplement:

d. Passive

Čiṭṭʰī +Ø	*likʰ* +Ø	+ī	*ǯā*+Ø	*rah* +Ø	+ī	*hai*
letter(fem) SG.NOM	write PERF.PART	FEM.SG	go CONV	remain PERF.PART	FEM.SG	be-PRES.3.SG

'The letter is being written'.

and

Čiṭṭʰī +yã̃	*likʰ* +Ø	+ī	*ǯā*+Ø	*rah* +Ø	+ī	*haĩ*
letter(fem) PL.NOM	write PERF.PART	FEM.PL	go CONV	remain PERF.PART	FEM.PL	be-PRES.3.PL

'The letters are being written'.

ǮĀ 'go' is the passive auxiliary, here in the form of converb ≈ gerund; RAHĀ 'remain' is the progressive auxiliary, which takes the converb ≈ gerund of the lexical verb)

Hindi is thus a non-ergative language: its transitive verb admits a DirO, and the meaning of a transitive verb typically has 'cause' as the generic component. It has no ergative case, either, but it does have an ergative construction—with a transitive MV in a past (perfective) form and the SyntSubj in the instrumental, the MV agreeing only with the DirO. With an imperfective MV, Hindi uses a nominative construction, and the verb agrees then with the SyntSubj. (In other words, Hindi manifests split ergativity.)

The SyntSubj's privileges in Hindi are:

1) the dependence on the MV;

2) non-omissibility,

3) the linear position before the MV and other actants;

4) the control of agreement of the MV (in an imperfective form);

5) the nominative case (again, with an imperfective MV);

6) the "passivizability" (that is, being the target of promotion by the passive).

The things are substantially different in Archi. Although, just like Hindi, Archi has a monoactantial agreement—if the MV is in the one of the least marked synthetic forms, as in (19), the actant of the MV that controls its noun-class agreement is itself—in contrast to Hindi—always in the nominative; it is not omissible and its syntactic position is targeted by an actant-manipulating voice-like transformation (as before, this actant is boxed in the examples; it is the SyntSubj, as will be shown).

(19) Archi (Daghestanian; Kibrik 1977, 2003: 332–368; Roman numbers stand for noun classes)

a. *Buwa* +Ø+Ø *da*+qʻa 'Mother came'.
mother(**II**) SG NOM **II** come-PERF

b. *Dija* +Ø+mu *buwa* +Ø+Ø *χir* *a*+r+ u
father(**I**) SG INSTR mother(**II**) SG NOM behind do.**II**.do-AOR

lit. 'Father Mother behind did' ["behind do" is an idiom meaning 'bring with oneself']. = 'Father brought Mother with him'.

c. *Dija* +Ø+mu *dos* +Ø+Ø *χir* *a*+w+u [⇒ *aw*]

father(I)   SG INSTR   friend(I) SG NOM   behind do.**I**.do-AOR
'Father brought a friend with him'.

d. *Dija   +Ø+mu   dos +til+Ø   χir   a+b+u*
father(I)   SG INSTR   friend(I) PL NOM   behind do.**III**.do-AOR
'Father brought friends with him'.

All plural nouns belong to noun class III; the verb AS 'do' shows class III agreement with the plural *dostil* 'friends'.

e. *Dija   +Ø+n   buwa   +Ø+ɬu anχ   +Ø+Ø   a+Ø+w*
father(I)   SG GEN   mother(II) SG COMIT   fight(Noun, IV) SG NOM   do .**IV**. do-PERF

lit. 'Father's with.Mother fight was.done'. = 'Father fought with Mother'.

f. **Dija +Ø+mu   kunne. ~ Dija +Ø+mu   kummul   kunne.*
father(I) SG ERG   eat-AOR   father(I) SG ERG   food-SG.NOM   eat-AOR
'Father ate'.

g. (i) *Balah +Ø+Ø   b+erχin 'Trouble gets forgotten quickly'.*   *dita+b+u*
trouble(III) SG NOM   soon.**III**.soon   **III** forget
  (ii) *Arša   horōk   ej +b+u   iškul   +Ø+Ø   da   +b+lu*
Archi-INESS   long.ago very.**III**.very   school(III) SG NOM   open .**III**. open-AOR
    'A school opened in Archi very long time ago'.
  (iii) ***D**+ez   un   malgan*
    II   I-DAT   you(SG(II)-NOM) be.dear-PRES
    'You [singular female] are dear to me'. = 'I love you'.

The SyntSubj in Archi has six privileges:

1) it depends only on the MV;

2) it is non-omissible, while all other actants of the MV can be absent (cf. (19f), where a generic noun 'food' must be used in the nominative);

3) it is positioned immediately before the MV after all other actants;

4) it controls—almost exclusively—the noun-class agreement not only of the MV, but also of circumstantials and even of certain actants, as in (19g), where the adverb *dita**b**u* 'soon', the particle *ej**b**u* 'very' and the actant ***d**ez* 'to.me' agree in noun class with the SyntSubj;

5) it is always marked by the nominative;

6) Archi has a "converse" voice,[iv] which promotes the AgCo to the SyntSubj, while demoting the former SyntSubj to the DirO:

(20) **a**. *Buwa +Ø+mu   χʷalli+Ø+Ø   b+a+r +ši   b+i*
mother(II) SG   INSTR bread(III) SG NOM   **III** do IMPF CONV III be-PRES
    lit. 'By.mother(AgCo) bread(SyntSubj) doing is'. = 'Mother is baking bread'.
  vs.
    b. *Buwa +Ø+Ø   χʷalli+Ø+Ø   b+a +r +ši   d+i*
mother(II)   SG   NOM   bread(III) SG NOM   **III** do IMPF CONV II be-PRES
    lit. 'Mother(SyntSubj) bread(DirO) doing is'. = 'Mother is baking bread'.

NB: **1**. The two sentences in (20) contrast in that (20a) answers the question "What is happening?", while(20b) constitutes an answer to the question "What about Mother?": in a sentence of this type, the SyntSubj must be Thematic.

2. In (20b), the auxiliary agrees with the SyntSubj, while the converb—with the DirO.

To sum up: Archi does not have an ergative construction, since its SyntSubj is always in the nominative; it does not have an ergative case, either: its AgCo is in the instrumental. But like Lezgian, Archi is an ergative language.

4.3 Polyactantial Agreement of the Main Verb

Language type 3: In language **L**, the MV agrees simultaneously with two actants, using two sets of agreement markers. In some languages, the MV can simultaneously agree with three or even four actants. However, in order to simplify, I consider here the MV's agreement just with two actants —L_1 and L_2, one of which is thus the SyntSubj and the other one, the DirO. This introduces into our inquiry an additional dimension: the necessity to distinguish between SyntSubjs and DirOs.

For Type **3** languages, two situations must be distinguished: either a monoactantial MV uses only one set of agreement markers, or it uses alternatively both (as a function of the lexical unit).

Subtype 3a. In **L**, the transitive MV agrees simultaneously with two actants, but a monoactantial MV features only one type of agreement.

Subtype 3b. In **L**, the MV agrees simultaneously with two actants, and a monoactantial MV features both types of agreement.

In a Subtype 3a language, the only actant of an intransitive MV is its SyntSubj, so that the researcher has to decide between two actants of a transitive biactantial MV. Such a situation is found in many languages; I select two for an examination—Georgian and Basque.

Georgian. In contrast to Lezgian and Archi, a transitive Georgian MV agrees—in person and number—simultaneously with two of its actants, which are, therefore, the SyntSubj and the DirO (for simplicity's sake, I leave out the agreement with the IndirO—rather than with the DirO—possible with some verbs). We have to settle accounts between these two: which one is boss—i.e. the SyntSubj? A transitive verb has two sets of agreement markers: Sets I and II. Only the markers of Set I are exclusively used for the actant of a monoactantial MV, which stands in most cases in the nominative, cf. (21a); it is a SyntSubj. But this fact by itself is not sufficient to consider Set I prefixes as exclusively subject markers: on a transitive verb, they can in principle cross-reference the DirO: precisely this, as we will see, happens in Basque. One has to compare both actants of a transitive MV as to their case-marking and mutual linear order. In the least marked transitive clause, with the MV in a tense of the present series, the actant cross-referenced by Set I markers is in the nominative and precedes the MV and the other actant, just as the SyntSubj of an intransitive MV. The other actant, which is in the dative, in a communicatively neutral sentence either follows the MV, or precedes it while following the nominative actant. Therefore, the first—nominative—actant is the SyntSubj of the transitive MV, so that Set I markers must be considered to be subject markers. As a result, the SyntSubj in Georgian is the element cross-referenced by subject markers; it is boxed in the examples of (21), and the subject markers are boldfaced.

(21) Georgian (the morphological representation is drastically simplified)

a. Intransitive Verbs: Present and Aorist

| K'ac+Ø +i | berd+eb | +**a** | '[The] man ages'. ~ |
| man SG NOM | age | PRES | |

| K'ac+eb+i | berd+eb | +**i**+**an** | '[The] men age'. |
| man PL NOM | age | PRES **IND** | |

| Me | **v**+berd+eb +i +**Ø** | 'I age'. | | Čven | **v** +berd+eb +i +**t** | 'We age'. |
| I- | **1**$_{SUB}$ age PRES IND | | | we-NOM | **1**$_{SUB}$ age PRES IND | |

| K'ac+Ø +i | da+berd+Ø +**a** | | '[The] man aged'. ~ |
| man SG NOM | PERF age AOR **3.SG**$_{SUB}$ | | |

| K'ac+**eb**+i | da+berd+Ø +**nen** | | '[The] men aged'. |
| man PL NOM | PERF work AOR **3.PL**$_{SUB}$ | | |

| Me | da+**v**+berd+Ø+i +**Ø** | 'I aged'. ~ | Čven | da+**v**+berd+Ø +i +**t** | 'We aged'. |
| I-NOM | PERF1$_{SUB}$ age AOR IND **SG**$_{SUB}$ | | we-NOM PERF1$_{SUB}$ work AOR IND **PL**$_{SUB}$ | | |

b. Transitive Verbs: Present and Aorist

(i) K'ac+Ø +i   m +xat'+av+**s**   me ~ g+xat'+av+**s** šen ~ xat'+av+**s** mas/mat
man   SG NOM **1SG**$_{OBJ}$ draw PRES **3.SG**$_{SUB}$ I-DAT   **2**$_{OBJ}$ you$_{SG}$-DAT   he-/they-DAT
'[The] man draws me ~ you$_{SG}$ ~ him/them'.

(ii) K'ac+eb+i   m +xat'+av+**en**   me ~ g+xat'+av+**en** šen ~ xat'+av+**en** mas/mat
man   PL NOM **1.SG**$_{OBJ}$ draw   PRES **3.PL**$_{SUB}$ I-DAT   **2**$_{OBJ}$ you$_{SG}$-DAT   he-/they-DAT
'[The] men draw me ~ you$_{SG}$ ~ him/them'.

(iii) Me   **v** +xat'+av+**Ø**   mas/mat   'I draw him/them'.
I-NOM   **1**$_{SUB}$ draw PRES **SG**$_{SUB}$ he/they-DAT

(iv) K'ac+Ø+ma   da +m +xat'+Ø +**a** me da+g+xat'+Ø+**a** šen   da+xat'+Ø+**a**
man   SG ERG   PERF 1.SG$_{OBJ}$ draw AOR **3.SG**$_{SUB}$ **2**$_{OBJ}$ you$_{SG}$-DAT   he-/they-NOM
'[The] man drew me ~ you$_{SG}$ ~ him/them'.

(v) K'ac+eb+ma   da +m +xat'+Ø +**es**   me ~ da+g+xat'+Ø+**es** šen ~ da+xat'+Ø+**es** is /isini
man   PL ERG   PERF 1.SG$_{OBJ}$ draw AOR **3.PL**$_{SUB}$ I-NOM   **2**$_{OBJ}$ you$_{SG}$-NOM   he-/they-NOM
'[The] men drew me ~ you$_{SG}$ ~ him/them'.

(vi) Me   da+ **v** +xat'+Ø+e is/isini   'I drew him/them'.
I-ERG   PERF **1**$_{SUB}$ draw +Ø   he/they-NOM

The Georgian SyntSubj has six privileges:

1) It depends only on the MV;

2) it is non-omissible;

3) it controls the same type of agreement for intransitive and transitive verbs, imposing sub-
ject affixes;

4) normally, it precedes the MV and the other actants;

5) with the MV in one of the present series tenses, the SyntSubj is in the nominative: (21a–d);
with a transitive MV, the DirO is in the dative: (21c–d). This is the most common
nominative construction, such as seen in many languages;

6) Georgian has a passive, which confirms the subjecthood of the noun in the ergative:

(22) *Gogi+m*   *es*   *st'at'ia+Ø+Ø*   */st'at'i+eb+i*   *da+c'er+Ø+a*
      Gogi **ERG**   this   paper   SG **NOM** / paper   PL **NOM**   PERF write AOR 3.SG
        'Gogi wrote this paper/these papers'.

    vs.

    Es   *st'at'ia+Ø+Ø*   */st'at'i+eb+i*   *da+c'er+il+i*   *iq+o*
    this   paper   SG **NOM** /paper   PL   **NOM** PERF write PASS.PART SG   be AOR.3.SG
    Gog+is   **mier**
    Gogi GEN **by.means**
     'This paper/These papers was/were written by Gogi'.

       NB: The Georgian MV does not reflect the plural of an inanimate SyntSubj; that is why **dacʿerili**
       'written' and **iqo** 'was' are in the singular for both 'paper' and 'papers'.

If the MV is in an aorist series tense, the case marking of the SyntSubj and the DirO changes
to, respectively, the ergative and the nominative, as in (21e–h), although their syntactic status
does not change. A transitive Georgian MV in an aorist series tense and its two main actants
form, of course, an ergative construction. (Just like Hindi, Georgian manifests split ergativity:
the ergative construction appears only with aorist series tense forms; elsewhere we have the
nominative construction.) In accordance with the convention concerning SyntSubj property
"violations," the appearance of the ergative instead of the "canonical" nominative does not
make the definition of the SyntSubj in Georgian any more problematic.

Georgian has the ergative construction and the ergative case, but it is not an ergative
language.

Basque. The Basque transitive MV also agrees simultaneously with at least two of its actants.
As in Georgian, there are two sets of agreement affixes, the prefixes I$_{NOM}$ cross-referencing the
N$_{NOM}$ and the affixes II$_{ERG}$ cross-referencing the N$_{ERG}$; for the single—nominative—actant of
an intransitive MV only the affixes of set I$_{NOM}$ are used. But here comes the important
difference with Georgian: with a transitive MV, one of its two actants is always in the
ergative; there is no tense-induced ergative split—that is, no nominative construction that
helps us identify the SyntSubj; Basque has an ergative construction in all tenses. In Basque,
we cannot know which affixes are subjectival. Therefore, in the following example, the
boldfaced agreement affixes are specified by the noun they cross-reference: N$_{NOM}$ *vs*. N$_{ERG}$; for
instance, "3$_{NOM}$" as a gloss of a marker **m** means '**m** cross-references the N$_{NOM}$', etc.

(23) Basque (*s* = /ś/, *tx* = /č/, / *z* = /s/)

    a. **Intransitive Verbs**

      (i) Gizon+a +Ø+Ø   *etorri*   *d* +*Ø* +*a*   'The man has come'.
        man   DEF SG NOM   come-PERF.PART 3_{NOM} SG_{NOM} be

      (ii) Gizon+a +k +Ø   *etorri*   *d* +*ir* +*a*   'The men have come'.
        man   DEF PL NOM   come-PERF.PART 3_{NOM} PL_{NOM} be

      (iii) Ni+Ø   *etorri*   *n* +*aiz*   'I have come'.
        I   NOM   come-PERF.PART 1.SG_{NOM} be

    b. **Transitive Verbs**

      (i) Gizon+a +Ø+k   *kotxe+a +Ø+Ø*   *saldu*   *d* +*Ø* +*u* +*Ø*+*Ø*
        man DEF SG ERG car DEF SG NOM sell-PERF.PART 3_{NOM} SG_{NOM} have 3_{ERG} SG_{ERG}
      'The man has sold the car'.

      (ii) Gizon+a +Ø+k   *kotxe+a +k+Ø*   *saldu*   *d* +*it* +*u* +*Ø*+*Ø*
        man DEF SG ERG car DEF PL NOM sell-PERF.PART 3_{NOM} PL_{NOM} have 3_{ERG} SG_{ERG}
      'The man has sold the cars'.

      (iii) Gizon+e +k   *kotxe+a+Ø+Ø*   *saldu*   *d* +*Ø* +*u* +*Ø*+*te*
        man PL.DEF ERG car DEF SG NOM sell-PERF.PART 3_{NOM} SG_{NOM} have 3_{ERG} PL_{ERG}
      'The men have sold the car'.

      (iv) Gizon+e +k   *kotxe+a +k+Ø*   *saldu*   *d* +*it* +*u* +*Ø*+*zte*
     man PL.DEF ERG car DEF PL NOM sell-PERF.PART 3_{NOM} PL_{NOM} have 3_{ERG} PL_{ERG}
      'The men have sold the cars'.

      (v) Ni+k   *kotxe+a +Ø+Ø*   *saldu*   *d*+*Ø* +*u* +*t*
        I   ERG car DEF SG NOM sell-PERF.PART 3_{NOM} PL_{NOM} have 1SG_{ERG}
        'I have sold the car'.

      (vi) Ni+k   *kotxe+a +k+Ø*   *saldu*   *d*+*it* +*u* +*t*
        I   ERG car DEF PL NOM sell-PERF.PART 3_{NOM} PL_{NOM} have 1SG_{ERG}
        'I have sold the cars'.

A transitive MV cross-references its DirO by the same markers as an intransitive MV cross-references its SyntSubj. For this reason, in Basque, the existence of actant-shuffling modifications of the verb is really crucial. The language has two such modifications (Rebuschi 1978: 76–77 and 82–83; Rebuschi 1981: 92, 1982: 299*ff*, 1986; Rebuschi's data are quoted with simplifications): a passive and two detransitivizations, which target the SyntSubj's syntactic position.

Passive: a diathetic conversion "DirO′ ⇒ SyntSubj, SyntSubj′ ⇒ AgCo" (the prime means 'old').

The Basque passive is illustrated in (24), where the sentences correspond to the sentences in (23b):

(24) **a.** (i) | *Kotxe+a +Ø+Ø* | *gizon+a +Ø+k* | *saldu* | *+a+Ø* | *d +Ø* | *+a* |

| car | DEF | man | DEF SG ERG | sell-PERF.PART DEF SG | 3_{NOM} SG_{NOM} | be |

'The car is sold by the man'.

(ii) | *Kotxe+a +k+Ø* | *gizon+a+Ø+k* | *saldu* | *+a +k* | *d +ir* | *+a* |

car DEF PL NOM | man DEF SG ERG | sell-PERF.PART DEF SG | 3_{NOM} PL_{NOM} be

'The cars are sold by the man'.

(iii) | *Kotxe+a+Ø +Ø* | *gizon+e +k* | *saldu* | *+a +Ø* | *d +Ø* | *+a* |

car DEF SG NOM | man DEF.PL ERG | sell-PERF.PART DEF SG | 3_{NOM} SG_{NOM} be

'The car is sold by the men'.

(iv) | *Kotxe+a +k +Ø* | *gizon+e +k* | *saldu* | *+a +k* | *d +ir* | *+a* |

car DEF PL NOM | man DEF.PL | sell-PERF.PART DEF PL | 3_{NOM} PL_{NOM}be

'The cars are sold by the men'.

(v) | *Kotxe+a +Ø+Ø* | *ni+k* | *saldu* | *+a +Ø* | *d +Ø* | *+a* |

car DEF SG NOM | I ERG | sell-PERF.PART DEF SG | 3_{NOM} SG_{NOM} be

'The car is sold by me'.

(vi) | *Kotxe+a +k+Ø* | *ni+k* | *saldu* | *+a +k* | *d +ir* | *+a* |

car DEF PL NOM | I ERG | sell-PERF.PART DEF PL | 3_{NOM} PL_{NOM} be

'The cars are sold by me'.

Detransitivizations: they result in "SyntSubj$_{ERG.}$ ⇒ SyntSubj$_{NOM.}$" [v]

Basque has a progressive construction, marked by the Adj ARI 'being in the process of, doing' and using the auxiliary IZAN 'be'—even for transitive verbs, which become *eo ipso* intransitive (since a transitive verb uses as its auxiliary only UKAN 'have'): the SyntSubj, instead of the ergative, takes the nominative, as an intransitive SyntSubj should; the former DirO remains in the nominative, but loses its status as a DirO, since the verb becomes intransitive; the MV agrees only with the SyntSubj:

(25) **a**. (i) | *Gizon+a +Ø+Ø* | *kotxe+a +Ø/k+Ø saltzen* | *ari* | *d +Ø* | *+a* |

man DEF SG NOM | car DEF SG/PL sell-GER | doing 3_{NOM} SG_{NOM} be

'The man is selling the car/s'.

(ii) | *Gizon+a +k+Ø* | *kotxea/kotxeak saltzen* | *ari* | *d +ir* | *+a* |

man DEF **PL** NOM | car-DEF.SG/PL.NOM sell-GER | doing $3_{NOM}$$PL_{NOM}$ be

'The men are selling the car/s'.

The other detransitivization (called "antipassive" in Rebuschi 1981: 92) produces a resultative construction:

b. (i) | *Gizon+a+Ø+Ø* | *kotxe+a+Ø/k+Ø saldu* | *+a +Ø* | *d +Ø* | *+a* |

man DEF SG NOM | car sell-PAST.PART DEF SG | 3_{NOM} SG_{NOM} be

'The man is having.sold the car/s'.

(ii) | *Gizon+a+k+Ø* | *kotxea/kotxeak saldu* | *+a +k* | *d +ir* | *-a* |

man DEF PL NOM | car-DEF.SG/PL.NOM sell-PAST.PART DEF PL | 3_{NOM} PL_{NOM} be

'The men are having.sold the car/s'.

The four Basque SyntSubj's privileges are as follows: 1) it depends exclusively on the MV; 2) it is non-omissible; 3) it tends to precede the MV and other actants; 4) its role is targeted by the passive and is confirmed by detransitivizations.

In conclusion, Basque is a non-ergative language, but it does have an ergative construction (without split) and an ergative case.

As for Subtype **3b** languages, probably the best-known example here comes from Acehnese. According to Durie 1985: 190 and 1987, Acehnese has no syntactic processes: no voices, no raisings, no detransitivization, no switch-reference, etc.; word order is extremely flexible. The only reliable syntactic property of actants of the Main Verb that amounts to a privilege is verb agreement—cross-referencing of actants on the MV. It cross-references two of its actants (only if they are animate): one by a prefixal marker, the other by a suffixal marker. However, with a semantically monoactantial verb having just one syntactic actant both types of agreement occur, which means that in (26a) and (26b) we see two different types of actant— one controlling prefix agreement and the other controlling suffix agreement:

(26) Acehnese (Malayo-Polynesian; Durie 1985, 1987, 1988; *ê* = /e/, *ô* = /o/, *eu* = /ɯ/, *ë* = /ɤ/, *j* = /ʒ/)

    a. ***Lôn**+lôp*   'I enter'.   and ***Geu**+lôp*   'He enters'.
         1.SG   enter   3.SG   enter

    b. *Rhët+**lôn***   'I fall'.   and *Rhët+**geuh***   'He falls'.
         fall   1.SG   fall   3.SG

Thus, both types of actant are privileged in Acehnese, since they, and only they, control the agreement of the MV. Therefore, one of these actants must be the SyntSubj and the other, the DirO. To decide which one of the two is more privileged than the other and thus is th SyntSubj, we need to consider a biactantial verb in a sentence where both types of actant are expressed:

    c. ***Lôn**+ngieng+**geuh*** 'I see him/her'.   ~ ***Geu**+ngieng+**lôn*** 'He/She sees me'.
       1.SG   see   3.SG   3.SG   see   1.SG

Examining sentences with two privileged syntactic actants, we find that:

    • The prefixal marker on the verb is obligatory and cannot be linearly separated from the verb (26d-i), while the suffixal marker is not obligatory and can migrate to the outer edge of the verbal phrase (26d-ii):

    d. (i) *Gopnyan **lôn**+ngieng* 'Him I.see'.   ~ *Lôn **geu**+ngieng* 'Me he.sees'.
          he   1.SG   see I 3.SG   see
      vs.
        Lôn ngieng+geuh***   'I see.him'. ~ **Gopnyan ngieng+**lôn*** 'He sees.me'.
         I   see   3.SG   he   see   1.SG
    (ii) *Ka+leupah+**lôn** u keude baroe*.
       PAST reach   1.SG to town   yesterday
     'I reached the town yesterday'.
     Lôn ka+leupah u keude baroe. ≡ *Ka+leupah u keude baroe+**lôn***.

    • The imperative requires the prefixal marker and does not allow the suffixal one:

    e. (i) ***Neu**+peumeu'ah!* 'Forgive [me]!' ~ **Peumeu'ah! ~*
             Neu+peumeu'ah+lôn**!* 'Forgive.me!'

2.SG   forgive

(ii) ***Neu***+*peujêt*   *ie*   *nyoe*   *keu* *jih!* 'Make him drink this water!'
      2.SG make.drink water this   to   he

Only the prefix-referenced actant can be the Addressee of an imperative utterance.

• The prefix-referenced actant, and only this actant, can be introduced by the preposition **lê,** when following the Main Verb:

f. *Gopnyan lôn+tët* +*rumoh **lê** lôn* 'I burned down his house'.
    he   1SG   burn house I   lit. 'He I.house.burnt by I'.

Therefore, the prefix-referenced actant is more privileged in Acehnese: it is the SyntSubj. The other one, suffix-referenced, is the DirO. This simply means that in (26b) a literal gloss should be rather 'It.falls me/him'. Durie himself calls these two actants Agent and Undergoer, since 30 years ago the notions of SyntSubj and SyntObj were too vague to be of any use; Durie 1985: 190–191 correctly indicates that none of Acehnese clause elements corresponds to the characteristics of the "syntactic pivot," a moot concept used at the time instead of SyntSubj. However, Durie makes it absolutely clear that "Agent" and "Undergoer" are not genuine semantic relations, but clearly syntactic ones (see especially Durie 1987). Therefore, it can be safely concluded that, by calling the prefix-referenced actant the SyntSubj and the suffix-referenced one the DirO, I simply sharpen and, at the same time, generalize the terminology.

What is special about the Acehnese SyntSubj and DirO is their more direct link to semantic roles. In many such languages as English or Russian, a SyntSubj can fulfill various semantic roles: it can express an Agent (***John** beat up Paul*), a Patient (***John** got a beating*), a Cause (***John** really worries us*), an Experiencer (***John** likes boiled potatoes*), a Property Carrier (***John** is intelligent*), Time (***The next morning** saw John in Nevada*), and so on; to a lesser extent, the same is true of the DirO. But in Acehnese, the SyntSubj expresses only the volitional Actor, and the DirO only the non-volitional Undergoer. The semantic opposition of volitionality is extremely important; Acehnese has special derivational means to change the volitionality of a verb (Durie 1988: 7): *jak* 'go, walk' ~ ***teu***+*jak* 'walk without volition' or *seunang* 'be happy' ~ ***meu***+*seunang* 'make oneself happy, enjoy oneself'. However, such an alignment of syntactic relations to semantic roles by no means diminishes the importance of syntactic relations.

With the proposed terminological change, one can draw an interesting parallel between the Acehnese sentences of the (26b) type—that is, with a verb that has a DirO only, but no SyntSubj —and Russian impersonal constructions in which the only semantic actant of the verb is expressed by a DirO (the verb in these constructions expresses an incontrollable state):

(27) ***Menja***_{ACC} *tošnit/rvët*   lit. '[It] nauseates/vomits me'. = 'I am nauseated/I vomit'.

    Menja_{ACC} *znobit*   lit. '[It] chills me'.   = 'I have a chill'.

    Menja_{ACC} *trjasët*   lit. '[It] shakes me'.   = 'I shake'.

    Menja_{ACC} *proneslo*   lit. '[It] diarrhea-ed me'.   = 'I had diarrhea'.

    Menja_{ACC} *skrjučilo*   lit. '[It] completely.bent me'. = 'I was doubled up [in pain]'.

    Menja_{ACC} *razneslo*   lit. '[It] expanded me'.   = 'I got fat'.

I do not see any substantive difference between Acehnese *Sakêt-lôn* lit. '[It] hurts/sicks me'. = 'I am hurting/sick', which is an impersonal construction, and the Russian impersonal construction of the type *Menja rvët* lit. '[It] vomits me'. = 'I vomit'. The difference is quantitative: Russian has a handful of such impersonal verbs, while in Acehnese there are hundreds of them.

5 SyntSubj Problems Related to Impersonal Constructions

Often, the dubious treatment of an actant as the SyntSubj is due to the failure to recognize the presence of a zero dummy subject, a lexeme similar to the expletive and meteorological IT of English, but having an empty signifier. Let me consider two cases, in Icelandic and in Amele. Icelandic has a common type of sentences of the form in (28):

(28) Icelandic (Andrews 2001)

a. *Bát+Ø* +*inn* / *Bát+a* +*na* rak á land 'The boat/s drifted to shore'.

   boat SG.**ACC** DEF / boat PL.**ACC** DEF   drift-PAST.3SG to   shore lit. '[It] drifted the.boat/s to shore'.

b. *Bát+i* +*num* /*Bát+u* +*num* hvolf+di 'The boat/s capsized'.

   boat SG.**DAT** DEF /boat PL.**DAT** DEF   capsize PAST.3SG lit. '[It] capsized the.boat/s'.

c. (i) *Hann*   kasta+ði   stein+i +*num*/stein+u +*num* 'He threw with.the.stone/s'.

   he-NOM   throw PAST.3SG   stone SG.DAT DEF /stone PL.DAT DEF

  (ii) ***Stein+i* +*num*/*Stein+u* +*num* var**   kasta+ð 'The stone/s were thrown'.

   stone   SG.**DAT** DEF /stone PL.**DAT** DEF   be-PAST.3SG throw PAST.PART

   lit. '[It] was thrown with.the.stone/s'.

According to Andrews 2001, the boldfaced element in the sentences of (28) is the SyntSubj, since its behavior shows at least 13 features that it shares with the behavior of the "canonical" SyntSubjs of Icelandic: it controls coreference with the "subject" of an infinitive and the choice of the reflexive possessive pronoun *sinni* 'self's' (Rus. *svoj*), it can appear between an auxiliary and the past participle of the lexical verb (where only SyntSubj are admitted), etc. However, "not only are they not nominative in case, but the verb does not agree with them" (Andrews 2001: 93), while normal SyntSubjs in Icelandic control the agreement of the MV and are marked by the nominative. Therefore, I conclude that these suspicious clause elements are not SyntSubjs—even though they behave in many respects as prototypical SyntSubjs sometimes do under specific conditions. Otherwise, it is not clear what Andrews and many others who share his perspective on this issue understand by a subject: by all means, not a clause element that is the depending member of a particular SSynt-relation.

In reality, the sentences in (28a–b) and (28c-ii) manifest an impersonal construction with a zero subject: in (28a–b), this is the lexeme $\varnothing_{3SG}$ «ELEMENTS», denoting some slightly mysterious natural forces; in (28c-ii), this is the zero dummy subject lexeme $\varnothing_{3SG}$, which is semantically and phonologically empty. These zeroes are equivalent to Eng. IT, Ger. ES and *Fr.* IL. (Spanish and Russian also have, in such contexts, a zero dummy: for instance, Sp. *Se lee muchas novelas* lit. '[It] reads itself many novels' and Rus. *Zdes´ mnoj siženo* 'Here by.me [it is] sat'.) The correct glossing of (28a), (28b) and (28c-ii) would be 'It drifted the boat/s to shore', 'It capsized the boat/s' and 'It was thrown with the stone/s'. That is exactly how all these

constructions are described in an elementary manual of Icelandic for non-natives (Glendening 1983: 49–50).

In Amele (Roberts 1987, 2001), the MV can simultaneously agree with four types of actant (quadri-personal agreement). Agreement affixes are different for each type of actant; the agreement of the only one actant with an intransitive MV allows us to establish the Subject Agreement affix set and thus to identify the SyntSubj without problems: the SyntSubj in an Amele sentence is the noun that imposes the use of these particular agreement affixes.

A problem concerning the SyntSubj in Amele comes from the inflectional category of switch-reference: in case a sentence includes two (or more) verbs, the first being subordinated to the second ($V^1 \leftarrow$**synt**$–V^2$), a switch-reference grammeme on V^1 is supposed to indicate whether V^2 has a SyntSubj referentially identical to that of V^1. (For instance, in *John came in and sat down* both verbs have referentially identical SyntSubjs; in *John came, and I sat down* the verbs have referentially different SyntSubjs.) Cf. (29a):

(29) Amele (Trans-New Guinea family; Roberts 1987, 1988, 2001; *q* = /ɡ͡b/)

a. *Ija* *hu* +*f*   +*ig*   *mad*+*ig* +*en*   'If I come, [I] will.speak'.

   I   come if-**SAME-SUB** 1SG$_{SUB}$ speak 1SG$_{SUB}$ FUT

vs.

   Uqa *ho* +*oʔ*   +*b*   *fi*   *ija* *mad*+*ig* +*en*   'If he comes, I will.speak'.

   he   come if-**DIF-SUB** 3SG$_{SUB}$ if I   speak 1SG$_{SUB}$ FUT

b. *Ege* *wen*   *Ø* +*g*   +*en*   'We became hungry'. =

   we   hunger   give 1PL$_{OBJ}$ 3SG$_{SUB}$.REMOTE.PAST   lit. '[It] us hunger gave'.

c. *Ege* *ʔo*   +*ʔob*+*ob*   *wen*   *Ø* +*g*   +*en*

   we   REAL-GER walk 1PL$_{SUB}$.**SAME-SUB**   hunger give 1PL$_{OBJ}$ 3SG$_{SUB}$.REMOTE.PAST

   lit. 'We walking, [it] us hunger.gave'. = 'As we walked, we became hungry'.

d. *Eu jagel November na* *uqa* *odo*+*ʔo* +*b*   *ʔul*+*ig* +*en*

   this month   in he   do **DIF-SUB** 3SG$_{SUB}$ leave 1PL$_{OBJ}$ 3SG$_{SUB}$.REMOTE.PAST

   lit. 'This in November he having.done, left.it.to.us.he' [Roberts 1987: 304, (620)].

 Amele has an impersonal construction, which expresses physiological and psychological states of a person; this construction has a dummy zero subject $Ø_{(3, sg)}$ with which the verb agrees; this is shown by the agreement grammeme 3SG$_{SUB}$ on the light verb 'give' in (29b). The Experiencer (= the hungry people) appears as DirO (also identified by verb agreement), and the designation of the state itself—a noun or an adjective—is a quasi-object, very much like quasi-object noun in Persian verbal collocations; it is not cross-referenced on the verb. What is found in (29b) is an unproblematic construction similar to Russian impersonal constructions of the type *Nas*$_{DirO}$ *trjasët* lit. '[It] shakes us'. = 'We are shaking' or *Nas*$_{DirO}$ *klonit v son* lit. '[It] pushes us into sleep'. = 'We are sleepy', with a dummy zero SyntSubj. But in a two-clause sentence, such as that in (29c), the verb of the first clause, where the SyntSubj is EGE 'we', is marked as having the same SyntSubj as the verb of the second clause, while this latter has a dummy zero subject. This fact makes Roberts to remark that, although 'we' in the second clause is a DirO, it possesses some SyntSubj properties, in the

first place—controlling the feature "same/different subjects" (Roberts 2001: 204). But why do we have to say that the suffix **-ob** signals the same **SyntSubj** in the next clause? Roberts himself (1988) states that the switch-reference in Amele may track the sequence of **Themes** ("same Theme/different Theme") rather than that of SyntSubjs. The detailed examples given in Roberts 1987: 292–305 also points to rather semantic character of Amele switch-reference: thus, in (29d), the SyntSubj is, of course, the same, but the marker of DIF-SUB signals the change of world setting—a new situation obtains. Therefore, if we accept that switch-reference in Amele marks the preservation/change of Themes (or maybe of situations described?), the problem disappears: it suffices to replace the names of grammemes SAME-SUB and DIF-SUB in (29c–d) by SAME-THEME and DIF-THEME.

6 The Syntactic Subject: Its Synt-role *vs*. Its Sem- and Comm-roles

The problem of defining SyntSubj has arisen in part as a result of the failure to separate, on the one hand, the purely syntactic properties that define a syntactic element of a clause, and, on the other hand, some semantic and communicative properties of that same element. It is true that syntactic clause elements encode—in the ultimate analysis—semantic roles of the corresponding meanings and are controlled by communicative factors. This, however, is not a reason for abandoning syntactic relations—and, by all means, this is impossible. Simply in some languages the alignment of syntactic relations to semantic roles is very intricate; thus, in English, a SyntSubj can correspond to a large variety of semantic roles. But in other languages such alignment is more straightforward: thus, in a basic clause of Archi the SyntSubj cannot be an Agent, an Experiencer or a Cause. But even if in some cases there is a one-to-one correspondence between syntactic and semantic roles, this should not lead us to confusing them. Thus, speaking of Lushootseed (Salishan family), Beck (2000: 310) states "that although there is an unusually close fit between the semantic structure of an utterance and the syntactic role that each participant … is assigned by the grammar, this fit is not one-hundred percent and so the invocation of a syntactic role … seems justified." This close fit is not at all astonishing: the SyntSubj as the most privileged syntactic actant tends to express the most privileged semantic role of Agent and the most privileged communicative role available to a nominal—that of the Theme, which in its turn, tends to be Given, referential and definite.

Acknowledgments

Notes

ⁱ **The passive in Dyirbal**. Dixon 1972: 65–67 speaks simply of «the **-ŋay** form» and «the **-ŋay** construction»: 40 years ago, no theoretical tools were available to properly describe the phenomenon. But here are his own examples (the SyntSubj, called "pivot" by Dixon, is boxed):

(i) **a.** $\boxed{Bayi\ bargan{+}\varnothing}_{X^{\text{™}}\text{I}}$ *baŋgul*   *yaṛa+ŋgu*$_{Y^{\text{™}}\text{II}}$ *ḍurga+$\varnothing$* *+n*
            the   wallaby NOM   the   man INSTR   spear   **ACT** PRES/PAST
            'The man is spearing the wallaby'.

  b. $\boxed{Bayi\quad yaṛa{+}\varnothing}_{Y^{\text{™}}\text{I}}$ (*baŋun*   *bargan+du*$_{X^{\text{™}}\text{II}}$)   *ḍurga+**ŋa*** *+ɲu*
            the   man   NOM   the   wallaby DAT   spear   **PASS** PRES/PAST
            'The man is spearing (at the wallaby)'.

This is an obvious diathesis modification marked on the verb—that is, a voice. True, terminologically, it is not OK to call it "passive"—because of semantic connotations of the term *passive*, since the Dyirbal verb aquires in this diathesis an "active" meaning; many linguists, A.E. Kibrik among them, call it "antipassive." Formally, this voice marks the following diathetic modification:

X	Y
I	**II**

$\Rightarrow$

X	Y
II	**I**

It also turns a transitive verb into an intransitive one: in (i-**a**), the tense suffix **-n** is that of transitive verbs, while in (i-**b**), its counterpart, -**ɲu**, is used only with intransitives.

   This is, of course, a classic passive schema, not some "antipassive." What is "anti-" here is Dyirbal itself: being ergative, it is "anti-nominative," in that all its verbs are semantically oriented in a way that is a mirror image of our verbs. 'X spears Y' corresponds in Dyirbal to 'Y undergoes spearing (by X)'. If the term *passive* jars as applied to (i-**b**), the terms *direct voice* and *converse voice* could be used; in nominative languages, they appear as *active* vs. *passive*. (For a review of "antipassive" constructions in various language types, see Cooreman 1994.)

ⁱⁱ Of course, Russian also has some problematic SyntSubjs, for instance:

– In the sentence *Èto*$_{SG}$ *byli*$_{PL}$ *moi druz'ja*$_{PL}$ 'This were my friends' the copula agrees not with the SyntSubj ÈTO 'this', but with the nominal attribute.

– The sentence *Mne*$_{DAT}$ *xočetsja pokoj+a*$_{GEN}$ ≈ 'I want some peace' = lit. '[It] wants.itself to.me of.peace' does not have an overt SyntSubj, but manifests a dummy zero SyntSubj; the same is true for *Im*$_{DAT}$ *èt+ogo*$_{GEN}$ *xvataet* lit. 'To.them of.this [it] suffices'. = 'This is sufficient for them'. These are impersonal constructions.

– A number of verbs (usually with the prefix NA-) allow for the SyntSubj in the genitive:

(i) *Naexali sjuda* $\boxed{vsjakie_{NOM}}$ lit. 'Came here anybodys'. = 'God knows who came here en masse'. ~

   Naexalo sjuda $\boxed{vsjakix_{GEN}}$ [*idem*, but more colloquial and more depreciative with respect to the Actor].

ⁱⁱⁱ The name of the Tongan absolutive should not be confounded with the name *absolutive* often given to the nominative case in languages with the ergative construction: the Tongan absolutive is formally different from the nominative. Note, however, that the Tongan absolutive optionally alternates with the nominative in full referential NPs:

(i) *'Oku 'alu '**a**$_{ABS}$ e tamasi.* ~ *'Oku 'alu $\varnothing_{NOM}$ e tamasi* lit. 'Is leaving the boy'.

^{iv} **The "passive" in Archi.** It is to some extent similar to the "passive" of Dyirbal, Note 1; see Kibrik 1975 and 2003: 352–354. Ja. Testelec (1979) was probably the first to insist on the voice-like character of this verbal "alternation" and draw a parallel with Dyirbal. The passive, or converse, voice in Archi has two characteristic properties:

   • As in several other Daghestanian languages, this voice is possible only in the imperfective—durative, habitual, progressive or frequentative—aspect.

   • In this voice, the MV receives a DirO in the nominative, which is a kind of anathema for an ergative language; moreover, the MV agrees with this DirO—along with the SyntSubj, so that the MV becomes bipersonal.

^v **The Category of Transitivization.** DETRANS(itivizer) is a grammeme of transitivization, an inflectional category of the verb similar to, but different from, voice. It resembles voice in that it impacts the verb central actants, the SyntSubj and the DirO; it differs from voice in that it does not permute the DSyntAs of the verb with the respect to its SemAs, but only modifies their surface realization (see Mel'čuk 2006: 231*ff*). This category includes at

least three grammemes: NEUTRAL ~ DETRANS ~ TRANS(itivizer). Tongan, examples 0)–0), features the pair NEUTER ~ TRANS; Chukchi has even two detransitivizers: DETRANS-1 and DETRANS-2. DETRANS-1, expressed by the prefix **ine-/ena-**, lowers the Synt-rank of the DirO (which becomes an IndirO); DETRANS-2 (the suffix **-tku/-tko**) not only lowers the Synt-rank of the DirO, but it also makes its appearance in the clause undesirable and, at the same time, blocks the expression of all other objects and complements, which are allowed both with the basic form and with the DETRANS-1 form.

(i) Chukchi

a. *Γəm+nan   tə +ret   +ərkən+Ø   kimitʕ+ən   (tomɣ+etə)*
I   INSTR   1SG$_{SUB}$ transport   PRES   3SG$_{OBJ}$ load   SG.NOM   friend   SG/PL.DAT
'I$_{X™I}$ transport a.load$_{Y™II}$ (to.a.friend/to.friends$_{Z™III}$)'.

b. *Γəm +Ø   t +**ine** +ret   +rkən   kimitʕ+e   (tomɣ+etə)*
I   NOM   1SG$_{SUB}$ **DETRANS-1** transport   PRES   load   SG.INSTR   friend   SG/PL.DAT
'I$_{X™I}$ transport a.load$_{Y™II}$ (to.a.friend/to.friends$_{Z™III}$)'.

c. *Γəm +Ø   tə +ret   +**ətku** + rkən   (ˀkimitʕ+e   ˀtomɣ+etə)*
I   NOM   1SG$_{SUB}$ transport   **DETRANS-2** PRES   load   SG.INSTR   friend   SG/PL.DAT
'I$_{X™I}$ transport (a.load$_{Y™II}$) (to.a.friend/to.friends$_{Z™III}$)'.

(i-a) presents an ergative construction, obligatory in Chukchi for any transitive verb. In (i-b), we find a nominative construction, possible only for an intransitive verb: the SyntSubj, which remains 'I', is in the nominative; the DirO '[a] load' has become an OblO in the instrumental, thus losing its salience; the two OblOs are optional. Finally, (i-c) is again a nominative construction: the two OblOs—'load' and 'friends'—are incompatible with each other and even less salient than in the preceding sentence; their omission is preferred.

Roughly, sentence (i-a) answers the question 'What are you transporting and to whom?', (i-b), the question 'What are you doing?', and (i-c), the question 'What is your occupation?'

Degrees of transitivization/detransitivization, related to the degree of the impact of the denoted action upon the object, are not a rarity; here is another example—from Warlpiri (Australian family):

(ii) Warlpiri

a. *Maliki+ḷi   ka+Ø   +Ø   ŋarka+Ø   yaḷki+ṇi*   'The dog is biting the man'.
dog   ERG   PRES 3SG$_{SUB}$.3SG$_{OBJ}$ NEUTRAL   man   NOM   bite   NON-PAST

b. *Maliki+ḷi   ka +ḷa   +žinta   ŋarka+ku   yaḷki+ṇi*   'The dog is biting at the man'.
dog   ERG   PRES 3SG$_{SUB}$ 3SG$_{OBJ}$ DETRANS   man   DAT   bite   NON-PAST

In Warlpiri, DETRANS lowers the transitivity of the verb, turning its DirO into an IndirO; but the verb remains transitive: it still presents an ergative construction, with the SyntSubj in the ergative case.

References

Aikhenvald, Alexandra, Dixon, Robert M.W., & Onishi, Masayuki, eds. 2001. *Non-Canonical Marking of Subjects and Objects.* Amsterdam/Philadelphia: John Benjamins.

Andrews, Avery. 2001. Noncanonical A/S Marking in Icelandic. In: Aikhenvald *et al*., eds. 2001: 85–111.

Beck, David. 2000. Semantic Agents, Syntactic Subjects, and Discourse Topics: How to Locate Lushootseed Sentences in Space and Time. *Studies in Language*, 24: 2, 277–317.

Bhaskararao, Peri & Subbarao, Venkata, eds. 2004. *Non-nominative Subjects*, vols.1–2. Amsterdam/Philadelphia: John Benjamins

Cole, Peter & Sadock, Jerrold, eds. 1977. *Grammatical Relations* [*Syntax and Semantics 8*]. New York etc.: Academic Press.

Cooreman, Ann. 1994. A Functional Typology of Antipassives. In: Fox, Barbara & Hopper, Paul, eds., *Voice: Form and Function*, Amsterdam/Philadelphia: John Benjamins, 49–87.

Croft, William. 1994. The Semantics of Subjecthood. In: Yaguello, ed. 1994: 29–75.

Davies, William & Dubinsky, Stanley, eds. (2001). *Objects and Other Subjects. Grammatical Functions, Functional Categories and Configurationality*. Dordrecht etc.: Kluwer.

Dixon, Robert. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Durie, Mark. 1985. *A Grammar of Acehnese on the Basis of a Dialect of North Aceh*. Dordrecht/ Cinnaminson: Foris.

Durie, Mark. 1987. Grammatical Relations in Acehnese. *Studies in Language*, 11: 2, 365–399.

Durie, Mark. 1988. Preferred Argument Structure in an Active Language. *Lingua*, 74, 1–25.

Falk, Yehuda. 2006. *Subjects and Universal Grammar*. Cambridge: Cambridge University Press.

Iordanskaja, Lidija & Mel'čuk, Igor. 2009. Establishing an Inventory of Surface-Syntactic Relations: Valence-controlled Surface Dependents of the Verb in French. In: Polguère & Mel'čuk, eds. 2009: 151–234.

Keenan, Edward. 1976. Towards a Universal Definition of "Subject." In: Li, ed. 1976: 305–333.

Keenan, Edward & Comrie, Bernard. 1977. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, 8: 1, 63–99.

Kibrik, Aleksandr. 1975. Nominativnaja/èrgativnaja konstrukcii i logičeskoe udarenie v arčinskom jazyke [Nominative/Ergative Construction and Logical Accent in Arči]. In: *Issledovanija po strukturnoj i prikladnoj lingvistike*, Moskva: Izdatel´stvo Moskovskogo universiteta, 54–62.

Kibrik, Aleksandr. 1977. *Opyt strukturnogo opisanija arčinskogo jazyka. III. Dinamičeskaja grammatika*. Moskva: Izdatel´stvo Moskovskogo universiteta.

Kibrik, Aleksandr. 1997. Beyond Subject and Object: Towards a Comprehensive Relational Typology. *Linguistic Typology*, 1: 279–346.

Kibrik, Aleksandr. 2001. Grammatical Relations. In: Smelser, N. & Baltes, P., eds., *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam: Pergamon Press, 6342–6348.

Kibrik, Aleksandr. 2003. *Konstanty i peremennye jazyka* [*Constants and Variables in Language*]. Sankt-Peterburg: Aletejja.

Kozinskij, Icxak. 1983. *O kategorii "podležaščee" v russkom jazyke* ["*Syntactic Subject" Category in Russian*], *Predvaritel´nye publikacii Problemnoj gruppy po èksperimental´noj i prikladnoj lingvistike*, vypusk 156. Moskva: Institut Russkogo jazyka AN SSSR.

Li, Charles, ed. 1976. *Subject and Topic*. New York etc.: Academic Press.

Mel'čuk, Igor. 1988. *Dependency Syntax*. Albany, N. Y.: SUNY Press.

Mel'čuk, Igor. 2004. Actants in Semantics and Syntax I/II. *Linguistics*, 42: 1, 1–66, 42(2): 247–291.

Mel'čuk, Igor. 2006. *Aspects of a Theory of Morphology*. Berlin: Mouton de Gruyter.

Mel'čuk, Igor. 2009. Dependency in Natural Language. In: Polguère & Mel'čuk, eds. 2009: 1–110.

Otsuka, Yuko. 2000. *Ergativity in Tongan* [PhD Thesis]. Oxford: University of Oxford. See http://www2.hawaii.edu/~yotsuka

Otsuka, Yuko. 2010. DP Ellipsis in Tongan: Is Syntactic Ergativity real? *Natural Language and Linguistic Theory*, 28: 315–342.

Polguère, Alain & Mel'čuk, Igor, eds. 2009. *Dependency in Linguistic Description*. Amsterdam/ Philadelphia: John Benjamins.

Rebuschi, Georges. 1978. Cas et fonction sujet en basque. *Verbum*, 1: 1, 69–98.

Rebuschi, Georges. 1981. Quelques problèmes de syntaxe anglaise pour la grammaire dite "relationnelle". *Verbum*, 4: 1, 85–119.

Rebuschi, Georges. 1982. *Structure de l'énoncé en basque*. [Collection ERA 642, numéro spécial.] Paris: Université Paris 7.

Rebuschi, Georges. 1986. Diathèse et (non-)configuralité: l'exemple du basque. *Actance* 2, 175–208.

Roberts, John. 1987. *Amele*. London etc.: Croom Helm.

Roberts, John. 1988. Switch-reference in Papuan Languages: A Syntactic or Extrasyntactic Device? *Australian Journal of Linguistics*, 8: 1, 75–117.

Roberts, John. 2001. Impersonal Constructions in Amele. In: Aikhenvald *et al.*, eds. 2001: 201–250.

Schwartz, Arthur. 1976. On the Universality of Subjects: The Ilocano Case. In: Li, ed. 1976: 519–543.

Suihkonen, Pirkko, Comrie, Bernard, & Solovyev, Valery, eds. 2012. *Argument Structure and Grammatical Relations. A Crosslinguistic Typology*. Amsterdam/Philadelphia: John Benjamins.

Tchekhoff, Claude. 1979. *La construction ergative en avar et en tonguien*. Paris: Klincksieck.

Testelec, Jakov. 1979. Ob odnom javlenii diatezy v dagestanskix jazykax [On a Diathesis Phenomenon in Daghestanian Languages]. In: *Materialy IX mežvuzovskoj studenčeskoj konferencii po strukturnoj i prikladnoj lingvistike*, Moskva: Izdatel´stvo Moskovskogo universiteta, 50–53.

Testelec, Jakov. 2001. *Vvedenie v obščij sintaksis* [*Introduction to General Syntax*]. Moskva: Izdatel´stvo RGGU.

Trương Van Chình. 1970. *Structure de la langue vietnamienne*. Paris: Paul Geuthner.

Van Valin, Robert. 1981. Grammatical Relations in Ergative languages. *Studies in Language*, 5: 3, 361–394.

Zimmerling, Anton. 2012. Nekanoničeskie podležaščie v russkom jazyke [Non-Canonical Subjects in Russian]. In: Voejkova, Maria, ed., *Ot značenija k forme, ot formy k značeniju: sbornik statej v čest´ 80-letija A.V. Bondarko*, Moskva: Jazyki slavjanskix kul´tur, 568–590.

Pri vsem X-e: a Corpus Study of a Russian Syntactic Idiom with Concessive Meaning

Valentina Apresjan

The National Research University Higher School of Economics
Russia, Moscow 101000, ul. Myasnitskaya, dom 20
vapresyan@hse.ru
Vinogradov Russian Language Institute
Russia, Moscow 119019, ul. Volkhonka 18/2
valentina.apresjan@gmail.com

Abstract

The paper presents a corpus study of the concessive syntactic idiom *pri vsem X-e* 'with all X' in Russian. The study demonstrates (a) a strong correlation between the semantics of the idiom and its other linguistic properties; (b) pragmatic properties that are typical of syntactic idioms in general; (c) language-specific idiomatic status. In particular, the combination of concession and intensification in its meaning explains its status as a negative polarity item; the fact that it refers to properties of objects explains its co-referential requirements. Pragmatically, like many other syntactic idioms, *pri vsem X-e* entails a certain structured worldview. Its linguistic properties are not carried across languages: its closest English counterpart, the construction *with all X* does not manifest any of its properties. This proves that the status of a syntactic idiom and all its linguistic consequences in one language cannot be predicted on the basis of the data from another language, and thus ought to be established on an individual basis.

Keywords

Semantics, syntax, pragmatics, syntactic idiom, gradable property, negative polarity, anaphor, parenthesis, scalarity, entailment.

1 Introduction

The paper presents a corpus study of the Russian syntactic idiom *pri vsem X-e* which has concessive meaning. Concessive semantics is expressed in Russian by a variety of grammatical and lexical means (conjunctions – *xotja* 'although', prepositions – *nesmotrja na* 'despite', particles – *tem ne menee* 'nevertheless', verbs – *ustupat'* 'to concede', etc.), as well

as by means situated "between lexicon and syntax", namely, syntactic idioms, or as they are sometimes called, syntactic phrasemes. The approach to syntactic idioms in this paper is based on the treatment of this phenomenon in Meaning-Text theory (Mel'čuk 1995, Iordanskaja & Mel'čuk 2007, Iomdin 2010), as well as in Construction Grammar (Fillmore at al. 1988). For the purposes of the present study, the main properties characterizing a syntactic construction as an idiom are its non-compositionality, fixedness of its function component parts, and limited variability of its content component parts, with lexical constraints on the filling of the variables. The paper considers the semantic, syntactic and pragmatic properties of the *pri vsem X-e* construction, establishes its idiomatic status in contrast with its non-idiomatic English counterpart *with all X,* postulates connections between its semantics and syntax, and formulates certain pragmatic properties typical of different syntactic idioms. Corpus approach[1] facilitates cross-linguistic comparison, as well as provides statistic foundation for the suggested analysis.

2 Semantics

First of all, the idiomatic status of the Russian construction *pri vsem X-e* needs to be established. The main criterion for a syntactic idiom is its non-compositionality (Fillmore et al. 1988). However, according to (Iomdin 2010: 144), "non-standard constructions" have to satisfy an additional requirement in order to qualify for "syntactic phrasemes", namely, lexical constraints. Either the construction has to contain inflexible lexical elements, that are more than a single function word, or there must be semantic constraints on the filling of free variables. The idiom *pri vsem X-e* satisfies all the conditions. It is non-compositional, i.e. its meaning is not the sum of the meanings of *pri* 'with' construction*, ves'* 'all' quantifier and the noun; two of its elements are lexically bound *(pri* and *ves');* the variable noun X is restricted to a certain semantic set.

2.1 Non-Compositionality

The free construction with the preposition *pri* can have multiple meanings. In the meaning closest to that of the idiom *pri vsem X-e,* the PP with *pri* expresses either a causal or a concessive connection with the predication of the main clause:

(1) *Pri takom bol'šom assortimente odeždy vybrat' platj'e budet netrudno*
    'With such a large collection of clothes, it will be easy to choose a dress' [causal connection]

(2) *Pri takom bol'šom assortimente odeždy ej ne udalos' vybrat' platj'e*
    'With such a large collection of clothes, she was unable to choose a dress' [concessive connection]

The causal vs. concessive interpretation of *pri X-e* phrases is dependent entirely on the context. It would be natural to expect the idiom *pri vsem X-e* 'with all X' to possess a similar

---

[1] The study avails itself of the data from the Russian National Corpus and the Corpus of Contemporary American

semantic ambivalence, even more so because it appears semantically very close to the intensifier construction *pri takom X-e* 'with such X'. However, that is not the case, as the construction *pri vsem X-e* has only concessive interpretations. Consider the following sentences, where (3) with concessive interpretation is grammatical, but (4) with causal interpretation is not:

(3) *Pri vsex svoix talantax, on ne smog sdelat' karjeru*
'With all his talents, he failed to make a career'

(4) **Pri vsex svoix talantax, on sdelal blestjaščuju karjeru*
'With all his talents, he made a brilliant career'

All instances of parenthetical *pri vsem X-e* construction in the Russian National Corpus have concessive interpretation:

(5) *Pri vsem staranii rukovoditelej Instituta èksperimental'naja baza stareet* (S.Alekseev)
lit. 'With all efforts of-heads of-institute the experimental base ages'
'For all the efforts of the Institute leaders, the experimental base is growing obsolete'

(6) *Bannikov pri vsej sile, xitrosti, iskušennosti v intrigax imel slabost'* (V.Valeeva)
lit. 'Bannikov with all strength, cunning, sophistication in intrigues had a weakness'
'For all his strength, cunning and experience in plotting, Bannikov had a weakness'

Its exclusively concessive meaning confirms the status of *pri vsem X-e* as a non-compositional item, since, as demonstrated above, its closest non-idiomatic correlate, emphatic construction *pri takom X-e* possesses two equally probable interpretations.

In this respect, the Russian construction *pri vsem X-e* is different from its English counterpart *with all X,* which equally freely allows both causal and concessive interpretations; consider the following examples from the Corpus of Contemporary American:

Causal:

(7) *With all these options, it won't be hard to find the perfect pair of jeans*

(8) *With all this information, you can soon learn to visualize mountains and areas*

Concessive:

(9) *Uncle Cy with all his flaws was the closest thing he had to a father*

(10) *With all this money, they never seem to clean the place*

2.2   Negative Polarity

One more distinction of the *pri vsem X-e* construction which argues its non-compositionality and its idiomatic status, is its negative polarity. Though by no means a classical negative polarity item, *pri vsem X-e* tends to occur in explicitly or at least implicitly negative sentences:

(11) *Počemu, tovarišči, my, pri vsem našem veličii, ničego takogo ne proizvodim?*
(V.Aksjonov)
'Why, comrades, we, with all our greatness, do not produce anything like that'
[negative element 'not']

(12) *Molodye vrjad li potjanutsja tuda, gde nevozmožno, pri vsej slave, obespečit'
sobstvennuju starost'* (N. Golovanova)
'Young people are unlikely to relocate to places where it is impossible, with all the fame, to provide for one's old age' [negative elements - 'unlikely', 'impossible']

(13) *Nikolaj Trofimovič pri vsem dobrom otnošenii ko mne i k kartine otkazalsja
vystavljat' na premiju* (E.Rjazanov)
'Nikolaj Trofimovič with all his good attitude to me and to my movie refused to nominate it for an award [negative element 'refused']

As these examples demonstrate, the main clause is likely to contain negative elements, such as 'not', 'difficult', 'impossible', 'unlikely' and the like.

The English counterpart of *pri vsem X-e,* the construction *with all X,* is not a negative polarity item:

(14) *With all this help, we soon finished the work* [causal interpretation, no negation]

2.3   Gradable Properties

Another property of the idiom *pri vsem X-e* which sets it apart from the free construction *pri X-e* and its English counterpart *with all X is* the lexical constraint on X. Lexical constraints of this idiom require the noun X to denote **a gradable property** (such as 'elegance', 'predictability', 'beauty', 'hatefulness') or a **complex of properties** (such as 'flaws', 'virtues', 'ambitions', 'difficulties'). The property X can be an attribute of an agent or a non-agent Y, who / which, in its turn, can have a property Z, perform an action Z or undergo an action Z:

The object Y has a property X and a property Z:

(15) *Pri vsej svoej odarjonnosti, on bezdel'nik*
'With all his brightness, he is an idler'

(16) *Pri vsej svoej deševizne, èti mašiny očen' nadjožny*
      'With all their cheapness, these cars are very reliable'

The object Y has a property X and does an action Z:

(17) *Pri vsej svoej odarjonnosti, on s trudom zakončil školu*
     lit. 'With all his brightness, he with difficulty graduated from school'
     'Bright though he is, he had a hard time graduating'

The object Y has a property X and undergoes an action Z:

(18) *Pri vsej ego odarjonnosti, ego vygnali iz školy*
     lit. 'With all his brightness, him expelled from school'
     'Bright though he is, he was expelled from school'

The meaning of the idiom can thus be formulated as follows:

(19) *pri vsem X-e, Y (is) Z* 'Object Y has property X to a high degree or object Y has many properties X; the speaker thinks that usually, if an object has property like X, it does not have property like Z, or cannot do action Z, or cannot undergo action Z; object Y has property Z, or does action Z, or undergoes action Z'.

Again, the English construction *with all X* is different *from pri vsem X-e* in that it does not require X to be a property:

(20) *With all these universities out there, he doesn't know where to apply*

This sentence would have been ungrammatical in Russian because the noun 'university' does not denote a property:

(21) **Pri vsex universitetax, on ne znaet, kuda postupat'*
     'With all universities, he doesn't know where to apply'

3 Syntax

Syntactically, the idiom *pri vsem X-e* 'with all X' resembles, at the first glance, an adjunct with causal, temporal, concessive, or conditional meaning. In Russian, such adjuncts are typically formed with prepositions *pri* 'with', 'in case of' (as in the idiom under consideration), *s* 'with', and *v* 'in' (modifiers are italicized):

(22) *Pri takih nalogax melkij biznes ne vyderžit*
     'With such taxes, small business won't survive'

(23) *S takoj figuroj ona možet stat' model'ju*
     'With such a figure, she can become a model'

(24) *V takix obstojatel'stvax trudno rassčityvat' na uspex*
    'In such circumstances it is difficult to count on success'

3.1 Parenthesis

However, the idiom *pri vsem X-e* manifests different syntactic properties than regular adjuncts with the preposition *pri*. First, it is necessarily parenthetical; cf. (25) but not (26):

(25) *Pri vsej podderžke gosudarstva, èkonomike prixoditsja nelegko*
    'With all the state support, the economy is going through difficult times'

(26) *Rabota vypolnena pri vsej podderžke Fonda*
    'The work has been carried out with all the support of the Foundation'

3.2 Sentential Position

Likewise, regular adjuncts *pri X-e* and the syntactic idiom *pri vsem X-e* have different sentential positions. Adjuncts tend to occupy either sentence-initial or sentence-final position, with the latter prevailing[2] according to the general tendency for SVO order in neutral sentences in Russian.

(27) *Pri vysokix temperaturax virus pogibajet*
    'At high temperatures the virus dies'

(28) *Virus pogibaet pri vysokix temperaturax*
    'The virus dies at high temperatures'

Only when *pri X* is topicalized, it can occur in midsentence (note the special contrastive prosody in this case):

(29) *Virus ↑pri vysokix ↑temperaturax pogibaet*
    lit. 'The virus at high temperatures dies'

The idiom, on the contrary, as a typical parenthetical clause, favors midsentence and sentence-initial positions, and avoids sentence-final position[3]:

(30) *Pri vsem ego bol'nom voobraženii, vrač on prekrasnyj*

(31) *Vrač on, pri vsem ego bol'nom voobraženii, prekrasnyj*

---

[2] The Russian National Corpus registers approximately seven-fold numerical prevalence of sentence-final *pri*-adjuncts.

[3] Sentence-final usages form only two percent of all the usages of *pri vsem X-e* idiom.

(32) $^?$*Vrač on prekrasnyj, pri vsem ego bol'nom voobraženii*
    'Despite having sick imagination, he is a wonderful doctor'

3.3 Anaphor

Next, unlike regular adjuncts, *pri vsem X-e* idiom exhibits anaphoric relations between either the subject or the object argument of the noun X and the syntactic subject or the object of the main clause:

Subject-Subject co-reference:

(33) *Pri vsex nedostatkax$_i$, ona čelovek$_i$ nadjožnyj*
    lit. 'With all shortcomings$_i$ she$_i$ is a reliable person'

Object-Subject co-reference:

(34) *Pri vsej pomošči$_i$, Maša$_i$ ne potjanet lečenie v častnoj klinike*
    'With all the help [to her$_i$], Masha$_i$ won't be able to afford treatment in a private clinic'

Sometimes word order plays a role in establishing the reference of *pri vsem X-e* idiom. Thus, the syntactic "co-reference with the subject" tendency can be semantically overridden in favor of co-reference with the object of the main clause in those cases where the clause with the idiom immediately follows the object, helping to establish anaphoric relations:

(35) *Logiku antikrizisnyx mer$_i$ pravitel'stva, pri vsej ix$_i$ zaputannosti, rossijane v celom ponimajut*
    lit. 'Logic of anti-crisis measures$_i$ of-government, with all their$_i$ unclarity, Russians generally understand'
    'The Russians generally understand the logic of the government's anti-crisis measures$_i$, despite their$_i$ unclarity'

The co-reference requirement characterizes the Russian idiom *pri vsem X-e,* but not its English counterpart 'with all X':

(36) *With all this fine raw seafood, I was surprised to find that my favorite appetizer of all was the tender, emphatic skewers of beef heart* (Corpus of Contemporary American)

The syntactic co-reference requirement for the *Russian pri vsem X-e* idiom is a consequence of its semantics, namely, of the fact that it most frequently denotes a property of an object, which comes in contradiction with its other properties or its behavior.

4 Pragmatics

4.1 Scalarity and Polarity

Pragmatically, the idiom *pri vsem X-e* entails a scale, where the object Y is characterized by the property X in a very high degree, and by a different property Z in a certain unspecified degree. Thus, it registers a certain level of abnormality in the object, according to the speaker's opinion, since it is characterized by properties of different domains, or of different polarities, one of them in a high degree.

What concerns the co-existing properties themselves, *pri vsem X-e* cannot describe objects possessing properties which are exact polar opposites, i.e. belong to the opposite poles of the same domain, such as *kind* and *wicked, mature* and *childish, sad* and *cheerful.* In this respect, it is different from the double conjunctions *i...i* 'and...and' or the adverb *v to že vremja* 'at the same time'[4]:

(37) **Pri vsej svoej dobrote, on zloj*
　　'With all his kindness, he is wicked'

(38) *On i dobryj, i zloj*
　　'He is both kind and wicked'

(39) *On dobryj i odnovremenno zloj*
　　'He is at the same time kind and wicked'

For obvious reasons, *pri vsem X-e* cannot introduce properties of the same polarity and the same domain (which is possible for coordinative conjunction *i* 'and'):

(40) *Pri vsej svoej nasmešlivosti, on ironičnyj*
　　'With all his sarcasm, he is ironical'

(41) *On nasmešlivyj i ironičnyj*
　　'He is sarcastic and ironical'

It can assign properties of the same polarity but different domains:

(42) *Pri vsem svojom intellekte, on očen' skromnyj čelovek*
　　'With all his intellect, he is a very modest person'

However, *pri vsem X-e* cannot ascribe properties that are totally unrelated (again, unlike coordinative conjunction *i* 'and'):

---

[4] However, some polar opposites, such as **tall and short, *wide and narrow, *fat and slim* cannot be conjoined under any circumstances, because it contradicts one's knowledge about the real world, where objects cannot simultaneously possess several observable characteristics that contradict one another.

(43) **Pri vsej svoej tolščine, on byl glupyj*
 'With all his bulk, he was stupid'

(44) *On byl tolstyj i glupyj*
 'He was fat and stupid'

Thus, *pri vsem X-e* requires the two co-existing properties to belong either to different polarities of close, but not coinciding domains, or belong to the same polarity in different domains. Yet this is a semantic requirement that in each case has to be "endorsed" pragmatically, i.e. the co-existence of those properties in one object should be possible, but unusual.

4.2 Evaluation and Anthropocentricity

One more pragmatic peculiarity of *pri vsem X-e* is that even neutral properties tend to gain positive or negative flavor when used in this syntactic idiom, and this positive or negative evaluation is the result of the anthropocentric perspective it entails. In this perspective, every property can be evaluated as either positive or negative, good or bad, convenient or inconvenient for people. Consider pragmatic awkwardness of sentences that resist evaluative interpretation:

(45) $^?$*Pri vsej svoej beskonečnosti, Vselennaja ne bezgranična*
 'With all its infinity, the Universe is not limitless'

Consider also the following pair of sentences with the idiom *pri vsem X-e,* where the same noun X is impossible in an objective non-evaluative context and becomes appropriate in an "anthropocentric" evaluative context:

(46) $^?$*Pri vsej svoej uzosti èta jubka očen' korotkaja*
 lit. 'With all its narrowness, this skirt is very short'
 'Though the skirt is narrow, it is very short'

(47) *Pri vsej uzosti svoix vzgljadov on čelovek vpolne zdravomysljaščij*
 lit. 'With all the narrowness of his views, he is quite a sensible man'
 'Though he is narrow-minded, he is quite sensible'

5 Conclusion

In the conclusion, this study proves that syntactic idioms are language-specific, and this should be reflected in their lexicographic treatment, as the presence of a correlate construction in another language does not necessarily signify the presence of a corresponding syntactic idiom.

Different syntactic idioms share certain properties; one of them is the tendency towards negative polarization. Another important property concerns pragmatics, namely entailment. Many syntactic idioms introduce scalar gradation, establish connections between situations or in other ways structure the representation of the world in the mind of the speaker.

Thus, *pri vsem X-e* introduces both a scale of properties and an idea of their connection with the behavior of the object they characterize; *let alone* introduces a scale of objects that possess certain properties to different extents; the Russian syntactic idiom *X-X, a Y Z (Kto-kto, a Vanja ne podvedjot* 'Who-who, but Vanya will not let one down' = 'Don't know about others, but John will not let one down') introduces a scale.

The presence of an entailment reflecting the speaker's opinion and assessment of the situation, that is typical of syntactic idioms, is likely the consequence of their idiomatic status. While regular syntactic means of expression present an "objectified" picture, syntactic idioms, like other phraseological means of expressions, reflect certain wisdoms, certain worldviews, even though in a much more abstract form than lexical idioms or proverbs.

Acknowledgements

Bibliography

Apresjan V. 2009. Concession in Russian: Semantics as a reflection of Rhetoric. In *Proceedings of the 4th International Conference on Meaning-Text Theory.* Montreal, June 24-26, 2009. http://mtt.upf.edu/mtt2009/ProceedingsMTT09.pdf 15-24.

Fillmore C.J., Kay P. & M. C. O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: the Case of Let Alone. *Language.* 502 (64, 3). 501-538.

Iomdin L. 2010. Konstrukcii malogo sintaksisa. In *Teoretičeskie problemy russkogo sintaksisa: vzaimodejstvie grammatiki i slovarja,* Apresjan Ju. (ed.). Jazyki slavjanskix kul'tur. Moscow. 59-280.

Iordanskaja L., Mel'čuk I. 2007. *Smysl i sočetajemost' v slovare.* Jazyki slavjanskix kul'tur. Moscow.

König E. 1991. Concessive Relations as the Dual of Causal Relations. *Semantic Universals and Universal Semantics.* D. Zafferer (ed.) Foris Publications. Berlin, New-York. 190-209.

Mel'čuk, I. 1995. Phrasemes in Language and Phraseology in Linguistics. In M. Everaert, E J. van der Linden, A. Schenk, and R. Schreuder (eds.), *Idioms: Structural and Psychological Perspectives.* Hillsdale, New Jersey, Hove, UK: Lawrence Erlbaum Associates. 167— 232.

Functional Macrocategory and Semiautomatic Inheritance of Semantic Features: a Methodology for Defining Nouns

María Auxiliadora Barrios Rodríguez

Facultad de Filología. Universidad Complutense de Madrid
Paraninfo s/n. 28040-Madrid
auxibarrios@filol.ucm.es

Abstract

The aim of this paper is to discuss a possible new methodology when defining concrete nouns such as *earring, ring* or *scarf* by their automatic inheritance of semantic features from their macro-categories, in this case *accessories*. The semantic analysis of the meaning by primes, molecules and lexical functions allows us to arrive at a set of definitions which are more coherent for one lexical field. This methodology could be helpful when building ontologies.

Keywords

Semantic analysis, ontology, definitions, lexical functions, semantic primes

1 Introduction

The aim of this paper is to discuss a possible new methodology when defining concrete nouns labeled semantically.

On the one hand, the principle of lexical inheritance (Mel'čuk & Wanner, 1996; Barrios, 2010) claims the inheritance of verbs by some groups of nouns labeled semantically. For instance, one of the meanings of the noun *accessory* combines with the verb *to wear*, so we assume that the nouns labeled as 'accessories', such as *earring, ring, scarf, hat, sunglasses*, also combine with the verb *to wear* (*to wear an earring, ring, scarf, hat, sunglasse*s, etc.). This principle was implemented in a database, known as *BADELE.3000*, which contains the 3,300 nouns most frequent in the Spanish spoken in Spain (Barrios & Bernardos, 2007). This linguistic resource was useful when building a geographic ontology (Barrios and Vilches, 2010).

The aim of this paper is to discuss the possibility of applying this principle not only to the collocations but also to the genus proximum; should the first word of the definition of *earring, ring, scarf, hat, sunglasses* be "accessory"? The semantic label (such as "accessory") as the first part of the definition is a usual practice within the Meaning-Text Theory (MTT)

framework. However, the Natural Semantic Metalanguage (NSM) proposal has a different perspective.

As Aristotle's tradition defends, the definition of taxonomic categories, such as *fish,* is independent of their hyponyms, such as *tuna, sardine* and *salmon*; whereas the definitions of these last nouns depend on the word *fish*. Indeed, *fish* can be defined as "a creature that lives in water and has a tail and fins", and *tuna, sardine* and *salmon* usually are defined including the hyperonym *fish* (*tuna:* a large *fish* that lives in warm water; *sardine*: a kind of small sea *fish*; *salmon*: a large silver-coloured *fish*). When building ontologies, concepts such as 'fish' and its hyponyms are defined in a correct way. The problem arises when defining some other collective nouns such as *accessories*. Some scholars would claim that we are dealing with two different types of words – the noun *fish* and all the natural class nouns are *taxonomic categories*, whilst the noun *accessories* and a large set of nouns describing objects are *collective super-categories concepts* (Wierzbicka, 1985). A third class, the functional macro-categories *furniture, vegetables* or *cosmetics* have not been found useful when defining words such as *table, carrot* or *lipstick*: the definition of *vegetables* includes a list of words ('vegetables are a set of things such as *carrots* or *peas*'); consequently, *carrots* or *peas* cannot include the meaning of *vegetables* (Goddard, 2009) which means that *carrot* should not be defined as 'an orange coloured *vegetable* that has a crisp texture'. Iomdin et al (2011) multilingual research proves that classes existing in several languages often overlap and include different items. The question is: how could these words be defined?

This article sets out to explore the extension of the principle of lexical inheritance regarding the semantic features of the functional macro-category 'accessory' (henceforth, macro-categories will appear between simple quotes). If people think that a *handbag* is an accessory, why is it that *handbag* cannot be combined with *to wear*? What are the common semantic components shared by nouns labeled as 'accessories'? Can the meaning of 'accessories' be considered as forming part of the meaning of *earring, ring* or *scarf*? Could this meaning be inherited when building the definitions of these nouns? This paper presents the possible meanings of *accessories* and some results after the implementation of the proposed methodology in a Spanish lexicon database.

2   Problem: the Complexity of Functional Macro-Categories

Wierzbicka distinguishes three types of macro-categories: a) **Taxonomic concepts** (*bird, dog, bed, bicycle*) stand for specific kinds of things and are based on similarity between the referents; b) **Functional concepts** (*pet, weed*) stand for things of any kind that have a specific kind of function and are based on the identity of function; and c) **Collective supercategory concepts** (*furniture, vegetable, fruit, clothes*) stand for collections of taxonomic concepts and are based on contiguity (Wierzbicka, 1985, 269). Goddard reintroduces this distinction in his proposal of the functional macro-categories (corresponding to the third group above) such as *vegetables*, *furniture* and *weapons*. Following his terminology, here *furniture, vegetable, fruit, clothes* and *accessories* are called **functional macro-categories**.

For artifacts, the function of the object denoted by the noun is seen as a component of the meaning (Apresjan, 2000). It has also been claimed that the definition of this kind of nouns should be formulated in terms *of "made for + an intended function"* (Wierzbicka, 1996). This function can even be viewed as the basis for the existence of a supra-category noun (called functional concepts above). Therefore, "toys and weapons are not taxonomic concepts at all,

because they do not designate kinds of things. Rather, they specify a kind of function" (Wierzbicka, 1985: 264).

This paper tries to shed light on whether not only nouns of artifacts and functional concepts, but also nouns of functional macro-categories, such as *accesorios* 'accessories', can be related to any specific function, to confirm the hypothesis of Goddard. Then, the focus of our study is the lexical field of accessories (related to clothes, shoes, jewelry), leaving aside uses of *accessory* as an adjective or as part of compound nouns *(car accessories, mobile accessories).* Evidence from the collocational phenomena suggests that the function of an accessory could be expressed by the collocation *to wear.* Should this be the case, we ought to consider that this function should be present in the meaning of every noun labeled as an accessory. However, the meaning of the noun *accesorio* is attached to some words related to the field we are interested in, that cannot be combined with the verb *llevar puesto* 'to wear'. In the following section, we will turn our attention to these peculiarities.

3   Unexpected Results During the Automatic Lexical Inheritance Process

The Spanish database called BADELE.3000 implemented the principle of lexical inheritance in order to prove the semantic motivation of most of the collocations. As a consequence of this implementation, 9,000 of a total of 20,700 collocations contained and formalized by means of lexical functions were automatically obtained in this resource. The database was also designed to collect unexpected results, keeping in mind the linguistic viewpoint of combinatory non-predictability.

When classifying Spanish words for clothes, shoes and accessories in this database, the author followed the verb patterns of three verbs: *ponerse* 'to put something on (oneself)', *llevar puesto* 'to wear something' and *quitarse* 'to take something off (oneself)'. Of the set of 66 nouns labeled as '*ropa y complementos*', 'clothes and accessories' (meaning in a reductive way 'clothes, shoes and accessories'): 23 were accessories, 9 were shoes, and 34 were 'clothes' (the last one included a subset, "underwear"). These labels were 3 of a total of 320 in a hierarchy of 9 semantic labels, and shared the same level (N4), as Figure 1 shows:

Figure 1: Hierarchy of Semantic Labels in the Spanish database BADELE.3000

After taking advantage of the inheritance properties of the aforementioned three verbs *(to put something on, to wear* and *to take something off),* 198 collocations (66 x 3) were automatically obtained. Each of these three labels, in turn, shared specific verbs, such as

sentar (una prenda de ropa) bien 'to suit [a piece of clothes] someone', *calzar (algún calzado)* 'to wear a shoe', considering that 'wear' is less restrictive than *calzar,* which is used only for wearing shoes; *lucir (un accesorio)* 'to sport an accessory', etc. These verbs were attached specifically to each one of the labels.

Figure 2 shows some verbs that combine with all the nouns labeled as 'clothes', such as *make, wear out, suit someone, be tight for someone, put something on:*

Figure 2: Inheritance of different verbs for the semantic label 'clothes'

As it was mentioned above, the database was designed with the proposal of verifying the predictions of the combinations. After the inheritance, 10 nouns labeled as 'accessories' were rejected because these nouns cannot be combined with *ponerse* 'to put something on', *llevar puesto* 'to wear' and *quitarse* 'to take something off': *agenda* 'diary', *cartera* 'briefcase', *bolsa* 'bag', *bolso* 'handbag', *mochila* 'backpack', *cartera* 'wallet', *monedero* 'purse', *maleta* 'suitcase', *paraguas* 'umbrella', *abanico* 'fan'. All these nouns are known in Spanish as *accesorio* 'accessory' or even *complemento* 'complement'.

Figure 3 shows some of the unexpected results:

Figure 3: Unexpected results of Verb Combinations for the semantic label 'accessory'

Thus the question arises: why can one put on a scarf, but not a handbag? Is there any semantic difference between these two accessories?

4 Analysis of the Semantic Features of Accessories and its Hyponyms

This section focuses on the analysis of semantic features in order to clarify the need for a systematic method of definition. Although the proposal of Apresjan et al (2008) based on Meaning Text-Theory (MTT) principles (Mel'čuk, 1996) is elegant and allows a multilingual approach, there are not enough tools to avoid circularity in the definitions; some other researchers build their dictionaries starting from the combinatorial phenomena and definitions are the last step. The semantic primes used in the Natural Semantic Metalanguage framework (Wierzbicka, 1996), however, try to define primes first (indecomposable words that can correspond to a similar concept in all languages), then molecules (words whose meanings can be described by primes). The difficulty with this approach lies in defining by a very simple set of 60 words called primes (see Appendix).

Our study aimed to understand the meaning of the functional macro-category 'accessories' and then to describe it by primes and molecules (marked with [m]). Our initial hypothesis was that there could be three different groups of 'accessories' (the three mentioned in paragraph 1 related to clothes, shoes, jewelry). As artifacts nouns are generally attached to a concrete function (this thing was made to X), we started describing some meanings as shown in (1):

(1) a. Something of one kind. This something was made [m] to be in touch with a small part of one finger [m]. This something was made [m] to look [m] good.

 b. Something of one kind. This something was made [m] to cover[m] the neck [m]. This something was made [m] to cover [m] a small part of the chest [m]. People wear it when it is cold [m].

 c. Something of one kind. This something was made [m] because people want to have many things inside this something. Women [m] have this something near them if they are not at home [m].

The meaning in a) corresponds to *ring;* b) to *scarf* and c) to *handbag.* These nouns represent three different groups of 'accessories'. Let's see why we proposed these three different groups: our first step was to describe the component of the meaning shared by these nouns, trying to elucidate whether there were semantic differences that could explain different possible syntactic patterns shown in Figure 3.

The semantic component shared by the three groups is to 'look good'. ***Accessories1*** and ***Accessories2*** share 'to wear', and ***Accessories2*** and ***Accessories3*** share the property of each of these things to have a particular function (such as "to protect the eyes from the sun's harmful rays", the sunglasses; "to know what time it is", the watch; etc.) From the perspective of practical knowledge of the world, the three groups are generally located in the same place in large department stores, because these objects are made for personal use and can be purchased as gifts. This could explain the polysemy of *accessories*.

		Pattern 1	Pattern 2	Pattern 3	Pattern 4
Accessories1	*ring, earring, necklace, bracelet*	It makes some part of a body 'look good' / this thing 'looks good'	People 'wear' it		It can be a gift. They are grouped and labeled in large department stores.
Accessories2	*scarf, sunglasses, watch, hat*	It makes people 'look good' / this thing 'looks good'	People 'wear' it	It has a 'function'	It can be a gift. They are grouped and labeled in large department stores.
Accessories3	*handbag, wallet, diary, briefcase*	It 'looks good'		It has a 'function'	It can be a gift. They are grouped and labeled in large department stores.

Table 1: Three different types of *accessories*: semantic and pragmatic features

However, the main focus of this study is to shed light on the meaning of the word *accessory*. We tried to define it by observation of the sets of referents and by introspection about the common features shared by the components of every set. Table 2 shows our proposal:

Accessories1	*ring, earring, necklace, bracelet*	Many things of many different kinds. These things were **made [m] to be in touch with** a part of the body. If people **wear [m]** one of these, they can think this part of the body **looks [m] good**.
Accessories2	*scarf, sunglasses, watch, hat*	Many things of many kinds. These things were **made [m]** because people want **to do something** with them. If people do not **wear [m]** these things they cannot do these somethings with these things. Many times these things look [m] good. If people wear [m] one of these things, they can think they **look [m] good**.
Accessories3	*handbag, wallet, diary, briefcase*	Many things of many kinds. These things were **made [m]** because people want **to do something** with these things. Many times these things **look [m] good**.

Table 2: Proposal of three different meanings of *accessories*

Spanish *accesorio* 'accessory' is a polysemous word with three different meanings. Explanations in Table 2 contain both primes and molecules; we use the expression 'to do something with this thing' because the meaning 'function' seems to be too complex to be expressed via primes and molecules.

5 Proposals

5.1 Proposal 1: Inheritance of Functional Macro-Category Semantic Features

Here we try to apply a systematic method of definition based on shared semantic features. The semantic features present in the definitions of Table 2, should be present in the definitions of the concrete nouns corresponding to each one of them. Table 3 shows the features to be inherited by the first set of nouns:

Meaning	Concrete nouns	Semantic features to be inherited
Accessories1	*ring, earring, necklace, bracelet*	1. **made [m] to be in touch with** (a part of the body) 2. **to wear [m] it** 3. (this part of the body [m]) **looks [m] good**

Table 3: Semantic features to be inherited by *ring, earring, necklace* and *bracelet*

Table 4 shows the semantic features of *accessories1* present in the meaning of *ring, earring, necklace* and *bracelet* after the semi-automatic inheritance. The meaning of nouns of *accessories1* (corresponding to jewelry and fashion jewelry) is quite homogenous: they share the semantic features: 'made to be in touch with', 'wear', 'part of the body' and 'look good'; differences between them are due to the different part of the body. The cause of this systematization could be that every one of these pieces was made for a similar proposal: to make something look good.

ring	This something was **made [m] to be in touch with a small part of one finger**. If people **wear** [m] this something, they can think their **hand** [m] **looks [m] good**.
earring	This something was **made [m] to be in touch with a small part of the ear**. If people **wear** [m] this something, they can think their **face** [m] **looks [m] good**.
necklace	This something was **made [m] to be in touch with the neck [m].** If people **wear** [m] this something, they can think their neck [m] **looks [m] good**.
bracelet	This something was **made [m] to be in touch with** a **small bottom part of the arm** [m]. If people **wear** [m] this something, they can think this part of the arm [m] **looks [m] good**.

Table 4: Proposal of definitions of some concrete nouns labeled as *accessories1*. Shared features: 'it was made to be in touch with', 'part of the body', 'wear', 'look good'.

Table 5 shows the semantic features of *accessories2* present in the meaning of *scarf, sunglasses, watch* and *hat*. There are more differences than similarities between these definitions; the meaning of nouns of *accessories2* is less homogenous: all these nouns only share the semantic feature 'wear' and the condition to wear it, so that it could fulfill its role:

scarf	This something was **made [m] to cover [m] the neck** [m]. This something was **made [m] to cover [m] a small** part of the chest [m]. People **wear** it when it is cold [m].
sunglasses	This something was **made [m] to cover [m] the eyes** [m] when the sun [m] is above [m]. When people **wear** it, they do not feel something bad in the eyes [m].
watch	This something was **made [m]** because people want **to know** now the time. When people **wear** it, this something covers the bottom [m] part of one arm [m].
hat	This something was **made [m] to cover [m] the head** [m]. When it is cold [m], if people **wear** it, they do not feel something bad in the head [m]. When the sun [m] is above], if people wear it, they do not feel something bad in the head [m].

Table 5: Proposal of definitions of some concrete nouns labeled as *accessories2*. Shared features: 'it was made for something specific' (a particular function), 'wear'

Table 6 shows the semantic features of *accessories3* present in the meaning of *handbag, wallet, diary* and *briefcase*. They do not share any semantic feature, each one was made to do something specific (a particular function):

handbag	This something was **made [m] because** people want **to have many things inside** this something. Women [m] have this something near the place they are if they are not at home [m].
wallet	This something was **made [m] because** people want **to have money** [m] **inside this thing**. This something was made small because people want to have this something near.
diary	This something was **made [m] because** people want **to write** [m] many things in this thing. These can be the things these people want to do during one day [m]. These can be the things these people want to do during one month [m]. These can be the things these people want to do during one year [m].
briefcase	This something was **made [m] because** people want to have **many things inside** this something. When people work [m], people have this something in the room [m].

Table 6: Proposal of definitions of some concrete nouns labeled as *accessories3*. Shared features: 'made for something specific' (particular function)

5.2 Proposal 2: Definition of *Accessories*

Definitions of the functional macro-category *accessory* (see Table 2) and concrete nouns (see Tables 3-5) are written with primes and molecules. This analysis is interesting from a semantic point of view, but it might not be very practical from a lexicographic point of view. However, these definitions allow us to rewrite the meanings in a traditional way: based on the

combinatory (to wear, to carry) we would propose to reduce the three meanings of *accessories* to two definitions useful for a traditional dictionary. The definitions could be described as shown in Table 7:

Accesorios₁ *(Accessories1,2* **of Table 2)**	Cosas de distinto tipo, como *joyas o piezas de bisutería, bufanda, gafas de sol, sombrero*, frecuentemente de bonita apariencia, que la gente se pone para estar más atractiva. (Many things of many kinds such as *some piece of jewelry, scarf, sunglasses, hat,* usually good-looking, that people wear [m] to look [m] good)
Accesorios₂ *(Accessories3* **of Table 2)**	Cosas de distinto tipo, como *bolso, monedero, agenda, cartera,* frecuentemente de bonita apariencia, que la gente lleva a mano. (Many things of many kinds, such as *handbag, wallet, diary,* usually good-looking, that people carry and have close at hand)

Table 7: Proposal of two different meanings of *accessories*

6   Discussion and Conclusions

While it may hold true that "in the past 'lexicography' meant practice, whereas 'semantics' meant theory" (Wierzbicka, 1985, 5), NLP applications call for a combination of both disciplines. This paper aims to find a semantic explanation that allows us to improve a lexicographic resource. Contrary to the common practice in lexicography, most of the functional macro-category nouns, such as *accessories*, should not be used as a superordinate in the definitions of the hyponyms. The reason is that these categorizations do not correspond to linguistic, but rather to extra-linguistic phenomena. A semantic analysis of every category should be developed in order to uncover all the semantic features belonging to each of these nouns. When the referents named and grouped by this category share a common function and there is linguistic evidence (such as the collocation *wear some accessories*), we can presume that the meaning of the functional macro-category nouns is present in the definition of their hyponyms, not necessarily as a superordinate but in some other way.

Accessories can be related to three specific functions: one is shared by the first group of nouns (*ring, earring, necklace, bracelet*), 'to wear it'; the second one is shared by the second group of nouns (*scarf, sunglasses, watch, hat*), 'to wear it if someone wants this thing to fulfill its function', the third one is shared by the third group of nouns (*handbag, wallet, diary, briefcase*), 'to carry this thing and to have it close at hand'. Verbs that express these functions (*to wear, to carry*) are values of the Lexical Function **Real**, which means 'to fulfill the requirement of something'. These verbs should appear not only in the definition of *accessories* but in the definitions of every concrete noun labeled as an *accessory*.

The semantic analysis of the meaning by primes, molecules and lexical functions allows us to arrive at a set of definitions which are more coherent for one lexical field. This methodology could be helpful when building ontologies. Some problems remain unresolved, such as the inheritance of the feature 'close at hand', expressed in different (and not so clear) ways (see also *diary* in Table 5, where this feature in not considered to be a part of the meaning because

it does not seem to be necessary for this concrete noun), or the missed feature 'personal' in *accessories₃* (too complex to be expressed by primes). The component 'if people do not wear [m] these things they cannot do this something with these things' can be absent in a watch: you can put your watch on the table and be watching the time. Despite this, semi-automatic inheritance of shared semantic features can be developed with databases such as BADELE.3000, when defining concrete nouns related to functional macro-categories.

Acknowledgements

Bibliography

Apresjan, J. 2000. *Systematic Lexicography.* Oxford: Oxford University Press.

Barrios, M. A. 2010. *El dominio de las funciones lexicas en el marco de la Teoria Sentido-Texto.* Barcelona: Estudios de Lingüística del español http://elies.rediris.es/elies30/

Barrios, M. A., & Vilches, L. 2010. "Is it possible to enrich ontologies with a specialized domain linguistic resource?" Breathnnach, U., de Barra-Cusack F. (eds.) *Terminology and Knowledge Engineering Conference. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges.* Fiontar. Dublin City University.

Barrios, M. A., & Bernardos, S. 2007. Badele.3000: An implementation of the lexical inheritance principle. *Wiener Slawistischer Almanach. Sonderband,* 69, 97–106.

Goddard, C. 2009. "*Vegetables, Furniture, Weapons: Functional macro-categories in the English Lexicon*". Melbourne: ALS.

Iomdin, B.L.; Piperski, A. Ch.; Russo, M.M.; Somin, A.A. 2011. How different languages categorize everyday items. *International Conference on Computational Linguistics. Dialogue 2011* http://www.dialog-21.ru/digests/dialog2011/materials/en/pdf/27.pdf

Mel'čuk, I., & Wanner, L. 1996. Lexical Functions and Lexical Inheritance for emotion lexemes in German. In L. Wanner, *Lexical Funtions and Lexicography and natural language processing*, 209–278. Amsterdam/ Philadelphia: John Benjamin.

Mel'čuk, I. 1996. Lexical Functions: a tool for the description of lexical relations in a lexicon. In L. Wanner, *Lexical functions in lexicography and natural language processing*, 37–102. Amsterdam / Philadelphia: John Benjamins.

Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis.* Ann Arbor: Karoma Publishers.

Wierzbicka, A. 1996. *Semantic Primes and Universals.* Oxford: Oxford University Press.

Adverbial Partials in Russian (*vdvoe* 'twice as much/half', *napolovinu* 'half' and others)

Igor Boguslavsky

Institute for Information Transmission Problems
of the Russian Academy of Sciences
Bolshoy Karetny per. 19, Moscow, 127994, Russia
igor.m.boguslavsky@gmail.com

Abstract

In Russian there is a large group of words that quantitatively describe an object by referring it to another object. Some of these words name a specific part of this other object, such as *polovina* 'half*', tret'* 'a third', *procent* 'per cent', etc. These words often have adverbial derivatives (which we call "adverbial partials") – *napolovinu* 'by half', *na tret'* 'by a third', *na N procentov* 'N%', *častično* 'partly', etc. Each word denoting a part has two valencies: "what constitutes a part", and "a part of what". We study how these valencies are filled and how the meaning of adverbial partials is incorporated in the meaning of the rest of the sentence. At the end of the paper we describe a curious semantic difference between the adverb *napolovinu* and the prefix *polu-* both meaning 'half'.

Keywords

Adverbial derivatives, quantifiers, *more, less, half, per cent.*

1 Partial Expressions in Russian

There is a group of words that denote a quantitative part of a whole (we will call them partial expressions, or partials), such as *polovina* 'half*', tret'* 'a third', *procent* 'per cent'. They are interesting as far as their valencies are concerned.

Partials have two valencies: "what constitutes a part" and "a part of what". Normally, only the second of them is expressed by a subordinated noun phrase: *četvert' arbuza* 'a quarter of a watermelon', *30% doxoda* '30% of the profit'. The first one is usually connected to the partial via a lexical functional verb:

(1) *Brak sostavljaet* [Oper$_1$(*procent*)] *60% partii*
      'Defective goods constitute 60% of the batch'.

(2) *Partija na 60% sostoit* [Labor21(*procent*)] *iz braka*
    'The batch is 60% defective'[1].

However, in some contexts the first actant can also be expressed by a subordinated genitive noun phrase: cf. *60% braka* '60% of defective goods' in (3).

(3) *Partija soderžit* [Oper2(*procent*)] *60% braka*
    'The batch contains 60% defective goods'.

There are cases when the valency of the whole is not marked morphologically or syntactically and it is even difficult to figure out what is part of what. For example, sentence (4) is ambiguous: either 17% of the Spanish exports of olives are supplied to Scandinavia, or 17% of olives consumed in Scandinavia are exported from Spain.

(4) *Ispanija postavljaet v Skandinaviju 17% maslin*
    'Spain supplies to Scandinavia 17% of olives'.

In this paper we are concerned with adverbial derivatives of partial words, such as *na chetvert'* '1/4', *napolovinu* 'half', *na (skol'ko-to) procentov* 'N%', and some closely related adverbs and adverbial expressions, such as *vdvoe* '2 times', *vtroe* '3 times', *v (skol'ko-to) raz* 'n times'. The question that we are interested in is how their meaning is incorporated in the meaning of the rest of the sentence. This question will be discussed in Section 2. In Section 3 we will describe a curious semantic difference between the adverb *napolovinu* and the prefix *polu-* both meaning 'half'. The difference is due to their different capacity to select a whole of which a half is taken.

2 Adverbial Partials in Different Contexts

We present three types of contexts in which adverbial partials (AP) can be used: the telic context, the context of comparison and the context of the predicate *sostoit iz* 'consists of' and its synonyms.

2.1 Telic Context

Consider sentences (5a–c) in which APs affect telic verbs or adjectives. All of them denote a situation or a property that has reached a natural endpoint.

(5a) *Zadača napolovinu rešena*
     'The problem is half-solved'
(5b) *Ty ubedil menja na 95%*
     'You convinced me 95%',
(5c) *Ja otčasti s vami soglasen*
     'I partly agree with you'.

Here the scope of AP, i.e. the whole of which only a part is realized, is the endpoint of the situation. (5a) means that the degree to which the situation "the problem is solved" takes place

---

[1]   We will return to this example below, in 2.3.

is half the degree it will have when the endpoint is reached. Such contexts also admit full degree adverbials, such as *polnost'ju* 'fully', *soveršenno* 'completely', or *sovsem* 'entirely', but do not admit adverbials of the type *vdvoe, vtroe, v 5 raz,* which are very typical of the next context type.

2.2   Context of Comparison

Examples of the context of comparison are sentences (6a)–(6c).

(6a) *Za dva mesjaca ceny vyrosli vtroe*
'In two months the prices tripled',
(6b) *Vaši resnicy stanut gušče na 60%*
'Your eyelashes will be 60% thicker',
(6c) *My na četvert' umen'šili rasxody na otoplenie*
'We cut a quarter of our heating expenses'.

In this class of examples, there are two values of the same parameter, and one of them is characterized by means of the reference to the other one. APs here define a value as equal to a certain part of another value, which serves as a reference point of comparison. (6a) means that after two months the prices became three times what they were at the start of this two months' period. (6b) asserts that the thickness degree of the eyelashes will be 60% more than the current degree. In (6c) heating expenses at different moments are compared, and one of them is also calculated on the basis of another one.

At some step of semantic decomposition, the comparison is always reduced to meanings "more" or "less". Therefore, one should first of all analyze combinations of AP with these meanings. In this connection, in subsection 2.2.1 we will discuss the semantic relationship between *more* and *less*, and their argument structure, and in 2.2.2 we will show how these meanings combine with the meaning of APs to form larger semantic units.

2.2.1   *Argument Structure of "more" and "less"*

While in mathematics the relations ">" and "<" are strictly symmetrical, in natural language words *more* and *less* are basically used for comparison, and comparison is asymmetric. When we say *The watermelon weighs three kilograms more than the melon*, we usually characterize the watermelon taking the melon as a reference point. If we reformulate the sentence by means of *less* (*The melon weighs three kilograms less than the watermelon*), the quantitative relationship between the fruit remains the same, but their roles in the pair "item to be characterized – reference point" change. In the second phrase, the watermelon is taken as a reference point and used to characterize a melon. The reference point is always the second term of comparison, the one with which the comparison is made. *More* and *less* denote the same relation but differ in the perspective. They are a classical example of converse terms.

In (Apresjan 1974 (1995): 123), *bol'še* 'more' is described as a three-place predicate with the arguments: 'what is more', 'more than what', and 'how much more'. We are inclined to consider the 'how much more' element as a circumstantial. The reason is as follows.

This element can be realized by at least three kinds of phrases:

(a) additive phrases, e.g. *bol'še na 3 kg* 'three kilograms more',

(b) multiplicative phrases, e.g. *v 5 raz bol'še* '5 times as much/many',

(c) an intermediate class of phrases that need both addition and multiplication for their interpretation, e.g. *bol'še na 50% <napolovinu, na četvert', na dve treti>* '50% (a quarter, two thirds) more'.

The problem is that if the (c)-phrases are used, *bol'še* and *men'še* sentences have noticeably different meanings; cf. sentences (7a)–(7b) with an (a)-phrase, that refer to the same situation, and (8a)–(8b) with a (c)-phrase, that are referentially different:

(7a) *Arbuz vesit na 3 kg bol'še, čem dynja*
'the watermelon weighs 3 kg more than the melon'
= the weight of the watermelon is equal to the weight of the melon **plus 3 kg**;
(7b) *Dynja vesit na 3 kg men'še, čem arbuz*
'the melon weighs 3 kg less than the watermelon'
= the weight of the melon is equal to the weight of the watermelon **minus 3 kg**;

(8a) *Arbuz vesit na 50% bol'še <na 50% tjaželee>, čem dynja*
'the watermelon weighs 50% more than the melon' [e.g. the watermelon has 6 kilograms and the melon has 4 kilograms];
(8b) *Dynja vesit na 50% men'še <na 50% legče>, čem arbuz*
'the melon weighs 50% less than the watermelon' [e.g. the watermelon has 6 kilograms and the melon has 3 kilograms].

This difference between (8a) and (8b) is due to the fact that percentages are calculated for different entities: in (8a) the weight of the watermelon is equal to the weight of the melon **plus 50% of the weight of the melon**, while in (8b) the weight of the melon is equal to the weight of the watermelon **minus 50% of the weight of the watermelon**.

If "how much more/less" phrases are considered to be actants of *bol'še* and *men'še*, the latter cannot be converse. An essential property of converse terms is that with their valencies filled, they should denote the same situation, which is not true for (8a)–(8b). Therefore, these phrases are circumstantials and should be assigned a semantic definition of their own. This definition, however, varies depending on the semantics of the predicate they are attached to.

2.2.2 *"How much more/less" Adverbials*

Phrases such as *na 3 kg* '3 kg (more/less),' *v 3 raza* 'three times (more/less)' and *na 30%* '30% (more/less)' play a similar role with respect to meanings 'more' and 'less' and should be described in a uniform way. Their description consists of two parts: (a) the semantic context: a relevant meaning component of predicate P to which the adverbial is syntactically connected; (b) the semantic contribution of the adverbial in the given context. Here are these descriptions. In square brackets an example is given.

P na N [*Cena vyrosla* (P) *na 100 rublej* 'the price rose 100 rubles']

(a) P contains the component 't_1 is more than t_2', where t_2 is the reference point.

(b) Addition: '$t_1 = t_2 + N$'

    P na N [*Cena upala* (P) *na 100 rublej* 'the price fell 100 rubles']

(a) P contains the component 't_1 is less than t_2', where t_2 is the reference point.

(b) Subtraction: '$t_1 = t_2 - N$'

    P v N raz [*Cena vyrosla* (P) *v 3 raza* 'the price tripled']

(a) P contains the component 't_1 is more than t_2', where t_2 is the reference point.

(b) Multiplication: '$t_1 = t_2 * N$'

    P v N raz [*Cena upala* (P) *v 5 raz* 'the price is 5 times less than it was before']

(a) P contains the component 't_1 is more than t_2', where t_2 is the reference point.

(b) Division: '$t_1 = t_2 / N$'

    P na chast' N [*Cena vyrosla* (P) *na 30%* 'the price rose 30%']

(a) P contains the component 't_1 is more than t_2', where t_2 is the reference point.

(b) Multiplication and addition: '$t_1 = t_2 + t_2 * N$'

    P na chast' N [*Cena upala* (P) *na 30%* 'the price fell 30']

(a) P contains the component 't_1 is more than t_2', where t_2 is the reference point.

(b) Multiplication and subtraction: '$t_1 = t_2 - t_2 * N$'

Some remarks:

    1.   All Russian "how much" adverbials combine freely with both "more" and "less" predicates, having different interpretations in these contexts, as shown above. In English, it is not always the case. While *vdvoe* 'twice' equally easily combines with increasing and decreasing contexts (*vdvoe tjazhelee tebja* 'twice your weight' – *vdvoe legče tebja* 'half your weight'), *twice* is biased towards the increasing contexts: *twice as high* – [??]*twice as low*.

    2.   It is easy to see that *napolovinu* 'by half' and *na 50%* are synonymous to *vdvoe* in the 'less' (or 'decrease') contexts and to *v poltora raza* in the 'more' (or 'increase') contexts:

    (9a) *Doxody upali napolovinu <na 50%>* '
        'the profits fell by half' [= *vdvoe*]
    (9b) *Doxody vozrosli napolovinu <na 50%>*
        'the profits rose by half' [= *v poltora raza*].

However, this is not a linguistic fact but rather a consequence of a mathematical truth: "subtracting a half is equivalent to dividing by 2" and "adding a half is equivalent to multiplying by 1.5".

Curiously enough, speakers sometimes find it difficult to correctly use expressions like *napolovinu* 'by half', *na tret'* 'by a third', etc. in increasing and decreasing contexts. Errors of this kind can be found even in the texts of irreproachable speakers. Here is an excerpt from a paper by a well-known philologist published in "Continent", 2000, №104:

"Do six por sčitalos', čto v jazyke est' tol'ko dva vyraženija, kotorye polnostju oboznačajut sami sebja. Eto slovo "slovo" i predloženie "Eto – predloženie". Vse drugie slova ne oboznačajut samogo slova, i vse drugie predloženija ne oboznačajut samogo predloženija. Teper' etot kratčajshij spisok samoznačaščix (avtoreferentnyx) jazykovyx obrazovanij možno **uveličit' srazu na tret'** (emphasis ours – IB), pribaviv k nemu odnoslovie "odnoslovie"."

'It was believed until now that in the language there are only two expressions that denote themselves. It is the word 'word' and the sentence 'This is a sentence'. All other words do not denote the word itself, and all other sentences do not denote the sentence itself. Now, this shortest list of autoreferring expressions can be increased by a third by including to it an odnoslovie (one-word-composition – IB) 'odnoslovie'.

It is true that the new member of the list makes one third of it, since the new list has three elements, but the old list of two elements is increased by half and not by a third.'

It is well-known that linguistic competence is closely related to the background world knowledge. Here we see another manifestation of this axiom: in order to use words correctly, speakers need to have some mathematical knowledge, at least elementary.

3.   An interesting case is the verb *razbavljat'* 'to dilute'. When we dilute milk with water, we take milk and add to it some amount of water. The action seems to be of the increase type, since it consists in adding something. However, the verb combines with AP as a decrease predicate. Take the recommendation (10):

   (10) *Razbav'te moloko vodoj napolovinu*
        'dilute milk with water by half'.

What does it mean? How much water should be added – an amount equal to the amount of milk or a half of this amount? Obviously, the first alternative is correct. The recommendation is to prepare the mix that contains half milk and half water. We should add as much water as we have milk. The initial quantity should be doubled. *Napolovinu* can be replaced by *vdvoe:*

   (11) *Vdvoe razbav'te moloko vodoj*
        'dilute milk with water by half'.

The situation is somewhat paradoxical. As we saw above, when something increases by half, the initial amount is multiplied by 1.5. When something decreases by half, the initial amount is divided by 2. In our case, the initial quantity increased, however the factor is not 1.5, but 2 (as in a decrease scenario).

As it often happens, if a word does not behave as we expect it would, we should look at its meaning more attentively. In fact, *dilute* contains both a component of adding and a component of decreasing, and it is the latter that is central to its meaning. To dilute milk with water means **to decrease the concentration of milk by adding water to it**. To dilute it by half means to decrease the concentration by half, by adding some water.

2.3   Context of *sostoit iz* 'consists of' and Its Synonyms

The verb *sostojat' (iz)* 'consist (of)' requires complete enumeration of components. Sentence (12)

> (12) *Komissija sostoit iz uchenyx i inženerov*
>   'the committee consists of scientists and engineers'

implies that there are no other members of the committee. Still, this verb co-occurs with the markers of partiality, such as *na N%*:

> (13) *Čelovecheskoe telo na 60% sostoit iz vody*
>   lit. 'human body 60% consists of water',
>   'human body is 60% water'.

The explanation is that the exhaustivity component in the verb meaning is weak and can be superseded by an explicit contradictory meaning, such as the meaning of AP. However, the AP meaning should not contradict the idea of exhaustivity too much. The admissible percentage should be rather high. One cannot say *Nastojka na 2% sostoit iz spirta 'the tincture is 2% spirit'. To describe this situation, one should use a verb that does not have the exhaustivity implication: *Nastojka soderzhit 2% spirta* 'the tincture contains 2% spirit'.

3   *Napolovinu* vs. *polu-* 'half'

Here we will compare two seemingly very similar elements: the prefix *polu-* and the adverb *napolovinu,* which are both closely related to the meaning 'half'[2]. In some contexts they are easily interchangeable:

> (14a) *Ego otec byl napolovinu francuz, napolovinu avstriec*
>   'his father was half French, half Austrian'.
> (14b) *Ego otec byl polufrancuz, poluavstriec.*

Still, there are important differences. First, the sentence *P napolovinu Q* implies that there exists another Q_1 such that *P napolovinu Q_1* also takes place. Sentence (15)

> (15) ?*Ona posmotrela na nego napolovinu ispuganno*
>   'she looked at him half-frightened'

is not complete. Another feature of her look should be given. On the other hand, sentence (16)

> (16) *Ona posmotrela na nego poluispuganno*
>   'she looked at him half-frightened'

is quite OK and does not ask for continuation.

Second, *napolovinu* cannot ascribe more than two characteristics to the same object. (17) is unacceptable for the same reason why nothing can consist of three halves.

---

[2]   A detailed description of the meaning and use of the prefix *polu-* is given in (Иомдин 2003).

(17) **Ona posmotrela na nego napolovinu ispuganno, napolovinu udivlenno, a napolovinu voprositel'no*
    'she looked at him half frightened, half surprised and half inquiringly'.

Prefix *polu-* is not bound with this restriction:

(18) *Ona posmotrela na nego poluispuganno, poluudivlenno, poluvoprositel'no*
    'she looked at him half frightened, half surprised and half inquiringly'

The source of these differences between *polu-* and *napolovinu* is rooted in the scope of the meaning 'half'. Both words introduce a half of some whole, but the question is what this whole is. It turns out that each word selects the whole in a different way.

When we say that she looked at him *napolovinu ispuganno, a napolovinu udivlenno* we claim that **among the properties of her look**, one half was a manifestation of fright, and another half – a manifestation of surprise. Obviously, no third half can exist. When we say that she looked at him *poluispuganno i poluudivlenno,* we are saying something different. We claim that **among the properties of fright**, only a half was present in her look. It was not real fright, it was half-fright, something similar to fright, but not proper fright. Similarly, it was not real surprise that could be seen in her look, it was half-surprise, i.e. something that had some properties of surprise, but not all. Her look was a mixture of different feelings. Each of them was represented by only some of its properties (literally, by half of them), and there is no reason why this mixture cannot contain more than two components.

Acknowledgements

Bibliography

Апресян, Ю.Д. 1974 (2 ed. 1995). *Лексическая семантика.* Москва: Школа "Языки русской культуры".

Иомдин, Б. Л. 2003. Семантика русской приставки ПОЛУ- // *Русистика на пороге XXI века: проблемы и перспективы. Материалы международной научной конференции (Москва, 8–10 июня 2002 г.).* Москва, ИРЯ РАН, с. 109–113.

Sex-Based Nominal Pairs in the French Lexical Network:
It's Not What You Think

Candice Delaite (1), Alain Polguère (2)

(1) ATILF CNRS
candice.delaite@atilf.fr
(2) Université de Lorraine & ATILF CNRS
alain.polguere@univ-lorraine.fr
44 avenue de la Libération, BP 30687
54063 Nancy Cedex
France

Abstract

We examine French nominal pairs of the type *berger* 'shepherd' ~ *bergère* 'shepherdess' and *chien* 'dog' ~ *chienne* 'bitch dog'. These pairs are considered from the viewpoint of their lexicographic modeling in an Explanatory Combinatorial lexical database: the French Lexical Network (FLN). We first present, in section 1, the linguistic problem we are dealing with; we identify two main types of semantic relations involved in these lexical pairs: sex-based quasi-synonymy (that is analyzed in detail in section 2) and sex-based contrastive opposition (is dealt with in section 3). In the form of conclusion, in section 4, we look at the implications of our modeling of such relations on the graph structure of the FLN.

Keywords

Lexical relation, lexical function, Explanatory Combinatorial Lexicology, grammatical gender, lexical database, French Lexical Network (FLN).

1 Statement of the Problem

1.1 N_{masc} ~ N_{fem} Lexical Pairs Based on Sex Difference

This paper deals with the problem posed to lexicographic description by French nominal pairs that will be referred to as N_{masc} ~ N_{fem} and that are characterized by the following four properties:

1. both lexical units N_{masc} and N_{fem} denote animate beings of the same kind, except for a difference in sex;

2. N_{masc} carries the masculine grammatical gender and denotes an animate being whose sex is, necessarily or by default, male;

3. N_{fem} carries the feminine grammatical gender and denotes an animate being whose sex is, necessarily or by default, female;

4. N_{masc} and N_{fem} are generally, but not necessarily, morphologically related.

Here are a few examples of some $N_{masc} \sim N_{fem}$ pairs: *berger* 'shepherd' ~ *bergère* 'shepherdess', *hôte* 'host' ~ *hôtesse* 'hostess', *nageur* '(male) swimmer' ~ *nageuse* 'female swimmer', *chien* 'dog' ~ *chienne* 'bitch dog', *étalon* 'stallion' ~ *jument* 'mare', *fiancé* ~ *fiancée*.[1]

We examine these lexical pairs in the context of the RELIEF lexicographic project that targets the building of the French Lexical Network or FLN (Lux-Pogodalla & Polguère, 2011; Gader *et al.*, 2012). The FLN is a model of the French lexicon that is being built according to Explanatory Combinatorial Lexicology's principles (Mel'čuk *et al.*, 1995); its main characteristic is to be formally structured as a multidimensional lexical graph, instead of being a simple "lexicographic text". The bulk of the FLN structuring results from the weaving of Meaning-Text lexical function relations (Mel'čuk, 1996).

Since our perspective is lexicographic in essence, it is useful to begin with a brief summary of how "classical" dictionaries handle $N_{masc} \sim N_{fem}$ pairs, focusing on the pairs that are morphologically related.

The three main general public dictionaries of French – *Petit Robert*[2], *Petit Larousse*[3] and *Trésor de la Langue Française*[4] – offer heterogeneous descriptions of morphologically related $N_{masc} \sim N_{fem}$ pairs. In some cases, there are individual entries for each pair member: for instance, *marquis* ~ *marquise* or *poète* 'poet' ~ *poétesse* 'poetess', and in other cases, both lexical units are grouped under a unique entry – for instance, *coiffeur* ~ *coiffeuse*, that are being described under a common entry named "**coiffeur, -euse**." The above-mentioned dictionaries seem to converge in their descriptive choices, for which it is normally possible to find a logical explanation. Thus, for the few cases that we have cited:

- the choice of a separate entry for *marquise* is almost imposed to dictionary makers by the polysemy that developed around the feminine noun: *marquise* (*d'une gare*) 'canopy (of a railway station)', *marquise* (*au chocolat*) 'chocolate cake', with no corresponding sense within the masculine vocable;

---

[1] We do not provide glosses for French lexical units that exist with identical meaning in English as borrowings. The pairs listed above form a rather heterogeneous set from a semantic and morphological viewpoint. However, we believe that they have to be considered and dealt with together, as will be shown shortly.

[2] CD-ROM edition, 2011.

[3] Online consultation: http://www.larousse.com/fr/dictionnaires/francais-monolingue.

[4] Online consultation: http://atilf.atilf.fr.

- the derivational suffix *-esse* possesses a special status in French morphology that makes *poétesse* appear as morphologically more "remote" from *poète* than *coiffeuse* appears to be from *coiffeur*.[5]

It is obvious that the lexicographic treatment of sex-based semantic derivations will depend on the grammatical characteristics of each language. Because there is no gender inflection in English, the morphological nature of the relation holding between the source and target of such derivation poses no problem. Furthermore, there are relatively few derivations of that kind based on the use of the suffix *-ess* (borrowed from French). For these reasons, English dictionaries always create two separate entries, one for the noun that denotes a male animate being and one for the noun that denotes the corresponding female being: see, for instance, *lion ~ lioness* in the *American Heritage*,[6] *Longman Dictionary of Contemporary English*[7] or *Cobuild*.[8]

Spanish is a more interesting case here, as Spanish nouns possess grammatical gender, and Spanish grammar contains gender inflection for adjectives, participials and articles. As a side effect, we find many $N_{masc} \sim N_{fem}$ pairs in Spanish, like in French. Similarly to French dictionaries, the *Diccionario de la lengua español* (Real Academia Española)[9] and the *Diccionario de uso del español* (Moliner, 2000) propose a heterogeneous treatment of $N_{masc} \sim N_{fem}$ pairs: two separate entries for *modisto* 'couturier, (male) fashion designer' ~ *modista* 'couturière, female fashion designer', but one single entry for *pastor* 'shepherd' and *pastora* 'shepherdess' named "**pastor, ra**" in the first dictionary and "**pastor, -a**" in the second. The case of Spanish is however very different from French because Spanish common nouns, in standard cases, possess a suffix that carries grammatical gender – cf. Mel'čuk (to appear) for an analysis of the morphological status of Spanish gender nominal suffixes. In the FLN, we treat morphologically related $N_{masc} \sim N_{fem}$ pairs in a systematic fashion, based on the **derivational** (and not inflectional) nature of the morphological link that unites them. Let us examine this point briefly.

1.2 Derivation and not Inflection

It is obvious that, from a strictly formal viewpoint, pairs of wordforms such as *berger* 'shepherd' ~ *bergère* 'shepherdess' display inflection-like characteristics: the signifier *bergère* looks like a "feminine of" *berger*. However, as shown by Mel'čuk (2000), there are at least five reasons why one should consider that no inflectional mechanism is involved here.

1. A description based on inflection would force us to postulate an inflectional category of nominal gender in French, that would apply only to nouns denoting sexed animate

---

[5] The suffix *-esse* is archaic and rather rare in modern French; it is almost no longer used to produce neologisms. See the etymological notice of the *Trésor de la Langue Française,* in its entry -ESSE[2]. In comparison, the derivation based on *-eur → -euse* is extremely productive.

[6] Online consultation: http://www.ahdictionary.com.

[7] Online consultation: http://www.ldoceonline.com.

[8] On-line consultation: http://dictionary.reverso.net/english-cobuild/teach.

[9] 20th edition, on-line consultation: http://www.rae.es/rae.html.

beings, whereas typical inflection applies globally to all lexical units of given parts of speech (see grammatical gender for French adjectives).

2. This hypothetical inflection would be very irregular – cf. *professeur* '(male) teacher' ~ *professeure* 'female teacher', but *coiffeur* ~ *coiffeuse* – and non-systematic – cf. [*un*] *poisson*$_{masc}$ '[a] fish', sex-neutral masculine noun that does not have a feminine counterpart such as *[une] poisson*$_{fém}$ or **poissonne*.

3. For these reasons, it is not uncommon that even native speakers hesitate on the proper choice of term, which is rather strange if we were to be in presence of inflection. Additionally, diatopic variations that are so common for N$_{masc}$ ~ N$_{fem}$ pairs, are clearly perceived as lexical rather than grammatical in essence – cf. *écrivaine* 'female writer', that is almost systematically used in Québec when talking about a woman writer, *vs.* the feminine noun [*une*] *écrivain*$_{fem}$, that is much more commonly used in France.

4. As mentioned in section 1.1, each element of a N$_{masc}$ ~ N$_{fem}$ pair tends to develop its own individual polysemy; there are therefore many pairs of vocables in French that are autonomously structured, which contradicts the inflectional modeling of the phenomenon we are examining.

5. Finally, derivational rather than inflectional description of French N$_{masc}$ ~ N$_{fem}$ pairs can be done in a very simple and direct manner. The only justification one can find for the approach commonly adopted by standard dictionaries, which tend to model these pairs as if they were inflectionally related (section 1.1 above), is the significant economy in printed pages it entails for printed versions of the dictionaries.

These observations logically lead us to consider that morphologically related N$_{masc}$ ~ N$_{fem}$ French pairs are cases of pure **derivations**, where nominal masculine lexical units denoting a male animate being are used as a source from which feminine lexical units denoting the corresponding female being is derived. Dictionaries and lexical databases, such as the FLN, should explicitly and systematically reflect this fact in their macro- and micro-structure.

1.3   Hypothesis: Two Types of N$_{masc}$ ~ N$_{fem}$ Pairs

Now that it is clearly established that pairs of nominal lexical units such as *boulanger* '(male) baker' ~ *boulangère* 'female baker', *fils* 'son' ~ *fille* 'daughter', *chat* '(male) cat' ~ *chatte* 'female cat', etc. are indeed displaying cases of morphological derivations and that no inflection mechanism is involved, we will identify the different types of such derivations and indicate how each type has to be modeled in the FLN.

Let us recall that we are interested in the links of semantic derivations that are not necessarily expressed by morphological means and that we are considering both morphologically related pairs such as those examined in section 1.2 above as well as pairs such as *frère* 'brother' ~ *sœur* 'sister' or *étalon* 'stallion' ~ *jument* 'mare'.

We take the lexicographic perspective of the FLN construction and of its graph structure weaving by means of lexical function links. In this context, it is of paramount importance for

all N_{masc} ~ N_{fem} lexical pairs to be described in a homogeneous way, whether or not they display a morphological connection.

We believe two relations exist that can hold between a masculine nominal lexical unit denoting a male animate being and a feminine nominal unit denoting the corresponding female being. These relations are illustrated by the following lexical pairs:

1. the pair *avocat* '(male) lawyer, advocate' ~ *avocate* 'female lawyer', that which displays, according to us, a synonymy-like lexical relation;

2. the pair *étalon* 'stallion' ~ *jument* 'mare', which displays a lexical relation that is, in essence, contrastive rather than quasi-synonymic.

We are now proceeding to the presentation of these two types of relations: sex-based quasi-synonymy in section 2 and sex-based contrastive opposition in section 3.

2   Sex-Based Quasi-Synonymy

2.1   Characterization

In lexical pairs of the *avocat* '(male) lawyer' ~ *avocate* 'female lawyer' type, one of the two lexical units, the masculine or the feminine noun (section 2.2 below), functions as generic. Thus, *avocat* in (1a) below necessarily denotes a male individual, while the sex can be unspecified in (1b).

(1)   a. *Le jeune avocat s'est avancé résolument vers les journalistes.*
  'The young lawyer walked resolutely towards the journalists'
  b. *Il a demandé la présence d'un avocat.*
  'He asked for a lawyer'

The plural form *avocats* in (2a) can designate a group of both men and women, and (2b) as well as (2c) can be used by a woman – i.e. she can either use the masculine or feminine noun.

(2)   a. *Tous les avocats du barreau parisien ont protesté.*
  'All the lawyers of the Parisian Bar have protested'
  b. *Je veux devenir avocat.*
  'I want to become a lawyer'
  c. *Je veux devenir avocate.*
  'I want to become a female-lawyer'

All this shows that there is some form of semantic unevenness between *avocat* and *avocate*:

(3)   a. *avocat* = '(male) individual whose profession consists in [...] or corresponding profession'
  b. *avocate* = 'female individual whose profession consists in [...] or corresponding profession'

In other words, these lexical units are both semantically ambivalent[10], denoting either an individual or a profession, but not in the same fashion. Whether it is used to denote an individual or the profession she exercises, *avocate* implies the female sex, which is not the case for *avocat* regarding the male sex. The "male" semanteme that is embedded in the definition of this latter lexeme must be considered being a weak semantic component (Mel'čuk *et al.,* 1995: 95); it is this semantically weak nature that is being signaled by the parentheses in the gloss (3a) above.

The complex semantic ratio holding between lexical units such as *avocat* and *avocate* entails that lexical units that form pairs of this type are not perceived as being in opposition vis-à-vis the denotation of sex. Elements of the pair *avocat ~ avocate* are not contrasting on the ground that one would denote a "male" and the other – the corresponding "female". Because of this, we consider such lexical units as quasi-synonyms and we call this type of relation *sex-based quasi-synonymy.*

It we take into consideration the semantic unevenness identified above, we need to distinguish between the following two cases.

First of all, a speaker who says (4a) instead of the prototypical sentence (4b) chooses to be more specific with regards to the designation of sex.

> (4) a. – *Je veux une avocate!*
>   'I want a female lawyer'
>   b. – *Je veux un avocat!* [man or woman]
>   'I want a lawyer'

Consequently, we encode the semantic derivation link *avocat → avocate* by means of the **Syn$_\supset$^{sex}** lexical function, which stands for «richer synonym with regards to the sex».[11]

Secondly, a speaker who says (4b) rather than (4a) chooses to be less specific with regards to the designation of sex. We therefore encode the semantic derivation link *avocate → avocat* by means of the **Syn$_\subset$^{sex}** lexical function, which stands for 'potentially less rich synonym with regards to the sex'. We are considering here a synonymy that is **potentially** less rich because the speaker can of course also use *avocat* in order to specifically signify a male individual.

Notice that the two lexical functions that have just been introduced correspond to two reciprocal relations:

Syn$_\supset$^{sex}(L1) = L$_2$ entails that **Syn$_\subset$^{sex}**(L$_2$) = L1, and vice versa.

There clearly is a considerable number of lexical pairs that are based on these quasi-synonymy relations. Let us mention, among others:

---

[10] On semantic ambivalence and its lexicographic treatment, see Milićević & Polguère (2010). Note that because this particular semantic ambivalence is systematic in French — all nouns denoting an individual who exercises a given profession can also denote the profession itself — Milićević & Polguère (2010) propose to not make the ambivalence explicit in the structure of the definition and to simply formulate the genus of this type of lexical unit using the general pattern 'individual who has a given profession'.

[11] The exponent ^{sex} in **Syn$_\supset$^{sex}** is a new standardized element we introduce in the formal language of lexical functions; therefore, it stands for Lat. *sexus* and not for Eng. *sex.*

- nouns denoting individuals who hold an official title or function – cf. *roi* 'king' ~ *reine* 'queen', *marquis* ~ *marquise* 'marquess', *président* 'president' ~ *présidente* 'female president' or *pape* 'pope' ~ *papesse* 'female pope';

- nouns denoting individuals who have a given citizenship, philosophy, religion, etc. – cf. *Français* 'French person' ~ *Francaise* 'female French person', *Juif* 'Jew' ~ *Juive* 'female Jew' or *bouddhiste*(N, masc) 'Buddhist' ~ *bouddhiste*(N, fem) 'female Buddhist';

- nouns denoting individuals who experience a given emotion or possess a given behavioral characteristic – cf. *amoureux*(N) 'person in love' ~ *amoureuse*(N) 'female person in love' or *lâche*(N, masc) '[a] coward' ~ *lâche*(N, fem) 'female coward';

- nouns denoting animals, such as *chien* 'dog' ~ *chienne* 'bitch dog' or *éléphant* 'elephant' ~ *éléphante* 'female elephant'.

We do not mean to imply that all nominal pairs that fit into the above categories are in $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$ vs $\mathbf{Syn}_{\subset}^{\mathbf{sex}}$ relation. In many instances, such as, for example, the pair of animal names *étalon* 'stallion' ~ *jument* 'mare', we observe another type of relation, contrastive in essence, which will be dealt with in section 3.

2.2 Masculine or Feminine Generic

It is important to stress that the semantic ratio between two lexical units that are linked by sex-based quasi-synonymy as well as the very existence of this type of quasi-synonyms for a masculine vs. feminine noun are not at all systematic in French, as demonstrated by the following two cases.

1. It can be the feminine noun that functions as generic, while denoting primarily a female animate being. For instance, *oie* 'goose' denotes a female animal or a species; its $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$ is *jars* 'gander', which cannot be used to denote a species.

2. A feminine noun denoting an animal can be completely neutral with regards to the denotation of the sex; therefore, it can have no corresponding masculine noun in French. Such is the case of *girafe* 'giraffe'. To specifically denote a male or female giraffe, one will have to say *girafe mâle* ~ *girafe femelle*.

It is, however, still possible to make generalizations. For instance, we believe that in French the generic is always the masculine noun **for nouns of N$_{\mathbf{masc}}$ ~ N$_{\mathbf{fem}}$ pairs denoting human beings**. We have not found any exception to this rule. It applies even for a pair such as *infirmier* '(male) nurse' ~ *infirmière* 'female nurse', about which one may easily think that the feminine noun is the generic. Rather than being a potential generic, the feminine noun is the default for extra-linguistic reasons, bearing to the fact that this profession has for a long time been exclusively exercised by women. But *infirmière* denotes specifically a female individual and cannot be used as generic; (5) below cannot be used in cases where one wants to refer to a mixed group of women and men.

(5) *Un groupe d'infirmières se tenait à l'entrée de l'hôpital.*
'A group of female nurses was standing at the entrance of the hospital'

Contrary to what one may first believe, it is indeed the masculine noun that is used in contemporary French as generic. For instance, *vêtements d'infirmiers* 'nurse clothes' in (6) below (found on the Web) can refer to both men's and women's clothes.

(6) *L'association a ainsi acheminé du matériel médical: 140 lits médicalisés manuels ou électriques, un bloc opératoire, 4 750 paires de lunettes de vue et des vêtements d'infirmiers.*
'The association has thus shipped medical equipment: 140 manual or automatic medicalized beds, one operating room, 4,750 spectacles and nurse clothes'

Other generalizations can probably be drawn about sex-based quasi-synonymy, but we do not yet have enough data in the FLN to be able to extract them. We now proceed to the second type of sex-based semantic derivation.

3   Sex-Based Contrastive Opposition

As mentioned above, some $N_{masc} \sim N_{fem}$ pairs do not fall in the scope of quasi-synonymy. For instance, it would be odd to state that *étalon* and *jument* are quasi-synonyms as they seem to be linked by a semantic opposition: a stallion is a male horse, in contrast with a mare, which is a female horse. Which lexical function should be used to account for the lexical relation connecting lexical elements of such pairs, given that the use of **Syn** is ruled out?

There is the **Contr** "contrastive", a lexical function which can be characterized as follows:

Contr(L_1) = L_2 if phraseological expressions exist that feature both lexical units L_1 and L_2 in contrastive opposition. For instance, **Contr**(*noir* 'black') = *blanc* 'white' because there are expressions in French such as (*photo*) *en noir et blanc* 'black and white (picture)', *Tout n'est pas tout blanc ou tout noir* 'Everything is not always black and white', etc.

This characterization, which is not based on the identification of an intrinsic semantico-syntactic content of **Contr**, entails that **Contr** is not an "ordinary" lexical function. A lexical unit that is the value of **Contr** for another lexical unit is not, strictly speaking, a semantic derivative of this unit. All this shows that pairs such as *étalon ~ jument*, which display an obvious semantic contrast, are not in a **Contr** relation. Indeed, the elements of these pairs are related through a semantic opposition that results directly from their lexical definition, rather than through the presence of phraseological expressions in the language that put them in contrast with each other.

Though not **Contr**, the semantic relation we consider here is much more related to contrastivity than it is, for instance, to antonymy – cf. the **Anti** lexical function. It is not based on the negation of a definitional semantic component lexical units it connects: compare true antonyms such as *présent*(Adj) 'that is here' ~ *absent*(Adj) 'that is **not** here' with non-antonymic contrastive pairs such as *étalon ~ jument*. The non-antonymic nature of the relation between *étalon* and *jument* stands out clearly if we draw the definitions of the two semantemes that

distinguish them: "female [X]" included in the meaning of JUMENT and "male [X]" included in the meaning of *étalon*:[12]

- 'female [X]' = '[X] whose sex enables her / it to give birth';

- 'male [X]' = '[X] whose sex enables him / it to participate in the fact that a female individual of the same species gives birth'.

The relation that concerns us being close to, but distinct from **Contr**, it should in theory be lexicographically modeled by means of two non-standard lexical functions:

    (7) **Corresponding female animate being** (*étalon*) = *jument*;

    (8) **Corresponding male animate being** (*jument*) = *étalon*.

However, these two non-standard lexical functions seem to meet most of the requirements for standardness (Polguère, 2007): they correspond to very regular relations in French (and probably in all natural languages), they are applicable to a very large and varied set of arguments, the returned values are many and varied, and finally they are often expressed by morphological means. For these reasons, we have decided to standardize these two symmetrical relations and make use of the two "normalized" encodings in the FLN: **Fem** for lat. 'feminus' and **Masc** for lat. 'masculus'.

Lexical function **Fem** and **Masc** have already been used in the DiCo project (Mel'čuk & Polguère, 2006), following the proposals made by A.-L. Jousse in order to normalize the encoding of some recurrent non-standard lexical functions (Jousse, 2010: 139–140). Technically, however, those were not the exact same **Fem** and **Masc** that are being used in the FLN because the DiCo used them indiscriminately for both families of semantic derivations that we have identified – synonymic and contrastive.

Lexical pairs that are connected by **Fem** *vs.* **Masc** relations can very well interact in the lexicon with a third lexical unit that connects to the first two as quasi-synonym. Such is the case for *cheval* 'horse', with regards to its relation to the *étalon ~ jument* pair. The lexical definition of *cheval* is similar in structure to those of lexical units such as *avocat* 'lawyer' examined in section 2. The masculine noun *cheval* denotes a certain type of animal which is, by default but not necessarily, a male. Figure 1 below visualizes the complete system of lexical relations connecting *cheval*, *étalon* and *jument*.

---

[12] For the sake of simplicity, we define two English semantemes, though, strictly speaking, the corresponding French semantemes "femelle" and "mâle" should be analyzed. The following definitions are loosely based on semantic treatment of lexical units denoting sexed beings that is proposed by Wierzbicka (1972: 34-56).
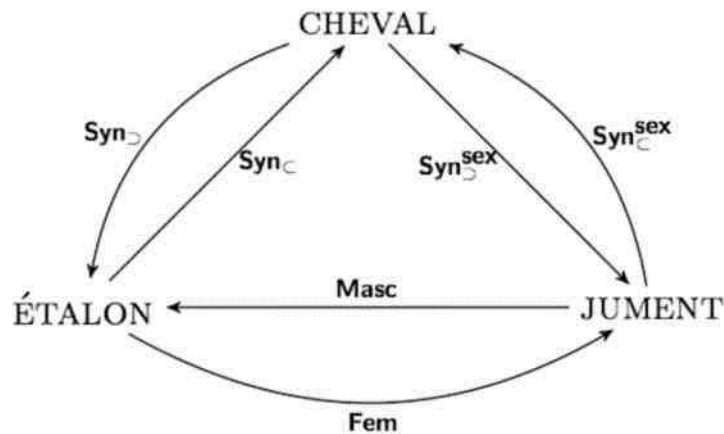
Figure 1: Lexical relations connecting *cheval* 'horse', *étalon* 'stallion' and *jument* 'mare'

To conclude this section, below we give two characteristics of sex-based contrastive oppositions in French that we have induced from the lexicographic work done on the FLN.

1. **Fem** and **Masc** relations are much less present in the lexicon than $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$ and $\mathbf{Syn}_{\subset}^{\mathbf{sex}}$.

This is easily explained by the fact that approximate synonymy is precisely the most common relation in the graph of any natural language lexicon.

2. $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$ *vs.* $\mathbf{Syn}_{\subset}^{\mathbf{sex}}$ is often realized morphologically, whereas it is hardly the case for **Fem** *vs.* **Masc**. In fact, we are yet to find a clear-cut case of a **Fem** that would be morphologically built out of its corresponding **Masc**, or vice versa.

Those are only preliminary observations, which ought to be systematically checked on a larger set of data.

4   Implications on the Structure of the FLN

Our analysis of sex-based semantic derivatives influences the development of the the FLN's structure in two ways.

Firstly, the FLN, unlike standard French dictionaries, systematically possesses distinct lexicographic entries for $N_{masc} \sim N_{fem}$ pairs. Because polysemy tends to develop independently for both vocables of these pairs, one should never try to unite them under a single entry, even when they make use of the same signifiers – cf. *lâche*$_{(N, masc)}$ '[a] coward' ~ *lâche*$_{(N, fem)}$ 'female coward'. The saving of printed paper is not a parameter in the context of the FLN and this method allows us to remain closer to what we believe is the actual structuring of the lexicon.

Secondly, paradigmatic links that connect lexical units involved in the system of $N_{masc} \sim N_{fem}$ oppositions are systematically encoded by FLN lexicographers using the appropriate lexical functions: $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$, $\mathbf{Syn}_{\subset}^{\mathbf{sex}}$, **Masc** or **Fem**. At the time of writing, the FLN contained 18,470 lexical units (senses) connected by a total of 26,959 lexical links; among those, 1,618 lexical function links encode lexical relations geared to $N_{masc} \sim N_{fem}$ pairs: 1,592 links for $\mathbf{Syn}_{\supset}^{\mathbf{sex}}$ *vs.*

Syn$_\subset^{\mathbf{sex}}$ relations and only 116 links for **Fem** *vs.* **Masc** relations. These statistics confirm what was said earlier: sex-based synonymy is significantly more present in the lexicon than the sex-based contrastive opposition.

Let us mention an interesting problem, which we will not develop for the lack of space. There are degrees in the synonymic *vs.* contrastive nature of links connecting $N_{masc} \sim N_{fem}$ lexical units. For instance, the *fils* 'son' ~ *fille* 'daughter' pair pertains to sex-based contrastive opposition because none of the two lexical units can be used as generic denoting someone's child without specification of the sex. However, the contrastive nature of the link appears to be weak for at least two reasons:

1. What dominates in the meaning of both lexical units is their relational nature '[X is] son/daughter of Y' and not, strictly speaking, the denotation of a certain type of individual;

2. The relation, central to the meaning, is identical in both cases.

To conclude, we would like to draw the reader's attention to Steffens (2011), a paper that we happened to come across right when the final version of the present text was being submitted. It takes a non-lexicographic approach to the question of $N_{masc} \sim N_{fem}$ pairs while reaching the conclusions that are astonishingly (and reassuringly) similar to ours.

Acknowledgements

Bibliography

Gader, N., V. Lux-Pogodalla & A. Polguère. 2012. Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. In *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)*, 109–125. Mumbai.

Jousse, A.-L. 2010. *Modèle de structuration des relations lexicales fondé sur le formalisme des fonctions lexicales*. Ph.D. dissertation, Université de Montréal & Université Paris Diderot (Paris 7), Montreal & Paris.

Lux-Pogodalla, V. & A. Polguère. 2011. Construction of a French Lexical Network: Methodological Issues. In *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, 54–61. Ljubljana.

Mel'čuk, I. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In Wanner, L. (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Language Companion Series 31, 37–102. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, I. 2000. Un FOU/une FOLLE: un lexème ou deux? *Bulag,* hors-série: 95–106.

Mel'čuk, I. To appear. The Notion of Inflection and the Expression of Nominal Gender in Spanish, *Studies in Language*.

Mel'čuk, I., A. Clas & A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*, Paris/Louvain-la-Neuve: Duculot.

Mel'čuk, I. & A. Polguère. 2006. Dérivations sémantiques et collocations dans le DiCo/LAF, *Langue française*, 150:66–83.

Milićević, J. & A. Polguère. 2010. Ambivalence sémantique des noms de communication langagière du français. In Neveu, F., V. Muni Toke, J. Durand, T. Klingler, L. Mondada & S. Prévost (eds.), *Actes de la section «Lexique et morphologie» du 2e Congrès Mondial de Linguistique Française (CMLF'10)*, 1029–1050. New Orleans.

Moliner, M. 2000. *Diccionario de uso del español*, primera reimpresión, edición abreviada por la Editorial Gredo, Madrid: Gredos.

Polguère, A. 2007. Lexical function standardness. In Wanner, L. (ed.), *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel'čuk*, Language Companion Series 84, 43–95. Amsterdam/Philadelphia: John Benjamins.

Steffens, M. 2011. *Homme/femme*, mode d'emploi. Existe-t-il des "antonymes de genre"? *Actes des XXVe Journées de linguistique*, 14–22. Université Laval, Québec.

Wierzbicka, A. 1972. *Semantic Primitives*. Linguistische Forschungen 22. Frankfurt: Athenäum.

An Information-Structure Account of Instrument-Subject Constructions

Christiane Fellbaum and Tova Rapoport

Computer Science, Princeton University
35 Olden Street, Princeton NJ (USA)
fellbaum@princeton.edu
Ben-Gurion University of the Negev
POB 653, Beer Sheva (Israel)
tovarap@bgu.ac.il

Abstract

This paper examines the Instrument-Subject Construction (ISC), addressing the question of the conditions on the Instrument's occupation of a position usually reserved for the Agent. We argue for an account that involves an interplay of the semantics of the Instruments themselves, the verbs selecting them and crucially, Information Structure. As a part of our analysis, we offer a typology of instruments based on the degree of Agent involvement vs. Instrument control over the event. The typology allows distinctions between the Instruments that can more freely occur in ISCs and those that are subject to specific conditions. We propose a key division between primary and secondary Instruments, demonstrate that ISCs are used only in particular contexts and, distinguishing among types of generic and eventive ISCs, conclude with an Information Structure explanation for the existence of ISCs.

Keywords

Instruments, Subjects, Verbs, Information Structure, Topic, Focus, Middle, Responsibility, Primary vs. Secondary Instrument, Instrument typology.

1 Introduction

The Instrument Subject Construction (ISC) is a construction in which the argument bearing the semantic role of an Instrument occurs in subject position, rather than in the more usual *with*-phrase. Consider the following pairs (most of our data has been harvested from the Web):

(1) a.   I cut the magazine with a kitchen knife.
     b.   The kitchen knife cut through the magazine much more easily.

(2) a.   The chef melted the chocolate with a Bunsen burner.
    b.   A Bunsen burner melted the chocolate.

(3) a.   I opened the door with the spare key.
    b.   The spare key opened the door.

Yet, as has been noted (Wojcik 1976; Marantz 1984; Brousseau & Ritter 1991; Schlesinger 1989 inter alia), not every Instrument *with*-phrase argument can appear in subject position:

(4) a.   She swept the floor with a broom.
    b.   *A broom swept the floor.

(5) a.   They signed the letter with the pen.
    b.   *The pen signed the letter.

We argue that several factors conspire to produce a well-formed ISC and offer an account of the constraints on ISCs that involve an interplay of the semantics of the Instruments themselves, the verbs selecting them and crucially, Information Structure.

2   The Instrument Type

Intuitively, the kitchen knife, Bunsen burner, spare key, broom and pen from the examples (1)–(5) are all Instruments. However, it is not easy to arrive at an exact definition of the class of Instruments. This class contains a varied range of elements, and a precise definition in terms of either syntactic diagnostics or semantic features is hard to pin down (Fillmore 1968; Nilsen 1973; Marantz 1984; Schutze 1996, Rissman to appear, inter alia). For our purposes, we adopt a notion of Instrument as an inanimate element that necessarily involves Agent implementation, whether continually throughout the event or solely at its initial point; the Agent is not required to be overt, i.e. linguistically present. As a diagnostic we assume well-formed paraphrases with the adverbial *with*-phrase and with the sentence '*Agent uses X for Verb-ing*'. We propose here a two-way classification of Instruments based on the degree of control over the Instrument that the Agent exercises:

A. Instruments that require manipulation and physical control by an Agent throughout the event fall into five subclasses:

   (a) Implements and mechanical tools, including tools designed for specific purposes:
        These knives killed some inmates.
        The Rabbit corkscrew opened two bottles.

   (b) Body parts of the Agent:
        His booted foot broke down the door.
        A slow finger traced the line where lace met flesh.

   (c) Materials; these may be used in conjunction with an implement:
        Although it polished the silver nicely, it was hard to use on intricate jewelry.
        The product cleans counters, stove tops, the interior of refrigerators, tile.

(d) Abstract entities involving an associated agentive event, including event nominals:
The combination opened the safe.
One kick of my foot opened the door.

B. "Autonomous" devices that require the initiation by an Agent but not manipulation over the entire duration of the event:

(a) Machines:
The washing machine washed all my filthy clothes.
The potato peeling machine peeled a truckload of potatoes in seconds.

(b) Substances:
How do we know polonium killed Litvinenko?
The detergent cleaned our cloth diapers very well.

(c) Natural Forces harnessed by an Agent:
Wind/water generates one third of the power in Holland.

Our proposed typology shares some aspects of Schlesinger's (1989) discussion of graded Agenthood (a "cluster concept") and Control. Invoking Lakoff's (1977) notion of Responsibility (Lakoff 1977), Schlesinger notes that ICSs draw attention away from the Agent. Our typology reflects that Control is strongly correlated with ISC formation.

3   Constraints on the ISC Subject

As mentioned, the set of proposed Instruments in the literature is varied. Our focus here is on accounting for the distribution of ISCs, and it turns out that such an account cannot be given solely in terms of any of the semantic properties that have been proposed for Instrumenthood.

Both Marantz (1984) and Rissman (to be published) attribute the constraints on ISCs to the verb's semantics. The verb determines whether an argument is assigned the Instrument role and the particular semantic type of that Instrument. Marantz introduces a distinction between "intermediary" and "facilitating" instruments, which he asserts accounts for the difference between (6b) and (7b):

(6) a.   Elmer unlocked the porcupine cage with a key.
    b.   A key unlocked the porcupine cage.

(7) a.   Elmer examined the inscription with a magnifying glass.
    b.   *The magnifying glass examined the inscription.

The key in (6) is an intermediary agent instrumental in the unlocking event: Elmer does something to the key, the key does something to the cage, and the cage unlocks. By contrast, Marantz states, the magnifying glass is an indispensable (facilitating) instrument in Elmer's examination of the inscription. The former class of Instruments, but not the latter, can serve as subjects. While it is the case that the magnifying glass does not act as an "intermediary agent" (i. e. it does not act on the inscription), it is not clear whether the indispensability of the magnifying glass is indeed the crux of the matter. Consider (8), from the Web:

(8)   The naked eye or even a magnifying glass will not decode the red fox [a stamp].

Whatever accounts for the difference between (6) and (7), it is not the "intermediary" vs. "facilitating" contrast that Marantz argues for: the magnifying glass in (8) does not act on the red fox. Moreover, one can, in principle, both examine and decipher something without a facilitating Instrument, so the requirement for such an Instrument is arguably not part of these verbs' lexical entry. The data indicate that an account of ISCs must consider factors beyond the nature of the Instrument, optionality of the Instrument, and in particular a "facilitating" vs. "intermediary" semantic distinction.

3.1 Primary vs. Secondary Instruments

Wojcik (1976), like Marantz, distinguishes a class of enabling (roughly equivalent to facilitating) Instruments, offering the following example (parallel to Marantz' (7a-b) above):

(9) a. John ate the food with a fork.
    b. *John's fork ate the food.

The implement here is not indispensable – one can eat with one's hands; yet the ISC is bad. To account for sentences like (9) we offer an alternative explanation. Consider, first, the following additional examples:

(10) a. *My cane walked to the door.
    b. *My new glasses saw everything.
    c. *My new hearing aid heard every note of the concert.
    d. *The straw drank a whole liter of cola.

The ISCs of (10) are all bad, despite the acceptability of the *with*-phrase and *use-for* paraphrases that points to the Instrument status of their subjects:

(11) a. Jane walked to the door with a cane. / Jane uses a cane for walking.
    b. I saw every sign with my new glasses. / I use my new glasses for seeing.
    c. I heard every note with my new hearing aid. / I use my new aid for hearing.
    d. I drank a whole liter of cola with a straw. / I used a straw to drink the cola.

Yet the subjects in (11) are not indispensable Instruments, and so the corresponding ISCs in (10) should be well-formed by Marantz' account. We argue that the reason for the unacceptability of the examples (10) has nothing to do with the optionality or indispensability of the Instrument, but rather that the subjects here are not in fact true Instruments of the actions. A more basic Instrument is covertly, and consistently, involved in each of these actions. The verbs in (10)–(11) describe physical actions: walking, seeing, hearing – that necessarily involve a body part. And so each verb has its own default, understood (and necessary) Instrument: legs, eyes, ears. It is these human parts that are the Implements in the relevant actions. Therefore, as expected, body parts can appear as subjects of ISCs with the appropriate verbs:

(12) a My strong legs climbed every meter of that mountain.
    b. My one good eye reads all the signs.
    c. My excellent ears heard every note of the concert.
    d. The naked eye will not decode the red fox.

These default body part Instruments are indispensable for the actions described, thus arguing against Marantz' and Levin's (1993) claim that such Instruments cannot be ISC subjects (or at

least against the relevance of optionality to Marantz' distinction). The primary body part Instruments can be ISC subjects, but the secondary aids that enable primary Instruments to perform the actions in question cannot, hence the unacceptability of the Implements as subjects in (10). This account makes sense in light of (Lakoff 1977) view that the Instrument subject assumes "responsibility" for the event. (The primary-secondary distinction we draw here differs from that of Nilsen, in which all implements are secondary in that they are always controlled by a primary body part). Both primary body parts and secondary aids are Instrument types, and this can be seen in the acceptable result of coordinating them in the *use-for* paraphrase:

(13) a.　I use my one good eye and a loupe for seeing the fine print.
　　　b.　I use my weak legs and a cane for hiking.

As for the impossibility of an *eat* sentence parallel to those in (12): eating and drinking, unlike walking or hearing, are not readily associated with a body part that is designed for and functions as the primary Instrument of these actions. For this reason, ISCs equivalent to those in (12) seem infelicitous: *My weak system couldn't eat the meat*; *My sore throat drank all the water.* By contrast, the following (Web) examples are fine: *Her stomach could not digest meat. My thirsty throat swallowed my first gulp.* The reason is that digesting and swallowing do have accessible primary Instruments. Along the same lines, we suggest regarding the Material argument of creation verbs like *knit* and *paint* as a secondary aid to the primary implement. Nilsen similarly considers Material as an Instrument. (14) shows that such creation verbs, just like the verbs in (13), select for both types of Instruments (and allow for their conjunction: *I knit sweaters only with the best needles and the finest lambswool*). And, similarly, only the primary Instrument can be the subject of an ISC, as shown in the contrast between (15) and (16):

(14) a.　I knitted the sweater with this lambswool / with these special needles.
　　　b.　I painted all these pictures with acrylic / with an old brush.

(15) a.　*This lambswool knitted that lovely sweater.
　　　b.　*Acrylic painted all these pictures.

(16) a.　These special needles knitted that lovely sweater.
　　　b.　An old brush painted all these pictures.

Thus, we can account for some classic ISC questions with the following constraint:

(17)　　　Constraint on ISC: The ISC subject must be a primary Instrument.

4　The Semantics of the Verb: a Brief Note

Brousseau & Ritter (1991) assert that the ISCs of Instrument verbs contain a variable for Instruments that is directly responsible for the action. For example, verbs like *cut* and *drill*, which form ISCs, entail the use of a specific Instrument: one can neither cut nor drill without an appropriate Implement. Yet the instrumental cause/change verbs do not, in fact, always form good ISCs, as (4b) and (5b) show. Moreover, ISCs exist with verbs that do not require a typical Implement, as (12) shows. As for a finer distinction, consider the following:

(18) a.　　A Bunsen burner melted the chocolate.= (2b)
　　　 b.　　A rock broke the window.

The verbs *melt* and *break* do not have a device or Implement Instrument as part of their meaning. Rather, they have a Means meaning component (as argued in Erteschik-Shir and Rapoport 2004, etc.), specifically "heat" and "forceful" respectively. Any element that can be used to realize these Means is a possible subject and, unlike the case with an Implement meaning component, does not require the presence of an Agent wielder. Thus, alongside the unacceptable (4b) and (5b), we have (19) below:

(19) a.　　A broom broke that chair.
　　　 b.　　The (laser) pen melted the wax.

Here, *a broom* and *the laser pen* realize the Means components of their respective verbs. Neither of these phrases is necessarily interpreted as a manipulated implement and therefore no Agent is entailed. Thus, it is not the type of Instrument alone that determines the acceptability of an ISC, but a combination of the Instrument type and the verb type. But this is still only a partial explanation of ISCs.

5　Information Structure

We argue that a successful account for the distribution of ISCs must be formulated in terms of context and Information Structure. This is particularly relevant to those ISCs with the verbs that entail the use of Implements, which usually require (overt) Agent manipulation. Examining the ISCs with verbs entailing specific Instruments, we are able to distinguish two types of sentences: generic ISCs as in (20) and eventive ISCs as in (21).

(20) a.　　This pen writes smoothly.
　　　 b.　　Our natural fiber broom sweeps large areas well.
　　　 c.　　Big needles knit loopy sweaters.
　　　 d.　　This knife slices Italian bread beautifully.

The sentences of (20) have the flavor of generic Theme-subject Middles: a property of the Instrument subject is understood as responsible for the action of the verb. Yet when these same elements occur in an episodic sentence, the result is not felicitous:

(21) a　　 *The pen signed the letter.
　　　 b.　　*The broom swept the floor.
　　　 c.　　*The needles knitted the sweater.
　　　 d.　　*The knife sliced the salami.

But it is not simply the case that, with Implement subjects, generic ISCs are good and eventive ISCs are bad. Contrast the sentences in (21) with those in (22):

(22) a.　　This pen signed the Declaration of Independence.
　　　 b.　　The needles I inherited from my grandmother knitted the baby sweaters.
　　　 c.　　This knife sliced my finger; that knife sliced the salami.

Consider also the contrast between *The knife killed the victims* and *Ladies and gentlemen of the jury, I will show that this knife killed the victims*. The second sentence is an example of a "courtroom" context, as discussed by Wojcik (Linguist List 4.445).

It seems that ISCs referring to actual, rather than generic events, are felicitous when they highlight a specific Instrument, either by contrast: e.g. with other Instruments, by negation, or with respect to a specific event. The key here is that in the felicitous examples, the context in which the Instrument is used is that of a known event. We know that someone signed the Declaration of Independence (with a pen), for instance; we know (in the courtroom scene) that someone has been killed with a knife. The Instrument, then, is not being introduced "out of the blue". In fact, it is hard to get an ISC with an out-of-the-blue, or an all-focus, sentence:

(23) a. So what happened yesterday?
         We opened doors. / We sliced salami.
     b. So what happened yesterday?
         *The key opened doors. / *The knife sliced salami.

We can also see the effect context has on a classic contrast (noted in Schlesinger 1989):

(24) a. *A rifle wounded the president.
     b. Two bullets wounded the President.

Compare (24a) to the following:

(25)   This old, cracked rifle wounded seven presidents.

A rifle, as with other agent-manipulated Instruments, can occur as an ISC subject, given the right contextual conditions. The bullets, on the other hand, are a device, a substance that once set in motion by an Agent, performs autonomously; like the Instruments in part B of our typology, they are therefore not subject to the same constraints.

As a further illustration of the role of context, consider the following minimal contrast:

(26) a. I gave Mary the key, **she opened the door**, and we went in.
     b. *I gave Mary the key, **it opened the door**, and we went in.

Contrast (26b) with the much better sentences of (27):

(27) a. I gave her the key, it wouldn't open the door, so we went away.
     b. I gave Mary a *new/different* key, *it* opened the door, and we entered.

Saying that a key opened the door, in most contexts, is vacuous. In (26b), *it opened the door* is equivalent to the uninformative 'This key opens doors'. In (27b), on the other hand, new, contrasting information is provided.

In general, it seems that when the Instrument assumes enough significance via its properties, it can be in subject position in lieu of the Agent that typically occupies this position. The context, whether discourse or actual, already introduces the existence or the possibility of an Instrument; it has already determined that there is something significant about the Instrument (and recall Lakoff's 1977 observation that in ISCs the Instrument assumes «responsibility» for the event). Thus, while the ISC apparently makes the Instrument a focal point, it is not, in fact, a Focus. As evidence, consider the following test for Topic (Reinhart 1981, Erteschik-Shir 1997):

(28) Topic test: Q: Tell me about this pen. / What i's special about this pen?
                    A: This pen/It signed the Declaration of Independence.

Note, too, that instrument subjects do not do well as answers to *wh*-questions, a test for Focus (Erteschik-Shir 1997, etc.):

(29) Focus test: Q: What signed the Declaration of Independence?
                    A: *This pen signed it.

But the instrument subject does offer a fine answer to a restricted, *which*-question, since it picks from a known set ("topic set" Erteschik-Shir 1997), provides another test for Topic:

(30) Q: Which of these writing implements signed the Declaration of Independence?
     A: THIS pen signed it.

The fact that the Instrument subject is a Topic falls out from its appearance only in those contexts where the possibility of an Instrument has already been introduced, in relation to a particular event. Thus, the subjects of (22), like the subjects of the well-formed Instrument Middles of (20) (and Theme Middles as well), are Topics. Assuming, then, that every well-formed sentence must have both Topic and Focus (Erteschik-Shir 1997), the Focus must be provided for each ISC. In fact, this is what has been argued for Theme Middles: Fellbaum (1985) states that Middles must be informative (hence, the importance of the adverbial, negation, or a verb referring to specific manner). Rapoport (2011) adopts this view, arguing that Middles need a Focus.

We therefore propose that for an ISC to be well-formed it must either contain a contrast (contrastive focus), or the action attributed to the Instrument subject must provide new information. In the case of generic ISCs like (20), the Instrument's ability to perform an action in a particular way provides this focus. Without this focus, Instrument Middles, like Theme Middles, are not felicitous:

(31) This pen writes *(well).
     This razor shaves *(smoothly).

A pen being able to write and a razor shaving are not unusual enough events to merit a statement. But a particular pen writing as opposed to another (*THIS pen writes*), negating the pen's inherent function (*This pen DOESN'T write*), or specifying a particular pen's properties (*This pen writes WELL*), are informative and thus licit statements.

6 Conclusion

We have offered an account for the constraints on the Instrument Subject Construction and those on its subject. As a part of our analysis, we have proposed a typology of Instruments in terms of degree of Agent involvement *vs.* degree of Instrument control over the event and have drawn a key distinction between primary and secondary Instruments as well as between types of generic and eventive ISCs. The ISC offers a distinct perspective on an event that could otherwise be described by an Agent-subject sentence. As Delancey (1991) puts it, the reason that the Instrument, rather than the Agent, is chosen as subject is a pragmatic reason of 'staging'. Noting the significance of the Instrument in the particular context given, we have demonstrated the Information Structure principles that underlie the particular staging of an

ISC. Doing so, we have attempted to answer the question: Why can Instruments usurp the position usually reserved for Agents?

Acknowledgments

Bibliography

Brousseau, A-M. & E. Ritter. 1991. A non-unified analysis of agentive verbs. *Proceedings of the Tenth West Coast Conference on Formal Linguistics.*

Delancey, S. 1991. Event construal and case role assignment. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*.

Erteschik-Shir, N. 1997. *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.

Erteschik-Shir, N. & T. Rapoport. 2004. Aspectual focus: Bare aspect. In J. Gueron & J. Lecarme (eds.), *The Syntax of Time.* Cambridge: MIT Press.

Fellbaum, C. 1985. Adverbs in agentless actives and passives. In W. Eilfort et al. (eds.), *CLS 21, Papers from the parasession on causatives and agentivity*. Chicago Linguistics Society.

Fillmore, C. J. 1968. The Case for Case. In Bach & Harms (ed.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston.

Lakoff, G. 1977. Linguistic Gestalts. In *Papers from the Thirteenth Regional Meeting of the Chicago Linguistics Society*.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago: University of Chicago Press.

Marantz, A. 1984. *On the nature of grammatical relations*. Cambridge, Mass.: MIT Press.

Nilsen, D. L. F. 1973. *The instrumental case in English; syntactic and semantic considerations*. The Hague: Mouton.

Rapoport, T. 2011. *Ms. The Disappearing middle: defining the middle to nowhere*.

Reinhart, T. 1981. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27.

Rissman, L. To appear. Periphrastic use: a modal account of instrumentality. *Proceedings of the 41st Annual Meeting of the North East Linguistics Society*.

Schlesinger, I. M. 1989. Instruments as agents: On the nature of semantic relations. *Journal of Linguistics* 25:1.

Schutze, C. T. 1996. PP attachment and argumenthood. *MIT Working Papers in Linguistics* 2.

Van Oosten, J. 1977. Subjects and agenthood in English. In *Papers from the Thirteenth Regional Meeting of the Chicago Linguistics Society*.

Wojcik, R. 1976. Where do instrumental NPs come from? In M. Shibatani (ed.), *Syntax and Semantics* 6. New York: Academic Press.

Lexemes with Inexpressible Semantic Valencies and Their Representation in the Dictionary[1]

Irina V. Galaktionova

Moscow State University
ig@philol.msu.ru

Abstract

The paper deals with Russian verbs *grabit'/ograbit'* 'to rob', *obvorovyvat'/obvorovat'* 'to burgle' that are usually considered to have an inexpressible (or blocked) semantic valency of the object taken by the robbers. Some forms that are governed by these verbs and supposedly act as the means of expressing this valency are analyzed. It is shown that a dependent in the *na* + accusative form (*ograbit' (bank) na million dollarov* 'rob (a bank) for a million dollars') used with all these verbs does not realize this valency but contains a description of the taken object. The paper looks closely at the word combinations like *grabit' nagrablennoe* 'steal the loot [rob the stolen]' which are possible only for the verb *grabit'*, and considers this valency to be filled here; thus the verb *grabit'* has to be regarded not as a word with an inexpressible valency but rather as a word that has two government patterns, where the valency of the taken object is expressible in only one of them.

Keywords

Inexpressible semantic valency, government pattern, lexicography

1 The Correspondence between Semantic and Syntactic Valencies

One of the questions raised within the Meaning↔Text Theory is the correspondence between the semantic and the syntactic valencies of a word (or rather a word that is regarded in its exact meaning, i.e. as a lexeme) and, in other words, the question of the morphological expression of its semantic valencies. A complete description of a lexeme in a dictionary implies the inclusion of such information into the dictionary entry, usually into a special section called "government pattern". The Explanatory Combinatorial Dictionary of Modern

---

Russian and the Active dictionary of Russian language that is currently being developed under the guidance of Ju. D. Apresjan and is described in Prospect Aktivnogo slovarja russkogo jazyka (2010) are examples of dictionaries containing this type of information.

It is known that there are some words in Russian language, whose (one or several) semantic valencies cannot be expressed by their own complements. In (Apresjan 1974: 146-149) examples of lexemes are given that demonstrate this feature (*grabit'* 'to rob', *obvorovat'* 'to burgle', *páxnut'* 'to smell' and others), as well as the examples of semantic lexeme groups where it appears to be systematic (verbs that describe reversal of a result of an action like *razdet'* 'to undress', *razgruzit'* 'to unload' and others). This phenomenon is also discussed in (Iordanskaja, Mel'čuk 2007). Some of such cases are examined in (Plungjan, Raxilina 2005). Ju. D. Apresjan in his recent paper (Apresjan et al. 2010) once again appeals to these already known facts, and also introduces some new lexical material. In addition to the words that absolutely cannot express their valency by means of their own complements, there are also some that can, even if it is uncommon for them. The actively discussed verb *promaxnut'sja* 'to miss' is an example of such a case, see (Plungjan, Raxilina 2005), (Percov 2006).

In this paper we will examine some of these lexemes that are supposedly unable to attach the complement that corresponds to their semantic valency (and not the ones where this valency remains unexpressed on a regular basis). We will try to answer the question: is this valency indeed inexpressible? It is especially important for the lexicographic representation of these words. In particular, it would define the format in which their government patterns are to be described. This question is important to me because, as a member of a team of authors, I work on the dictionary entries for the Active dictionary of Russian language, whose theoretical basis was developed by Ju. D. Apresjan in (Prospect 2010).

2 Inexpressible Valencies and Their Lexicographic Representation in the Active dictionary

Ju. D. Apresjan calls the inexpressible valencies *latent*, "meaning that they are expressed indirectly in the language" (Apresjan et al. 2010: 364). In his paper he gives two typical manifestations of latent valencies, illustrating them with examples.

The first one is the possibility to be attached to the verbs acting as lexical functions or having a similar meaning, on which the word with such a valency depends. This feature was noted by I. A. Mel'čuk, see (Iordanskaja, Mel'čuk 2007). For example the "mark" valency of the noun *ekzamen* 'exam' is not expressed by its own complement, but such a complement can be attached to the verb that governs *ekzamen*: *sdat' ekzamen na trojku* 'pass an exam with a C', *polučit' trojku na ekzamene* 'get a C on an exam'.

The second type of manifestation is the presence of a derivative noun that denotes the actant of the situation corresponding to this valency. For example, the verbs *skladyvat'* 'to add' and *vyčitat'* 'to subtract' have the "result" valency which is not expressed by means of their own complements, while a corresponding actantial noun exists: for the given examples these are *summa* 'sum' and *raznost'* 'difference'.

The lexicographic techniques within the Active dictionary of Russian language are regulated by the guidlines on composing dictionary entries developed by Ju. D. Apresjan, the head of the project; see (Prospect 2010: 124).

In this dictionary the actantial nouns are described in the section "DER" ("Derivaty" 'derivatives') regardless of whether the respective valency can or cannot be filled by means of the complement. In this regard describing words with latent valencies is no different from describing regular words. Government features of these two word groups, on the contrary, cannot be described in the same way. The ability of a lexeme to attach complements is to be included in the entry for that particular lexeme. For this purpose a special Government section ("Upravlenie") is used. It contains information regarding ways of morphological expression of valencies. If expression of some valency is uncommon but possible, the means of its expression must be included into the government pattern and can be marked as rare. This is apparently the case with the majority of valencies of the verb *promaxnut'sja* 'to miss'.

Evidently, this section can only include information about expressible valencies. If a specific valency is latent, the Government section can be supplemented with an optional note stating that this particular valency is usually not expressed by means of the complement (for example valency "means" of the verb *otvintit'* 'to unscrew' in its first meaning is supposed to be described this way[2]). In accordance with the guidelines, it is also possible to put more information regarding inexpressible valencies into the Commentaries section, such as information about the possibility of attaching a complement not to the word – generally the noun – that has a variable, but to the verb by which it is governed. For example, the noun *anketa* 'questionnaire' has an inexpressible valency "people who fill out the questionnaire" (compare *anketirovat' studentov* 'to conduct a questionnaire among students'), nevertheless, we can mention these people by attaching a complement. For instance, it can be attached to the verb *rasprostranit'* 'to distribute': *rasprostranit' anketu sredi studentov* 'to distribute a questionnaire among the students', but not **anketa sredi studentov* 'the questionnaire among the students'.

We will examine only one typical example illustrating inexpressible valencies that was introduced by Ju. D. Apresjan. It is the example of the verbs *grabit'* 'to rob' and *obvorovyvat'* 'to burgle'. The verbs with the meaning of reversal of a result of an action (*razdet'* 'to undress', *rasstegnut'* 'to unzip, unfasten, unbutton' etc.) will only be touched upon.

3 *Grabit'* 'to rob' and *Obvorovyvat'* 'to burgle'

Ju. D. Apresjan characterized the verbs *grabit'* 'to rob' and *obvorovyvat'* 'to burgle', which have three semantic valencies – those of the agent, the taken object and the victim – as follows: "Verbs like *grabit'* and *obvorovyvat'* (unlike *vorovat'* 'to steal', *izymat'* 'to seize' and *konfiskovat'* 'to confiscate') always leave the valency of the taken object unexpressed; indeed, *obvorovyvat' B* = 'to steal some object X from B', and it turns out that in order to correctly

---

[2] This meaning is shown in the guidelines through the word combination *otvintit' tabličku ot dveri* 'unscrew a sign from a door', which cannot be supplemented by an indication of the means used to attach the sign, cf *privintit' tabličku k dveri šurupami* 'screw on a sign to a door with screws', but not **otvintit' tabličku ot dveri šurupami <s šurupov>* 'unscrew a sign from a door with screws <from screws>'. Of course there are correct word combinations like *otvintit' šurupy* 'unscrew the screws', but it seems that another lexeme of the verb *otvintit'* is present here. This issue will be touched upon at the end of this paper.

interpret these verbs it is necessary to mention the actant X (the taken object) which cannot be attached directly to the verb" (Apresjan, 1974: 148)[3].

This statement is true not only for the verbs mentioned but also for their perfective forms. Indeed one cannot say in Russian **grabit' <ograbit'> proxožego na šapku* 'rob a stranger of a hat' or **obvorovyvat' <obvorovat'> kvartiru na dragočennosti* 'rob an apartment for valuables'. And there are no other morphological ways of referring to the taken object using word combinations acceptable in the language.

Next we will look at the forms governed by these verbs that need to be proven not to serve as a way of filling the valency of the taken object.

3.1 The Prepositional Phrase *na* + Accusative Governed by Verbs *grabit'/ograbit'*, *obvorovyvat'/obvorovat'*

In the ungrammatical examples provided above the prepositional phrase *na* + accusative (PP$_{na+ACC}$) (and not any other, as e.g. PP$_{ot+GEN}$ – *grabit' <ograbit'> proxožego ot šapki*) was used on purpose.

Of all the unacceptable forms this PP is distinct in two ways. Firstly, it is the very form that is present in the government pattern of a playful colloquial lexeme *ograbit'* 'to borrow or take', which is used only in the perfective and was described in the Active dictionary by M. Ja. Glovinskaja. For example: *Možno tebja ograbit' na paru tysč?* 'Can I rob you of a couple of thousand?'; *Ja vas ograblju na zakrutočku* (L. Ginzburg) 'Can I rob you of a cigarette'[4]. Secondly, there are some contexts where these verbs – in particular, *grabit'/ograbit'* – govern the PP$_{na+ACC}$.

Let us provide a few examples from the Russian National Corpus (it is necessary to mention that in the Corpus one can come across such word combinations only in texts – mainly the news articles – dated before the middle of the XX century).

(1) *Kladovaja Simbirskogo banka byla **ograblena na dva milliona rublej*** (N. Panov, 1935- 1950)
'The strongroom of the Simbirsky bank was robbed of two million roubles'.

(2) *Manufakturno-produktovyj sklad trjox lesopil'nyx zavodov **ograblen tysjač na desjat'** šajkoj banditov* (V. Šiškov, 1913-1932)
'The warehouse of three wood-sawing plants, where textiles and food items were kept, was robbed of ten thousand by a gang of criminals'.

---

3   I. A. Mel'čuk also notes that the verb *(o)grabit'* "does not admit the expression of the Thing Taken by the robbers: *Ivana ograbili *šuboj <*ot šuby, *na šubu, …>*, lit 'They robbed Ivan of his fur coat'" (Iordanskaja & Mel'čuk, 2007: 70).

4   The removal of the ban on valency expression when moving from direct meaning to the indirect ones was demonstrated by Ju. D. Apresjan using adjectives *skupoj* 'stingy' and *sčedryj* 'generous' (Apresjan et al., 2010: 364); cf. **skupoj <sčedryj> na den'gi* 'stingy <generous> with money', but a figurative *skupoj <sčedryj> na poxvaly* 'stingy <generous> with praise'.

(3) *Na exavšego na izvozčike vladel'ca aptekarskogo magazina Ivanova nabrosilsja s revol'verom neizvestnyj i,* **ograbiv** *ego* **na summu okolo 5 000 roubles.***, brosilsja bežat'* ("Reč", 1911)
'The owner of a pharmacy store was travelling by cab when he was assaulted by an unknown individual carrying a revolver, who, after having robbed him of 5000 roubles, had escaped'.

(4) *Pomnite togo parnja, gospoda, čto ubil kupca Olsufjeva,* **ograbil na poltory tysjači** (F. M. Dostojevskij, 1880)
'Gentlemen, do you remember that fellow who had killed merchant Olsufjev and then robbed him of 1500 (roubles)'.

The verb *obvorovyvat'/obvorovat'* – though rarely – also governs the PP$_{na+ACC}$

(5) *Nekto Sevostjanov nedavno esčo zateval torgovlju c Persieju železom v obširnyx razmerax, no na pervyx že porax prikazčik* **obvoroval** *ego* **na desjatki tysjač***, i delo lopnulo* (P. I. Ogorolnikov, 1873)
'A certain Sevastjanov had recently began an extensive iron trading with Persia, but was instantly robbed of tens of thousands by the counterman, and the business collapsed'.

And in modern texts there are more of such examples.

(6) *Internet-mošenniki legko grabjat banki na $ 10 tysjač v sutki* (bagnet.org)
'Internet scammers easily rob banks of $10 000 a day'.

(7) *Magazyn v Sos've obvorovali na 110 tysjač rublej* (serovglobus.ru)
'A' shop in Sos'va was robbed of 110 thousand roubles'.

A question arises: can it be said that in cases such as given in examples above, the valency of an object being taken is filled by the PP$_{na+ACC}$ attached to the verb?

In order to answer this question, we have to look first at the obvious fact that in all such cases the complement can only be expressed by a PP with a quantitative meaning like *na krupnuju <značitel'nuju, nebol'šuju> summu* 'of significant <small> sum of money', *na tysjaču rublej* 'of a thousand roubles', *na million dollarov* 'of a million dollars', *na milliardy* 'of billions'. An even if the words like *rubl'*, *dollar*, etc. are absent, a certain sum of money is implied. Word combinations like **ograbit' na šubu* 'to rob of a fur coat' or **ograbit' na dva kilogramma zolota* 'to rob of two kilograms of gold' are, as it was mentioned before, ungrammatical. Theoretically, one could justify this situation by the semantic restrictions imposed on the filling of the valency of the (taken) object.

It turns out, however, that the mentioned complements do not necessarily name the very object that was stolen. It seems like in the sentence (4) actual money in the amount of 1500 was stolen. But example (2) tells about a robbery of a warehouse, where there were definitely no money stored – the robbers took some goods whose value was indicated. In the sentences where the PP *na takuju-to summu* 'of such-and-such sum of money' is used, it is always implied that it is not the money (or at least not the money alone) that was taken. In other cases it remains unclear (and is seen only from the context) what it is that was actually stolen, just as it is in sentence (3).

Here is one typical example consisting of a headline and a beginning of a news article taken from the Web.

(8) *Parižskij muzej ograbili na 500 millionov evro. Neizvestnye poxitili pjat' kartin Pablo Pikasso, Anri Matissa i drugix izvestnyx xudožnikov v noč na četverg v galeree iskusstva v Pariže, stoimost' poxisčennogo dostigaet 500 millionov evro* (solovei.infol).
'A museum in Paris was robbed of 500 million euros. Certain unknown individuals stole five paintings by Pablo Picasso, Henri Matisse and other famous artists on the night between Wednesday and Thursday in the gallery of arts in Paris, the value of stolen art pieces amounts to 500 million euros'.

In this example – as well as many others – the content of the message clearly says that it is not the money that was stolen. Even if when it is about a bank robbery (example (1)), other valuables, apart from actual bank notes, can act as stolen objects.

Thus the PP$_{na+ACC}$ points to a quantitative estimation of stolen goods expressed monetarily, rather than the object itself. Therefore, this form cannot be regarded as a way of expressing the correspondent semantic valency of the verb *grabit'/ograbit'*.

Now let us look at the common usage of this dependent with the perfective verbs *ograbit'* and *obvorovat'*. Even if the verb used is in imperfective (example (6)), it can attach the discussed form only in case it points to a past event, meaning that one can already calculate the loss. Meanwhile, the object itself is directly involved in the situation which is named by the verb *grabit',* and if the PP$_{na+ACC}$ was expressing the object it would be easy to attach it to the verb when describing a directly observed situation: **Karaul, menja grabjat na tysjaču rublej!* 'Help! I'm being robbed of a thousand roubles!' which in reality is impossible.

There are other words in Russian language that can attach a PP like *na summu v neskol'ko tysjač* 'worth a few thousand', *na sto tysjač rublej* 'worth 100 000 roubles'. For example, the verb *kupit'* 'to buy' is used in word combinations like *kupit' tovara <saxara> na 500 rublej* 'to buy 500 roubles worth of goods <sugar>'. We can assume that two valencies of the verb are being filled here: that of a purchased product and that of the sum of money given to the seller. And it is the second valency that is being filled by the PP$_{na+ACC}$. However the very same PP is present in contexts where it is obviously governed not by a verb but by a noun. For example: *Na etom sklade xranjatsja tovary <produkty, lekarstva, tkani> na mnogie milliony rublej* 'This warehouse stores goods <groceries, medication, fabric> amounting to millions of roubles'. There are also nouns that use such means to fill one of their semantic valencies: *dolgi po zarplate na ogromnuju summu* 'salary debts amounting to a huge sum', *zajmy na obsčuju summu v million rublej* 'loans amounting to a million roubles'.

The verbs *grabit'/ograbit'* and *obvorovyvat'/obvorovat'* as opposed to, say, the verb *kupit'*, clearly do not have a semantic valency of a sum of money that could be filled by the PP$_{na+ACC.}$ It is also unreasonable to assume that the ability to attach such a complement is determined by grammatical (in terms of Ju. D. Apresjan) features of these verbs. It seems that systematic approach to describing this PP means admitting that it semantically belongs to the actant that cannot itself be attached to these verbs.

And regarding the verbs at hand, it means that the PP$_{na+ACC}$, which is syntactically governed by them, depends semantically on the word that denotes the taken object and is unable to

attach itself to these verbs. This may serve as an additional proof that they do possess the semantic valencies of the taken object. In other words, such dependents can be considered another manifestation of latent valencies.

What does it all mean for the lexicographic representation of the discussed verbs? It is evident that their valency of the taken object is indeed inexpressible by the PP$_{na+ACC}$, and, therefore, this PP cannot be included into the government pattern as a way of filling this valency. At the same time, the type of word combinations discussed here should be mentioned in the dictionary entry, if not in the Government section, then in the Combinatory section that contains conventional word combinations. Unfortunately, the dictionary format does not allow for such cases to be accompanied by proper commentaries.

3.2 The Accusative Noun Phrase with the Meaning of the Taken Object Governed by the Verb *grabit'*

The verb *grabit'* (but neither *ograbit'* nor *obvorovyvat'/obvorovat'*) demonstrates another type of word combinations, that give us an opportunity to discuss the question of inexpressibility of object valency. These are combinations of this verb with an accusative noun phrase (NP$_{ACC}$) such as *grabit' nagrablennoe* 'steal the loot [rob the stolen]', *grabit' nacional'nye bogatstva* 'steal [rob] national wealth'.

All these verbs, including *grabit'*, freely use the NP$_{ACC}$ to refer to a third participant – either the person that is being deprived of something or the place where this something is located (for the sake of convenience let us call this participant "the owner"): *grabit' proxožix <karavany, korabli>* 'rob strangers <caravans, boats>', *ograbit' bank <gorod>* 'rob a bank <a town>', *obvorovat' kvartiru* 'burgle an apartment', *obvorovyvat' magaziny* 'burgle shops'. But in the examples *grabit' nagrablennoe, grabit' nacional'nye bogatstva* NPs specifically refer to the taken object.

The Russian National Corpus gives a considerable number of examples of the NP$_{ACC}$ with this meaning. The most common (71 contexts) is undoubtedly the word combination *grabit' nagrablennoe* 'steal the loot [rob the stolen]' which is found in texts beginning from 1917 and up to the modern day and is primarily used as a quote. For example:

(9) *Oni tverdo uvereny, čto postupajut ne beznravstvenno, ibo ubeždeny v tom, čto znamenityj leninskij lozung "Grab' nagrablennoe!" prizyvaet ix k vpolne zakonnomu peredelu sobstvennosti* (Ju. Azarov, 2002)
'They firmly believe themselves not to be acting immorally as they are convinced that the famous Lenin's slogan "steal the loot" calls for a fairly legitimate redistribution of property'.

All the other occurrences from the Corpus are found mainly in texts dated to the XIX century. Examples of more recent occurrences are rare. On the whole, the most common is the word combination *grabit' imusčestvo* 'steal [rob] the property' (22 contexts), for example:

(10) *Zanjav xutor, načali grabit' imusčestvo kazakov* (M. A. Šoloxov, 1928-1940)
'Having occupied the farm, [they] began stealing Cossacks' property'.

There are also word combinations like *grabit' sobstvennost' <požitki, zapasy, dobro, (č'jo-l.) dostojanie>* 'rob the property <belongings, supplies>', *grabit' den'gi <sokrovisča, tovary, utvar'>* 'rob the money <treasure, goods, homeware>', *grabit' čužoe* 'rob someone else's'.

The less specific the noun's meaning, the more natural the word combination sounds. For example *grabit' (muzejnye) cennosti* 'steal [rob] the (museum) valuables' is much better than *grabit' den'gi* 'steal [rob] the money', the latter sounds like an archaism[5].

In modern Russian the verb *ograbit'* cannot be used this way, but only a hundred years ago it was possible:

(11) *Fomovskij zajavil policii, čto <…> on podvergsja napadeniju so storony trjox neizvestnyx zloumyšlennikov, kotorye **ograbili** u nego **pasport i košelek** s den'gami i skrylis'* ("Južnaja kopejka", 1912)
'Fomovsky told the police that he was attacked by three strangers who took [robbed him of] his passport and money and left'.

The most notable are the contexts where, alongside the taken object, its owner is also mentioned (and sometimes – just the owner). Such contexts can both be found in the Corpus and on the Internet, including news articles; see examples (12), (13), (14).

(12) *Ograblen on ne byl, da i **grabit' u nego** bylo nečego; p'jan tože ne byl* (L. K. Čukovskaja, 1978)
'He was not robbed, and there was nothing to rob him of; neither was he drunk'.

(13) *Zastrojsčiki v Sankt-Peterburge **grabjat imusčestvo** graždan **v** snosimyx **garažax*** (fleksagon.ru)
'Real-estate developers in Saint-Petersburg steal the citizens' property in the garages under demolition'.

(14) *Zaderžany dvoe zloumyšlennikov, kotorye **grabili imusčestvo** iz avtomobilej* (nikinform.com)
'Two criminals who were stealing property from cars were arrested'.

This kind of a government pattern, where NP$_{ACC}$ points to the taken object, and PP$_{u+GEN}$ (if it'is a person) or PP$_{iz+GEN}$, PP$_{v/na+LOC}$ (if it is a place) are used to denote the owner – in the modern language is no longer valid for the perfect *ograbit'* but is still present in the imperfect *grabit'*, while more lexical restrictions are imposed on filling the first of these two valencies.

This government pattern is archaic and is not quite standard, but probably due to the influence of such verbs as *krast'* 'to steal', *vorovat'* 'to thieve', which are very similar in meaning, and also due to the widespread slogan "*Grab' nagrablennoe!*" it still remains in Russian. It is possible that in the foreseeable future it will be lost in *grabit'*, as it happened with *ograbit'*.

---

[5] V. I. Belikov's answer to the question: "What can be robbed?" and a discussion on the topic can be found here: http://forum.lingvo.ru/actualthread.aspx?tid=54051.

Having looked at all these facts, can we say that the discussed word combinations prove that the verb *grabit'*, as opposed to *ograbit'* and *obvorovyvat'/obvorovat'*, actually expresses the valency of the taken object with its own dependent? There are at least three answers to this question.

Firstly, one could just consider all such word combinations as phraseological units, due to their lexical restrictions. However, since different words can act as dependents and because the semantics of these word combinations is compositional, this solution is only acceptable as a purely technical one, the one which is convenient when it comes to describing, but which does not get to the root of the problem.

Secondly, one could presume that in the cases like *grabit' imusčestvo* we are just talking about a different lexeme of the verb *grabit'*. In this case, the lexeme examined earlier does indeed have an inexpressible valency, while in the new lexeme the valency of the taken object is expressed by the NP$_{ACC}$.

And finally, it is possible that in word combinations *grabit' proxožix* 'rob the strangers' and *grabit' imusčestvo* 'rob the property' it is the same lexeme (all of the Russian dictionaries as well as the mentioned entry by M. Ja. Glovinskaja describe it this way). In this case, all the valencies of the verb *grabit'* in that particular meaning are, in theory, expressible, but the verb has two government patterns. In the first, the main one *(grabit' proxožix),* only the valency of the owner but not that of the taken object is expressed. In the second model, which is rare and probably resides on edge of the norm (or beyond it?), both valencies are expressible, but usually only one, that of the taken object, ends up being expressed while the owner's valency either remains unexpressed or is attached to the word denoting that object: *grabit' imusčestvo kolxoznikov* 'rob the farmers' property'.

There seems to be no reason to assume that the described word combinations contain two different lexemes of the verb *grabit'*. The differences in government do not correlate with any other differences, apart from that of either presence or lack of an aspectual pair: in all the cases the verb describes the same situation with the same three participants, and lack of an aspectual pair in word combinations like *grabit' imusčestvo* can be regarded as a transitional step in the verb's evolution. Thus the existence of word combinations like *grabit' imusčestvo* probably signifies that the verb *grabit'* cannot act as an example of a word with a completely inexpressible valency.

In lexicographic description of this verb one has to somehow locate its ability to govern a NP$_{ACC}$ denoting the taken object. The easiest way is to include a second government pattern in the entry, mark it as rare and point all lexical restrictions imposed on it. Choice of a specific way of describing the discussed word combinations would depend on the dictionary's format. One thing is clear: at least the mere fact of their existence has to be mentioned in the entry.

The verb *grabit'* demonstrates ambiguousness of the NP$_{ACC}$ governed by it. This ambiguousness is often seen in a variety of other verbs that are supposed to have inexpressible valencies. For example, the verb *zastegnut'* 'to button' can express valencies of both the first and the second objects: *zastegnut' kurtku na vse pugovicy <na molniju>* 'button all the buttons <the zipper> of the jacket' (in Russian both *kurtka* and *vse pugovicy <molnija>* are attached to the verb), while with *rasstegnut'* the valency of the second object is inexpressible, compare incorrect **rasstegnut' kurtku ot pugovic <s molnii>* 'unbutton all the buttons <the zipper> (and) the jacket' with correct *rasstegnut' kurtku* 'unbutton the

jacket'. However, in the same situation one could say *rasstegnut' vse pugovicy <molniju> na kurtke* 'unbutton all the buttons <unzip the zipper> of the jacket', where NP$_{ACC}$ specifically refers to the second object. Similar government pattern is seen in many other – though not all – verbs which mean reversal of an action's result. The question of whether such examples can be considered evidence of expressibility of the second object's valency demands a separate discussion. It is closely related to the question of the number of lexemes in a given entry and is thus directly connected to the real-world lexicographic practices.

For example, the earlier mentioned verb *otvintit'* 'to unscrew' has two such lexemes. The first one can be found in the word combination *otvintit' tabličku ot dveri* 'unscrew a sign from the door'. It has an inexpressible valency of the means and corresponds to the verb *privintit'* 'to screw on'. The second lexeme appears in the word combination *otvintit' šurup* 'unscrew a screw' where *screw* is not the means but an object, and this lexeme corresponds to the verb *zavintit'* 'to screw down' and has no inexpressible valencies.

4   A Few Words in Conclusion

The verbs covered or just touched upon in this paper are of a great theoretical interest with regard to the problem of inexpressible valencies.

The dictionary representation of these words, without a doubt, implies taking into account all the aspects of their behaviour. The amount and the kind of information to be included into a dictionary entry is defined by the size and the type of that particular dictionary, its theoretical foundation and the guidelines on composing dictionary entries. In case of the Active dictionary of Russian language, all of the above was defined by Ju. D. Apresjan.

The dictionary author often wants to convey all the aspects of a described word, which sadly turns out to be impossible, because of a particular dictionary's format. This paper provides an attempt to describe the intricacies concerning the verbs *grabit'/ograbit'* and *obvorovyvat'/obvorovat'* in just one aspect of how two of their three semantic valencies are being filled.

A real description of this kind of words in the Active dictionary will become available once the first issue of this dictionary, containing an entry for *grabit'* by M. Ja. Glovinskaja, is out. This issue is currently being prepared for publication.

Bibliography

Apresjan, Ju. D., I. M. Boguslavskij, L. L. Iomdin, V. Z. Sannikov. 2010. *Teoretičeskie problemy russkogo sintaksisa: Vzaimodejstvie grammatiki i slovarja*, Moscow: Jazyki slavjanskix kul'tur.

Iordanskaja, L. N., I. A. Mel'čuk. 2007. *Smysl i sočetaemost' v slovare*, Moscow: Jazyki slavjanskix kul'tur.

Percov, N. V. 2006. K suždenijam o faktax russkogo jazyka v svete korpusnyx dannyx, *Russkij jazyk v naučnom osvesčenii*, № 1(11): 227-245.

Plungjan, V. A., E. V. Raxilina. 2005. *Prizemlit'sja* i *promaxnut'sja*: semantičeskie mexanizmy syntaksičeskix ograničenij. In *Vostok – Zapad: Vtoraja meždunarodnaja konferencija po modeli "Smysl↔Tekst"*, Moscow: Jazyki slavjanskix kul'tur.

Prospect Aktivnogo slovarja russkogo jazyka. 2010. Moscow: Jazyki slavjanskix kul'tur.

Negation and Valencies of Russian Verbal Predicates

Leonid Iomdin (1), Boris Iomdin (2)

(1) Laboratory of Computational Linguistics
A.A.Kharkevich Institute for Information Transmission Problems,
Russian Academy of Sciences
Moscow, Russia
iomdin@iitp.ru

(2) Theoretical Semantics Department
V.V.Vinogradov Institute of Russian Language,
Russian Academy of Sciences
Moscow, Russia
iomdin@ruslang.ru

Abstract

The paper discusses the semantic interaction of the negation with certain types of verbal predicates in Russian, which involves, depending on the predicate type and its main valency structure, the emergence of new semantic valencies: the valency of the missing distance, the valency of the missing time span, and the valency of the missing quantity. The first type of valency was discovered by Juri Apresjan for accomplishment verbs like *dobežat'* ≈ 'run up to' as in *On dvadcat' šagov ne dobežal do ukrytija* 'he did not run the last twenty steps to the shelter'. We report that the phenomenon is of a much broader nature and range: other types of predicates may develop the valency of the missing time span, as in *Ona ne doučilas' v universitete dva goda* ≈ 'she did not complete her education at the university, dropping out two years earlier', the valency of missing quantity, as in *Emu nedoplatili tysjaču rublej* 'they underpaid him a thousand roubles', or, in sophisticated cases, more than one such valency.

Keywords

Valency structure, negation, ambiguity, semantics, lexicography

1   Introductory Remarks

Several years ago, Juri Apresjan drew the linguists' attention to some remarkable cases of valency structure change in Russian accomplishment verbs with the *do-* prefix. He showed that in the context of an overt or implicit negation such verbs may develop an additional semantic valency, "the remaining distance to the end point", as in *On ne dobežal do dereva vsego pjat' šagov* ≈ 'He did not run the only remaining five steps to the tree', or *Emu ostavalos' dobežat' do ukrytija pjat' metrov, kogda razdalsja vystrel* ≈ 'The distance he had to run yet was five meters when the shot rang out' (Apresjan, 2005: 9; 2006: 26). This valency seemed not to exist beyond the negative context. Even though sentences like *On dobežal pjat' šagov do dereva* ≈ 'He ran the five steps to the tree' are possible, the phrase *pjat' šagov* 'five steps' does not describe the distance to the end point remaining at the moment of speech, when this distance is zero.

In what follows we will try to demonstrate that Apresjan's observations reflect a more general phenomenon. It involves a broad class of lexical units by no means exhausted by the verbs mentioned above.

2   Valency Structure of Verbal Complexes Containing a Negation

2.1   Verbs with the *do-* Prefix

We will first look at other verbs with the *do-* prefix that convey the meaning of completing the action or process and see how they interact with the negation. The examples given throughout the paper are mainly taken from the Russian National Corpus (www.ruscorpora.ru), though some of them are borrowed from other Internet sources and very few are constructed by the authors.

We will start with the verb *doučit'sja* ≈ 'complete one's studies' which in our opinion behaves in a manner fairly close to the listed verbs of motion. Consider the following examples:

(1)   *V xudožestvennoj škole Vadim **ne doučilsja** odin god, tak kak ee vypuskniki ne polučali v te gody attestata i poètomu mogli postupat' tol'ko v xudožestvennye vuzy* (schools.keldysh.ru)
'In the art high school, Vadim dropped out one year earlier (lit. 'did not finish studying one year'), since its graduates were not entitled to a school-leaving certificate and could only enter art colleges'[1].

(2)   *Možno li sdat EGÈ, esli **nedoucilsja**[2] god do 11 klassa, čtoby polučit attestat?* (Otvety@Mail.ru)

---

[1]   English translations of examples are given for illustrative purposes and do not claim to be fully accurate or immaculately grammatical.

[2]   Here and below, we will not distinguish between the cases when the negative particle *ne-* and the verb following it are written as one word or two words, exactly reproducing the orthography of the source. Semantically, the two versions are identical, while the usage allows both, often showing no clear preference to either of them. The situation seems quite natural: many of the verbal complexes considered are in fact integrated lexical units for which no generally accepted orthographic pattern is available.

'May one take the centralized test if one dropped out one year earlier and not completed grade 11, in order to get a certificate?'

(1) and (2) provide convincing evidence to the fact that the verb *doučit'sja* develops, in the context of negation, an additional valency with the meaning 'time remaining to the completion of the result', similar to the respective additional valency of the remaining distance to the end point in motion verbs (and probably identical with the additional valency of the verb *dožit'* 'live up to', listed by Apresjan together with verbs of motion).

This valency, which can appropriately be called the **valency of the missing time span**, has certain features that deserve attention. First, its semantics is not directly inherited from the semantics of the verb *učit'sja* 'study', from which it is derived. The verb *doučit'sja* means something close to 'complete the studies successfully', whereas its negative counterpart *nedoučit'sja* (or, which is the same, the verbal complex *ne doučit'sja*) means 'give up studying not achieving the expected result'. In the meantime, the verb *učit'sja* does not have the valency of time span: the expression like *učit'sja god* 'study a year' is an instance of a durative construction, which belongs to the circumstantial and not to the actantial type. On the other hand, this supplementary valency of the missing time span is implemented by an NP in the accusative (or, if an explicit negation is available, the genitive): this is a rather unusual fact for Russian where such an instantiation is primarily characteristic of the object valency of the verb.

Certain other verbs with *do-/nedo-* prefixes, which develop the valency of the missing time span in the presence of a negation, exhibit similar behavior, e.g. *(ne) dolečit'sja* '(not) complete medical treatment' in (3), *(ne) dorabotat'* '(not) finish working' in (4), *(ne) doigrat'* '(not) finish playing' in (5), *(ne) doderžat'* '(not) keep enough' in (6), *(ne) dovarit'* '(not) cook completely' in (7) and *(ne) dosmotret'* 'not watch to the end' in (8):

(3) *... ja brosila pit' antibiotiki, **ne dolečilas'** dva dnja, sil net* (babyplan.ru)
     'I gave up taking antibiotics, dropped out of treatment two days earlier, I have no strength';

(4) *Vy u nas **ne dorabotali** po raspredeleniju god, kažetsja, s kopejkami?* [D. Rubina]
     'It seems that you have dropped out of obligatory work according to graduates' placement a bit more than a year earlier?';

(5) *Donetskij "Metallurg" i rossijskaja "Alanija" **ne doigrali** 15 minut*
     (novostimira.com)
     'Donetsk Metallurg team and Russian Alania did not play the last 15 minutes';

(6) *Projavljali plenku doma v poludoxlom projavitele, **nedoderžali** itogo minut 5.*
     (ofelijh.livejournal.com);
     'They developed the film at home in a developer that was half dead and underprocessed it 5 minutes on the whole';

(7) *Esli **nedovarit'** jajco xotja by minutu, ono okažetsja vsmjatku*
     'If you undercook an egg even a minute it will come out soft-boiled';

(8) *Vse rasstroilis', kogda **ne dosmotreli** fil'm do konca minut 10* (www.tourtrans.ru)
'Everyone was disappointed when they did not watch the last 10 minutes or so of the film.'

As in the case of the motion verbs, the negation need not be adjacent to the verb (or form a new verb from it). It may be implicit, as in (9), where the verb *ostat'sja* 'remain' introduces an implicit negation:

(9) *Mar'e Vasil'evne **ostalos' dorabotat'** tri nedeli* [M. Blagova]
'Maria Vasilyevna had three weeks left to work'

or refer to another verb placed higher in the syntactic tree, as in (10):

(10) *Prorabotala 39 let (mne "zlodejski" **ne** dali **dorabotat'** vsego odin god do "krugloj" daty – 40 let, do pribavki k pensii)* [V. Davydov]
'I worked for 39 years (the villains did not let me work just one more year until the good round figure of 40 years to get an increase to my pension).'

It should be noted that in principle the meaning component 'time remaining to achieve the expected result' can sometimes be expressed with the verbs considered even when no explicit or implicit negation is present:

(11) *U nas uvol'njajut 40%, dali 2 nedeli **dorabotat'*** [Internet forum on financial crisis, 2008]
'They fire 40% of our staff but they allowed us to work 2 more weeks';

(12) *Uvereny, čto Xiddink **dorabotaet** dva kontraktnyx goda?* [Sovetsky Sport, 2006.04.13]
'Are you sure Hiddink will see out the two years of his contract?'

(13) *On polučil na èkzamenax dolgoždannye trojki i so spokojnym serdcem perešel v druguju školu – teatral'nuju, gde prekrasno **doučilsja** dva goda* (http://rusakters.ru)
'He passed the exams with long-awaited C grades and moved with a calm heart to another school, the theater school where he smoothly completed his studies in two years' (literally, '…finished studying two years').

In exactly the same way, the meaning component 'distance remaining to the end point' is sometimes expressible in the absence of a negation. We already saw one such instance: *On dobežal pjat' šagov do dereva* ≈ 'He ran the five steps to the tree'. Here is another example:

(14) *Potixon'ku my **dobreli** poslednie 200 metrov do ostanovki i seli dumat', čto delat' dal'še* (http://turbina.ru).
'Little by little we dragged ourselves the last 200 meters to the bus stop and started thinking what to do next'.

We will postpone somewhat the discussion of the issue of whether or not this meaning component corresponds to a valency of the non-negated verb until we have considered a few more accomplishment verbs with the *do-* prefix and their interaction with a negation.

Let us compare sentences with the verb *dobrat'* ≈ 'gain additionally' with and without negation:

(15) *Tanja na dnevnoe otdelenie ne prošla, **nedobrala** odnogo balla, no eë začislili na večernij* [L. Ulitskaya]
'Tanja was not accepted to the regular (lit. daytime) department (of the university) as she was one credit short, but she was enrolled in the evening department'.

(16) *Tanja sdavala segodnja èkzamen i na ètot raz **dobrala** bally*
'Tanya took an exam today and this time earned all the necessary credits'.

We will now try to answer the following question: do NPs in the accusative or genitive which occupy the direct object position of such verbs instantiate the same semantic valency of the verb/verbal complex, or do they correspond to different semantic valencies? We believe that the latter is true. In particular, the semantic valency of *dobrat'* instantiated in (16) by the noun *bally* is a normal object, similar to objects of verbs denoting regular "monotonic" actions or processes (*čital knigi* 'read books', *uvidel bašni* 'saw towers'). Conversely, in (15) a different semantic valency of *nedobrat'* is present, which we propose to call the **valency of missing quantity**. Indeed, sentence (16) says that the subject made efforts to earn more credits while it says nothing about the number of credits absent and still required to achieve the result. This is however exactly what sentence (15) says.

The same is true of two more sentences with the same verb *dobrat' / nedobrat'*.

(17) *Ne budet doždja pod naliv — i srazu centnerov po sem' **ne dobereš'** na gektare* [B. Možaev].
'If there is no rain to fill the grain we will lose (lit. not harvest) about seven hundred kilograms (of crops) per hectare';

(18) *Prostite, **ne dobral** dva semestra po francuzskomu: putajus' v padežax* [A. Mikhajlov]
'Sorry, I failed to take the last two semesters in French so I confuse the cases'.

Here, *centnerov po sem'* 'about seven hundred kilograms' and *dva semestra* 'two terms' instantiate the valency of the missing quantity. It is, however, not known what total yield of crops is expected and how many semesters the course of French took.

If we proceed from our hypothesis, we will easily see that the negation affects the verb *dobrat'* in a way very similar to how it affects the motion verbs like *dobežat',* where the valency of the missing distance is added, and verbs like *doučit'sja* that get the valency of the missing time span[3]. The essential difference is, however, that in the former two cases the emerging valency is unrelated to the core valency structure of the non-negated verb, while in the latter case it is a derivative of the object valency of this verb.

To be sure, *dobrat'* is not the only verb with the *do-* prefix that can be affected by the negation in this way, which is evidenced by the following examples. In (19) through (21) it is *(ne)dosčitat'sja* ≈ 'find missing'.

---

[3] The valencies of the missing time span and the missing quantity may sometimes be difficult to discern: *On ne doučilsja v universitete dva mesjaca* 'he dropped out of university two months short of graduation' and *On ne doučilsja v universitete dva semestra* 'he dropped out of university two semesters short of graduation' are quite close in meaning. We believe, however, that the valencies are not identical, but leave the detailed discussion for the future.

(19) *Odnoj iz storon inoj raz možno **nedosčitat'sja** daže dvux pešek, ravenstvo sil ne budet narušeno* [E. Gik]
'One of the parties [in chess] may even find two pawns missing, the equality will not be violated'.

(20) *Činovniki **ne mogut dosčitat'sja** svyše 100 tysjač detej v Severokavkazskom federal'nom okruge* (http://www.webground.su)
'The officials cannot account for (lit. 'cannot count in full') over 100 thousand children in the North Caucasian Federal District'.

In both sentences, we can observe instantiations of the valency of the missing quantity (notwithstanding the fact that in (20) the negation does not directly refer to *dosčitat'sja)*. In contrast to that, in (21) we are dealing with a regular object valency.

(21) *V prošlom godu v kompanii nedosčitalis' značitel'noj pribyli* [K. Pal'šin]
'Last year, the company missed on the profits considerably (lit. 'found missing considerable profits')'

Similarly, in (22) we have the valency of the missing quantity and in (23) the normal object valency of the verb *nedoplatit'* 'underpay':

(22) *Čarujuscaja kartina bogatstva voznikla pered glazami Šlomo imenno togda, kogda emu **nedoplatili** rovno vosem'sot šekelej* [D. Rubina]
'The fascinating picture of wealth emerged before Shlomo's eyes at the very moment when they paid him exactly eight hundred shekels less'.

(23) *Tratja bol'se deneg na ènergiju, predprijatija prosto **nedoplačivajut** ljudjam zarplatu*
'By spending more money on energy, the enterprises simply underpay the people'.

Below, some more verbs with the instantiated valency of the missing quality are given: *dostavat'* 'reach' in (24)[4], *(ne)dorezat'* ≈ '(not) kill all' in (25), *(ne)dodavat'* ≈ '(not) give all' in (26), *(ne)dovesit'* '(not) give full weight, 'give short weight' in (27), and *nedopostupit'* ≈ 'come in less quantities' in (28).

(24) *Postav'te ciklamen v vederko s vodoj tak, čtoby uroven' vody **ne dostaval** 5-10 mm do verxnego kraja gorška* [B. Savelyev]
'Put the cyclamen into a bucket filled with water so that the water level is 5 to 10 mm below the upper rim of the pot (lit. does not reach 5-10 mm)';

(25) *Odna iz ošibok Ivana Groznogo sostojala v tom, čto on **nedorezal** pjat' krupnyx feodal'nyx semejstv. Esli on èti pjat' bojarskix semejstv uničtožil by, to voobsce ne bylo by Smutnogo vremeni* [Profil', 2003.06.23]
'One of Ivan the Terrible's errors was that he did not kill (all of the) five large feudal families. If he destroyed these five boyar families, there would be no Time of Troubles';

---

[4] This is a rare occasion where the negation cannot be merged with the verb with the *do-* prefix because it could be confused with a very different verb *nedostavat'* 'be insufficient'.

(26) *Vydajut saxar — iz devjati kuskov **ne dodajut** dva* [S. Grigoryev].
'If they distribute sugar, out of nine lumps they do not give two';

(27) *A prodavec obmanul pokupatelja na 1 rubl' 70 kopeek — **nedovesil** 2 kilogramma 670 grammov jablok* [Čelovek i zakon, 1978].
'The shop assistant cheated the customer by 1 rouble and 70 copecks: he gave him the short weight of 2 kilogram and 670 gram apples';

(28) *Za 9 mesjacev protiv rasčetov v bjudžet **nedopostupilo** 97 mlrd. rublej* [E. Gaidar] ≈
'In the period of 9 months, 97 billion roubles did not come to the budget as expected by calculation'.

Note that in (28) the semantic valency of the missing quality is not even generated by the direct object but by the subject of the verb *nedopostupit'*. Naturally, the general mechanism is the same.

It is extremely remarkable that the negative context is sometimes able to generate more than one valency of the missing quantity in some verbs with the *do-* prefix. Let us return to the verb *doučit'sja* and take a closer look at its semantic valencies. First of all, these are the valencies inherited from the verb *učit'sja* from which it is derived, namely (1) subject valency (*kto doučilsja* 'who completed the studies'), (2) counterpart (the teacher's) valency (*u kogo/gde doučilsja* 'with whom / where someone completed the studies'), and (3) the object valency (that of the academic discipline or subject matter: *čemu doučilsja* 'in what subject someone completed the studies'). Besides, the verb has the valency of the missing quantity, discussed above in detail, which primarily emerges in the negative context: (4) *skol'ko vremeni ne doučilsja* 'how much time earlier someone dropped out of studies'. However, this is not all.

Let us investigate the meaning of the expression *ne doučit'sja semestr*. The word *semestr* 'semester' may either fill the valency of the missing quantity *(Iz pjati let Ivan ne doučilsja semestr i brosil institut čerez četyre s polovinoj goda* 'Of the five years, Ivan missed one semester and dropped out of college after four years and a half'), or else it may instantiate one more valency (5) of the period, or a quantum, of study *(Ivan ne doučilsja pervyj že semestr i brosil institut, ne dožidajas' sessii* 'Ivan did not even complete the first semester and dropped out before the end-of-semester exams').

Now we are ready to consider one more pair of examples with *(ne) doučit'sja*:

(29) *Ona i **gramote** učilas', ne doučilas'; šit'ju učilas', portnixoj ne vyšla* [S. Sergeev-Censky]
'She learned to read and write but did not succeed (lit. 'did not complete the studies'), learned sewing but could not become a seamstress';

(30) *Mnogomu vy esce **ne doučilis'** i ko mnogomu ne privykli* [A. Kuprin] ≈
'You did not complete your studies with regard to many things and are not yet accustomed to many things'.

In (29), the word *gramota* 'ability to read and write' fills the valency of the subject matter. In contrast, the word *mnogoe* 'much; many things' in (30) does not instantiate the valency of the discipline (it is not asserted that someone learned many things). Rather, it represents the

valency of the missing subject-matter (what was not yet learned) which also emerges in the presence of a negation. The same valency is present in (31):

(31) *On ne doučilsja dvum predmetam <neskol'kim metodam, ...>*
   'he did not complete his studies in two subjects <several techniques, etc.>'

The following sentence can however be interpreted in two different ways:

(32) *No vperedi u tebja esce celaja žizn': esli ty xočes' stat' nastojascim čelovekom, to dolžen doučivat' to, **čemu nedoučilsja** v licee i čto dalos' by tebe v licee gorazdo legče.* [V. Avenarius on Pushkin's years of adolescence]
   'You have the whole life ahead of you: if you want to become a real man you have to finish learning what you could not complete in the lyceum and what would have come much easier to you in the lyceum'.

This sentence may either refer to started but unfinished disciplines (e.g. studied mathematics but did not complete the course), in which case we have the valency of subject matter, or it may refer to the untouched disciplines (e.g. did not study Latin in the lyceum, should take a course elsewhere): here is the instance of the valency of the missing subject matter.

It can now be seen that, at least with accomplishment verbs with the *do-* prefix, the negation which comes into contact with the meaning of quantity may in principle affect various verbal valencies, opening their antipodes and thus contributing to the diversity of predicate valencies and their instantiations (cf. Boguslavsky 2009). We believe that the situation is the following. When estimating the quantity or mass of elements that characterizes an action or a process at some temporal point, the speaker can be interested in three values: how many elements are activated, how many have been activated so far and how many will have to be activated yet. The latter value is most naturally highlighted by the negation; however, when verbal semantics represents an action or a process as anisotropic (this is definitely the case of the accomplishment verbs with the *do-* prefix since they are focused on the closing phase of the action or process) the negation may prove not to be compulsory: the remaining part of the elements corresponds to this closing phase.

Hence, the sentence *My **dobreli** poslednie 200 metrov do ostanovki* 'We dragged ourselves the last 200 meters to the bus stop' contains the valency of the missing distance, the sentence *On **dorabotal** zlosčastnye dva mesjaca i polučil pensiju* 'he worked the ill-fated two months that remained and received his pension' contains the valency of the missing time span and the sentence *Nam **doplatili** nakonec 500 rublej* 'they paid us the outstanding 500 roubles at last' contains the valency of the missing quantity.

2.2 Other Verbs

Even though the verbs with the *do-* prefix constitute the most typical lexical class for which the valencies of deficiencies are opened, the phenomenon is not confined to this class. Apparently, such valencies may emerge in very common situations. Let us compare the sentences

(33) *Na kontrol'noj Ivan rešil tri zadači*
   'On the test, Ivan solved three problems'

(34) *Na kontrol'noj Ivan ne rešil tri zadači <trex zadač>*
    'On the test, Ivan did not solve three problems'.

Whilst (33) is fully unambiguous, (34) has two different readings: (a) 'the number of problems solved by Ivan is not equal to three'[5] (it is not known how many problems that had to be solved Ivan did not solve) and (b) 'the number of problems unsolved by Ivan is equal to three' (how many problems he did solve is unknown).

In our view, the reading (b) of (34) is a case in which the valency of the missing quantity of the verb *rešit'* is instantiated.

The same valency is probably instantiated for the verb *net* 'there is none' in the sentence from a well-known story by Leo Tolstoy, *The Plum Stone*:

(35) *Pered obedom mat' sočla slivy i vidit, odnoj **net***
    'Before dinner Mother counted the plums and found that one was missing'.

In this sentence, the general presence of plums is not negated whilst the absence of one is asserted.

We believe that the valency of the missing quantity adequately explains the ambiguity effect in a well-known construction with the lexical unit *ne xvatit'* 'be insufficient' (which is always spelled as two words due to an orthographic whim).

(36) *Mne xvatilo sta rublej, čtoby kupit' bilet*
    'A hundred roubles was enough for me to buy the ticket'

(37) *Mne ne xvatilo sta rublej, čtoby kupit' bilet*

While sentence (36) is unequivocal, the negated sentence appears ambivalent: (a) 'a hundred roubles was not enough for me to buy the ticket' vs. (b) 'I had a hundred roubles less than was needed to buy the ticket'. In interpretation (a), which amounts to the fact that I had one hundred roubles but the ticket cost more, the NP *sto rublej* instantiates the regular object valency of the verb *xvatit'*. In intepretation (b) which implies that I had n roubles but the ticket cost $n+100$ roubles, this NP realizes the valency of the missing quantity[6].

It is worth noting that the existence of the valency of the missing quantity as opposed to the regular object valency is corroborated by certain combinatorial properties. We will list a few of them, primarily associated with the use of quantifiers (not trying to provide a semantic explanation).

In particular, the verb with the istantiated valency of the missing quantity cannot be used with quantifying words *nikogda* 'never' or *ni razu* 'not a single time'. E.g. the NP *dva balla* 'two credits', as in (38), may only be interpreted as a regular object, whereas the same NP in (39) is interpreted as one instantiating the valency of the missing quantity:

---

[5]   Most likely, this number is less than three. However, this fact is not established by the semantics of words or the sentence but with the help of logical inference, which we do not intend to discuss here.

[6]   A very close idea was proposed for the analysis of sentences like (37) in Rakhilina 2010: 328-329. It is even more clearly stated in the Rusgram project (see http://rusgram.ru/отрицание#7_4).

(38) *On nikogda <ni razu> ne dobiral dva balla (na peresdače)*
    'He never <not once> earned additional two credits (when retaking the exam)'

(39) *On vsegda <každyj raz, vsjakij raz> ne dobiral dva balla (na èkzamene)*
    'He always <every time, each time> did not earn additional two credits (when taking the exam)'

An NP instantiating the valency of the missing quantity cannot contain the numeral *oba* 'both': sentence (40) may only be interpreted as containing a regular object valency:

(40) *On ne dobral oba balla <oboix ballov>*
    lit. 'He didn't earn both additional credits'

In the same way, the valency of the missing quantity is excluded in the context of the particle *ni* ≈ 'not a single':

(41) *On ne dobral ni balla (ni odnogo, ni edinogo balla)*
    lit. 'He earned not a single additional credit'

Sentence (41) may only be understood as containing a regular object valency. In contrast, the word *skol'ko* 'how many' can only fill the valency of the missing quality for such a verb:

(42) *Skol'ko (ballov) on ne dobral?*
    'How many additional credits did he fail to earn?'

3   Valency Structure of Nominal Derivates

It is interesting to consider nominal derivates of verbs belonging to the class discussed above. In an earlier paper (Iomdin, Iomdin 2011) we reported on the ambiguity caused by uniform instantiation of different valencies in NPs like *gipoteza Sepira-Uorfa* 'Sapir-Whorf hypothesis' (the genitive NP *Sepira-Uorfa* points to the authors of the hypothesis) vs. *gipoteza Boga* 'the hypothesis of God' (the genitive NP *boga* refers to the content of the hypothesis). In the latter case, we have to do with the so-called self-derivate of the initial noun (Boguslavsky – Iomdin 2011).

The nouns formed from the verbal complexes discussed above behave in much the same way. Indeed, words like *nexvatka* 'deficit', *nedostača* 'shortfall', as well as other words of the same lexical field *(deficit* 'deficit', *rastrata* 'embezzlement', etc.) normally have two valencies: the subject valency (what is missing) and the valency of the missing quantity (how much is missing). These valencies are well distinguishable in collocations with lexical functional verbs:

(43) *Nexvatka kasaetsja* (= Func$_1$) *medikamentov*
    'The deficit affects medications'
    vs. *Nexvatka sostavljaet* (= Func$_2$) *desjat' tonn*
    'The deficit amounts to ten tons'.

The subject valency is normally expressed by a genitive NP *nexvatka prodovol'stvija* 'shortage of food', while the valency of the missing quantity is instantiated by a PP formed by prepositions *v* or, less frequently, *na (nedoimki v desjat' tysjač rublej* 'arrears of ten thousand

roubles'*, nedoplata na sto rublej* 'underpayment of a hundred roubles'), as well as an NP in the nominative, forming a so-called completive-appositive construction, as in (44):

(44) **Nedoves** *dvadcat' grammov sčitaetsja nedopustimym*
'An underweight of twenty grams is considered unacceptable'.

However, many of such nouns accept a genitive NP able to instantiate both the subject valency (45, 47) and the valency of the missing quantity (46, 48):

(45) **Nedostača** *xleba poslužila pričinoj načala graždanskoj vojny* [V. Bykov]
'The shortage of bread was the reason why the civil war began'.

(46) *V stolovoj každyj den' obnaruživalas'* **nedostača** *neskol'kix desjatkov porcij* [A. Zinovyev]
'Every day, the shortage of several dozens of rations was discovered in the canteen'.

(47) *Iz-za* **nexvatki** *denežnyx sredstv, pomescenij i kvalificirovannyx sudej reforma zatjagivalas'* [A. Afanasyev]
'Due to the deficit of money, premises and qualified judges the reform became stalled';

(48) *Gusevskoe samomnenie namnogo prevosxodilo ego ličnye gabarity, i on vsju soznatel'nuju žizn' stradal ot* **nexvatki** *pjati santimetrov rosta i desjati kilogrammov vesa* [O. Divov].
'Gusev's self-conceit much exceeded his personal dimensions, and he suffered throughout his life from the shortage of five centimeters of height and ten kilograms of weight'.

In certain cases, genitive instantiations of the two valencies may also invoke an ambiguity. For instance, *nexvatka pjati grupp tovarov* 'shortage of five groups of goods' may be interpreted as containing a subject valency ('the shortage that affected five groups of goods') or a missing quantity valency ('the shortage that amounted to five groups of goods'). It goes without saying that resolution of such ambiguity may be a challenge.

4   Conclusion

We have tried to show that Russian abounds in situations where the negation affects the valency structure of predicate words in a non-trivial way. Naturally, in the short paper we could not raise all of the relevant issues. In particular, we have not discussed the contribution to the semantic interaction of the negation and predicate valencies offered by the communicative structure of the sentence, the particular elements of the lexicographic definition (like assertion or presupposition), the scope of negation and the scope of the valencies themselves. Neither have we been able to (re-)consider the semantic interaction of the negation with non-actant entities, including the ambivalent durative constructions like *Ustanovka ne rabotala dva goda* lit. 'the equipment didn't work two years' ('worked less than two years' vs. 'was inactive for two years') much discussed by Jury Apresjan, Igor Boguslavsky, Elena Paducheva and others, or to account for the particularities of aspectual and temporal verbal meanings in this interaction. These are the issues of our future work.

Acknowledgements

Bibliography

Apresjan 2005 – Ю. Д. Апресян. О семантических правилах // *Болгарская русистика* 2005, № 3-4, с. 3-11.

Apresjan 2006 – Ю. Д. Апресян. Типы соответствия семантических и синтаксических актантов // *Проблемы типологии и общей лингвистики*. – СПб., 2006. – С. 15-27.

Boguslavsky 2009. – Igor Boguslavsky. Enlarging the Diversity of Valency Instantiation Patterns and Its Implications. Logic, Language, and Computation. *Lecture Notes in Computer Science Volume 5422,* 2009, pp 206-220.

Boguslavsky, Iomdin 2011. – Богуславский И.М.*,* Иомдин Л.Л. Актантные свойства автодериватов // *Грамматика и лексика у словенским језицима. Зборник радова с међународног симпозијума*. Нови Сад-Београд, 2011. С. 55-68.

Iomdin, Iomdin 2011: Boris Iomdin, Leonid Iomdin. Valency Ambiguity Interpretation: What Can and What Cannot be Done. // *Proceedings of the 5th International Conference on Meaning-Text Theory*. Barcelona, September 8 – 9, 2011. Igor Boguslavsky and Leo Wanner (eds.) ISBN 978-84-615-1716-9 P. 108-119. / http://meaningtext.net/mtt2011

Rakhilina 2010. *Construction Linguistics*. Москва, Азбуковник, 2010.

Syntactic Phraseme of the Type *X-t' ne pereX-t'*

Elizaveta Kuzmenko, Elmira Mustakimova

National Research University Higher School of Economics
Faculty of Philology
Moscow, Russia
eakuzmenko_2@edu.hse.ru
egmustakimova_2@edu.hse.ru

Abstract

This paper focuses on one of the syntactic phrasemes with partial verb gemination – *X-t' ne pereX-t'*. A description of its semantic and syntactic characteristics based on the materials from the Russian National Corpus and difficulties arising in the automatic processing of this structure are outlined.

Keywords

Semantics, microsyntax, syntactic phraseme, construction grammar

1   Introduction

There is an aspect of linguistic research which was not taken into account in the works of structuralists because it stays above the traditional segment approach to the language. This aspect is called construction grammar in terms of (Fillmore, 1988; Fillmore et al., 1988), microsyntax in (Iomdin, 2003) and phrasemes in (Mel'čuk & Iordanskaja, 2007). Phrasemes are characterized by interpenetration (which is different from the ordinary interaction) of its components (Rakhilina, 2010). Thus, they are situated between the grammatical and lexical aspects of the language, and this leads to some problems with the automatic analysis of the text. Various problems concerning phrasemes and text processing were considered in (Iomdin, 2003; Apresjan et al., 2010).

This work focuses on one of the syntactic phrasemes with partial verb gemination – *X-t' ne pereX-t'*. This phraseme has not been sufficiently investigated in phraseology studies. We are going to give a description of its semantic and syntactic characteristics based on the materials from the Russian National Corpus and outline the difficulties of processing this structure.

2   Meaning of the Phraseme

The Russian National Corpus gives 52 examples of this syntactic phraseme, among them 9 examples are ambiguous, the rest of the examples made the base for our study.

The number of clear examples, which is rather insignificant, shows that this phraseme belongs to language periphery, which makes it even more interesting to investigate.

The meaning of the phraseme

del-a-t'	*ne*	*pere-del-a-t'*
do-V-INF	NEG	COMPL-do-V-INF

and others of this type is defined in the following way: "There is a lot of Y that can undergo the action X, but some part of Y will not be affected by the action X, whatever efforts are put into it, because the quantity of Y is too big":

(1) *Vižu – mnogo raboty, <u>rabotat' ne pererabotat'</u>, i sily pribyvaet…* (G. Uspenskij. Vlast' zemli)
'I see – there is a lot of work, a person couldn't do it all, and the strength comes…'

The meaning of the phraseme partially corresponds to the meaning of the expression *mnogo Y* 'a lot of Y'. However, they differ in shades of meaning. The meaning of the phraseme is far more complex than the meaning of *mnogo Y. X-t' ne pereX-t'* means not only 'a lot of objects for an action', but also implies either that a lot of effort is needed for performing an action, or that the number of objects is itself the focus of discourse.

This phraseme has one actant Y, which is the object of an action. However, there is an insignificant number of examples that reveal deviations from the general definition and semantic governing. As this number is too insignificant to set off separate definitions, we find it appropriate to add two modifications:

1. If an agent Z is introduced, the phraseme takes the meaning "There is a lot of Y, so that if Z does X with Y, (s)he will not manage to do X with all Y":

(2) *Esli vy popali na staju okunej, vam ih taskat' ― ne peretaskat'.* (E.Vladimirova. Trud-7)
'If you come across the pack of perches, you won't be able to drag them all out.'

2. If the verb X is intransitive, an agent Z is introduced while Y disappears.

This case belongs to the periphery of the phraseme's combinatory power – there were only 3 examples. In such cases the phraseme takes the meaning "X is such a long action that Z will hardly be able to finish it":

(3) *Ty na graždanku smotriš', a nam tut služit' ne pereslužit'.* (V. Kornilov. Demobilizacija)
'You are looking forward to demobilization, but we will be serving here endlessly.'

There is a number of partially geminated constructions which have similar frames:

1. *vidimo-nevidimo* 'a lot of'

2. *tušit, tušit, ne potušit* 'trying hard to do X without any results achieved'.

The first structure is a full phraseme, the second is a syntactic phraseme, as their meaning cannot be composed of the meanings of their constituents and they have lexical and morphological restrictions for the constituents (**slyšimo-neslyšimo*). To prove that *X-t' ne pereX-t'* has the same kind of idiomaticity as these structures do, let us observe whether or not *X-t' ne pereX-t'* demonstrates any features of non-compositionality.

3 Non-Compositionality of the Construction

To prove the status of *X-t' ne pereX-t'* as an idiom, let us compare fragments (4) and (5):

(4) *Ptic perestreljat'.*
    'To shoot all the birds.'

(5) *Ptic streljat' – ne perestreljat'.*
    'There are so many birds that one cannot shoot all of them.'

The fragments (4) and (5) differ in two elements of the signifier: *streljat'* 'shoot' and *ne* 'not' (bold in the left part of the scheme below). However, the signified of the phrase (5) contains extra elements: the element 'so many' and the meaning of inability, which cannot be regularly derived from the meanings of *streljat'* and *ne*. More formally, this can be represented in the next scheme:

'shoot' + **'not'** + 'shoot all' → 'so many'$^{\text{dominant}}$ + 'not' + 'possible' + 'shoot all'.

The non-compositional units of the meaning are underlined in the right part of the scheme.

Another evidence of the idiomatic status of the phraseme is the combinatory power of the construction. For example, the verbs cannot take adverbs: **Knigi vnimatel'no čitat' – ne perečitat'* 'To read the books attentively and not be able to read them all'

A classification of phrasemes is given in (Mel'čuk & Iordanskaja, 2007). According to this classification and our previous proof, *X-t' ne pereX-t'* falls into the category of quasiphrasemes because it meets the following conditions:

1. The meaning of the construction contains the meanings of all its constituents and besides some unpredictable extra element.

2. The meanings of all the constituents are not semantically dominant.

The scheme presented above shows that the concerned phraseme conforms with the definition given here.

4 Restrictions on the Verbs

Not every verb of the Russian language can be used in this phraseme. Here are some restrictions:

1. If the verb with the prefix *pere-* obtains a meaning that is different from the intentional meaning, it cannot be used in the phraseme. For instance, *perevarit'* (from *varit'* 'to boil') is not 'to boil everything', but 'to digest' in its first meaning (which makes it impossible to use), 'to boil for too long' in the second (which is also not a completive

meaning, needed for the phraseme): **Ovosči varit' ne perevarit'* 'to boil the vegetables and not be able to boil them all'.

2. The meaning implies that a great number of objects Y already exists. Thus, the meaning of the construction is not compatible with the verbs of creating. If the object of the verb is a Patient (exists and undergoes changes), this verb can be used in the construction. If the object is the result of an action (does not exist and is being created), this verb cannot be used in the construction:

^{OK}*Knig/knigi čitat' ne perečitat'*.
'a person cannot read all the books'
**Pisem/pis'ma pisat' ne perepisat'*.
'a person cannot write all the letters'

3. Some verbs with prefixes are less likely to be used in the phraseme (here we consider the verbs in imperfective form like *vozdelyvat'* 'to cultivate' and not *sdelat'* 'to do'). For example, the phraseme allows the verb *sobirat'* 'to gather', but rejects the verb *vybirat'* 'to select':

^{OK}*Sobirat' ne peresobirat',*
**Vybirat' ne perevybirat'*

The possible reasons for these limitations are:

1. The prefix has the meaning that contradicts the meaning of the phraseme **Vybirat' ne perevybirat'* (here the prefix *vy-* has the meaning of isolating some part of the objects from the others, whereas the meaning of the phraseme implies that the objects represent a whole unity);

2. The prefix is two syllables long or the verb already contains two prefixes, so after adding the prefix *pere-* the word becomes "overwhelmed" with prefixes:

**Vozobnovlyat' ne perevozobnovit'* 'to resume';

3. The prefix forms a verb of a formal register, which contradicts the register of the whole phraseme: **Vbirat' ne perevobrat'* 'to absorb'.

Thus the phraseme tends to be selective in respect of prefixes of the verb.

5 Lexical and Morphological Identification of the Constituents

The phraseme always contains three elements: the first one is a verb in the active voice, usually transitive and in the imperfective aspect, the second one is a negative particle *ne* and the last one is a transitive verb with the prefix *pere-*, in perfective aspect and active voice.

The first and the third elements, which form the variable part of the phraseme, are cognate verbs, usually infinitives. However, there are some examples where the first element is in the imperative mood 2SG and the third element is in the indicative mood, non-past tense 2SG:

(6) *...rybi, diči stol'ko, čto beri — ne perebereš.* (V. Astaf'ev. Tsar'-Ryba)
 '…there is a lot of fish and fowl – you couldn't take it all.'

We suppose that the situation where both verbs of the phraseme are in 2SG, non-past tense is also possible:

(7) *Stol'ko marok: kleiš' ne perekleiš'.*
    'So many stamps – you couldn't paste them all.'

Other forms in the verbal paradigm seem unsuitable for this phraseme. Perhaps, such restrictions come from the ability of the imperative 2SG and the infinitive forms to imply a generic subject, whereas other tenses and mood forms imply a specific subject or situation.

6   Pragmatics

The phraseme can be used in informal communication to emphasize the amount of some objects or time needed for some action. Besides that, the phraseme usually conveys the speaker's attitude to the described situation, either very negative (8), or highly positive (9).

(8) *Zaxoronenij segodnya malo – vsego pyat', zato musora posle prazdnika opyat' na jego učastke vozit' ne perevozit'.* (S. Kaledin)
    'Today there have not been many burials – just five, but after the holiday there is again so much garbage at his plot – you couldn't transport it all'.

(9) *Vitalij dobyču vzjal: taskat' ne peretaskat' ― bylo čem gordit'sja!* (P. Aleshkovskij)
    'Vitalij took the prey; there was so much of it, you couldn't drag it all – there was enough to be proud of'.

The phraseme has two variants of a Patient depending on the case of the noun: accusative or genitive. The use of case is connected with the intention (the focus of interest) of the speaker. Let us observe the difference in the meaning of the same verb *taskat'* 'carry' when it is used in the phraseme with a Patient in genitive and in accusative:

1. **Gen.Pl + phraseme** is used to highlight the amount of objects, to which the action is applied:

(10) *Krome togo, <u>podobnyh scenariev</u> (GEN) «taskat' ne peretaskat'», i vo mnogih pri želanii možno obnaružit' «ruku Berezovskogo».*
     'Apart from that, you could pull similar scenarios endlessly, and in many of them you could find the trace of Berezovsky.'

In the example (10) the main idea is that the number of such scenarios is overwhelming, while the action itself (*taskat'*) is not so important. Note that the object is only possible in the plural form. In such cases the phraseme as a whole acts as an adverbial clause. In this variant of the phraseme the word order is fixed: **Čitat' knig – ne perečitat'*. However, the elements of the phraseme can sometimes alternate with pronouns:

(11) *Knig stol'ko – čitat' ih – ne perečitat'!* (authors).
     'So many books – you cannot read them all'.

2. **Acc + phraseme** is used to highlight the duration of the action being performed and the amount of effort which is put into the action:

(12) *Konečno, špaly byli ešhe ne vse položeny,* <u>*ix (ACC)*</u> *predstojalo taskat' ― ne peretaskat'*

‘Of course, the cross ties were not all laid yet, we were in for dragging them endlessly.’

In the example (12) the complexity of the action and its duration in the future are emphasized, while the amount of objects somehow stays in the background. In this variant of the phraseme the word order is not fixed, thus it is possible to say ᴼᴷ*Čitat' knigi – ne perečitat'*. In contrast to the genitive case, the object in accusative can be used both in singular and plural forms.

The phraseme is commonly used to describe a situation that is of high importance to the speaker or just a situation that worries the speaker though (s)he does not take part in it.

Corpus examples give the ground for an observation about register and style in which this phraseme is used. 82% of the examples are from press and publicistic prose, and that underlines the informality of this syntactic structure, so that the usage of the phraseme belongs to the area of persuasiveness and expressiveness. The rest of the examples come from conversations and dialect speech.

7   Translation

One of the major problems arising from the use of this phraseme is the problem of translation. As the reader may have noticed, throughout our paper we tried hard to come up with a decent translation for the phraseme while translating Russian examples into English, though the result was not very good from all points of view. The desired translation has to meet the following requirements:

- fully reflect the meaning of the phraseme;

- be comparable in length with the Russian equivalent;

- be idiomatic in order to convey stylistic and discursive features of the text.

The best option would be to find a similar phraseme in the language to which the presented phraseme is translated, however, this is almost impossible.

Existing translations of the phraseme are based on the meaning. This structure is translated literally, we can observe direct mapping of the components of meaning into other language's units:

1. Pavel na tretij den' povez kartošku. Pjatnadcat' meškov nagrebli, vo vsju imejuščujusja taru, a navalennaja v ogorode kuča edva podžalas' liš s odnogo kraja. Da ešče skol'ko kopat'! Eto značit, <u>vozit' ne perevozit</u>'. (V. Rasputin. Proščanije s Matjoroj)
‘Pavel took the potatoes away on the third day. They filled up fifteen sacks, all the bags they had, but the pile in the garden had barely got any smaller. There were so many more! That meant <u>they would not be able to take it all.</u>’ (Translated by Antonina Bouis)

The impossibility to do X with all Y is implied in the translation.

2. Vitalij upustil dobyču, a on – vzyal, i kakuju: <u>taskat' ne peretaskat',</u> bylo čem gordit'sya! Xorek dostal nož, no lajka ne unimalas'... (P. Aleškovskij. Žizneopisanije Xor'ka)

'Vitaly had missed his big chance and Skunk had taken it. And what a trophy this was. <u>You could never have hauled something this size any distance.</u> Skunk took out his knife, but the dog would give him no peace.' (Translated by Arch Tait)

The amount of Y is implied in the translation.

It seems that such a meaning-componential translation is not very suitable as it causes some loss of language brevity and expressiveness.

In the process of writing this paper we had to invent an intuitive translation for the phraseme and came up with the expression "to do X with Y endlessly". This option seems to sound non-professional, and it has a number of disadvantages. However, this one word *endlessly* reflects many components of the phraseme's meaning very briefly:

- the amount of Y;

- the amount of time needed to process all the Y – endlessness, in particular;

- the sequence of the previous statement – the impossibility to do X with all Y.

Moreover, in examples (3), (10), (12) the translation which takes into account the idea of endlessness seems to be even more appropriate than any other suggestions. This is probably caused by the fact that the main focus of discourse in these examples is not the objects, but the action.

Therefore, it can be noticed that this intuitive option appeals at least to two out of three traits of good translation mentioned in the beginning of this chapter.

8   Other Idiosyncrasies of the Phraseme

In sentences, the phraseme usually serves as a predicate of the impersonal clause, therefore, it forms the head of the clause. This syntactical structure recruits an actant Y from the definition of meaning, which serves as a Patient, in genitive or accusative case. As for the Agent Z from the modifications of the meaning, it is in dative case (dative subject syntactic relationship).

The punctuational design of the phraseme can be of different kinds:

1. The phraseme can be separated from the rest of the sentence by a dash.

(13) Tri centnera ryby ne podnimut ― ee lovit' ne perelovit'! (V. Astaf'ev. Tsar'-ryba)

2. The phraseme can contain a dash inside – example (2).

3. The phraseme is not separated from the rest of the sentence – example (8).

4. The phraseme can contain a comma inside.

(14) Nazyvajutsja konfety kratko i dohodčivo ―GNP: Gryzi, Ne Peregryzeš'!

5. The phraseme can be separated from the rest of the sentence by a colon – example (9).

The differences in punctuational design of the phraseme in the sentences outline the syntactical unity of the phraseme, though the bonds inside it are not so homogeneous. The phraseme serves

as an integral structure in the sentence, but inside itself it demonstrates that the bonds between the first constituent and the rest of the structure are weaker than those between the second and the third constituents.

The phraseme has a number of features in its intonational design. Before the phraseme, an intonational pause is usually required – it is used in order to emphasize the amount of objects or the efforts needed for an action, depending on the type of the phraseme. The phraseme as a whole is pronounced slightly faster than the rest of the utterance. The first constituent of the phraseme is pronounced with rising intonation, the second and half of the third constituents are pronounced at the same tone level, and at the end the intonation falls. Such changes of intonation are typical of the intonational frame No. 5 in the classification described in (Bryzgunova, 1980). This type of intonational frames is used to express and enhance the speaker's opinion on the subject.

9   Conclusion

Having analyzed the syntactic phraseme *X-t' ne pereX-t'* we can assume that, indeed, this phraseme constitutes a major problem for the automatic analysis because of its non-compositionality, difficulty in defining interior syntactic relations and idiosyncrasies in the exterior syntactic relations.

The problems, however, are not limited to the machine analysis of this phraseme. This structure also raises a set of translation difficulties. The laconic form of the phraseme contains the notions of conativity, plurality and action. Such complexity in its meaning has no analogy among signifiers in other languages, so it should be translated descriptively. The future research may examine the ways of overcoming these difficulties in more detail.

Acknowledgements

Bibliography

Apresjan, Yu., I. Boguslavskij, L. Iomdin, V. Sannikov. 2010. *Teoreticheskie problemy russkogo sintaksisa: Vzaimodejstvie grammatiki i slovarja*, 141—156. Jazyki slavjanskih kul'tur. Moscow.

Bryzgunova, E. 1980. Intonatsia. In *Russkaya grammatika*. Shvedova, N. (EIC), Moscow: Nauka.

Fillmore, Ch. J. 1988. The mechanisms of 'Construction Grammar'. In Proceedings of the Annual Meeting of the Berkeley Linguistics Society 14, 35—55.

Fillmore, Ch. J., P. Kay, M. C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: the case of let alone. In Language, Vol.64, No.3, 501—538.

Iomdin, L. 2003. Bol'šie problemy malogo sintaksisa. In *Trudy meždunarodnoj konferencii po komp'juternoj lingvistike i intellektual'nym tehnologijam Dialog'2003*, 216—222.

Mel'čuk I., L. Iordanskaja. 2007. *Smysl i sočetaemost' v slovare*, 215—245. Jazyki slavjanskih kul'tur. Moscow.

Rakhilina, E. 2010. *Lingvistika konstrukcij*. Moscow: Azbukovnik.

The Unit Graphs Framework:
A Graph-Based Knowledge Representation Formalism Designed for the Meaning-Text Theory

Maxime Lefrançois, Fabien Gandon

Wimmics, Inria, I3S, CNRS, UNSA
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis, France
Maxime.Lefrancois@inria.fr, Fabien.Gandon@inria.fr

Abstract

In this paper we are interested in the choice of a graph-based knowledge representation formalism that would allow for the representation, manipulation, query, and reasoning over linguistic knowledge of the Explanatory Combinatorial Dictionary of the Meaning-Text Theory (MTT). We show that neither the semantic web formalisms nor the Conceptual Graphs Formalism are suitable for this task, and we justify the introduction of the new Unit Graphs framework. We then detail the core of this formalism which is a hierarchy of unit types driven by their actantial structure. Finally, we define the new deep semantic representation level for the MTT, where the specialization of actantial structures of deep semantic unit types may correspond to a specialization of conveyed meanings.

Keywords

Linguistic Knowledge Representation, Meaning-Text Theory, Unit Graphs, Explanatory and Combinatorial Dictionary.

1   Introduction

In this paper we are interested in the choice of a graph-based Knowledge Representation (KR) formalism that would allow for the representation, manipulation, query, and reasoning over linguistic knowledge of the Explanatory Combinatorial Dictionary (ECD), which is the lexicon at the core of the Meaning-Text Theory (MTT) (c.f. for instance Mel'čuk, 2006). We envision two application scenarios of such a formalization:

- In a ECD lexicographic edition oriented project, we could enable the semi-automation of some of the lexicographers tasks. For instance, we could check that a set of constraints is satisfied, or we could suggest preliminary drafts of articles (e.g., lexical function key-value pairs, lexicographic definition sketches, government pattern).

- We could propose a syntax, which is a formal language based on knowledge engineering standards. Like WordNet today, the linguistic knowledge written with that syntax could be published to the web of linked data[1]. This would support their use as a highly structured lexical resource by consumers of the linked data cloud.

Most past or current projects that aimed at implementing the ECD did so in a lexicographic perspective. One important example is the RELIEF project (Lux-Pogodalla and Polguère, 2011), which aims at representing a lexical system graph named RLF (Polguère, 2009) where lexical units are interlinked by paradigmatic and syntagmatic links of lexical functions (Mel'čuk, 1996). In the RELIEF project, the description of Lexical Functions is based on a formalization proposed by Kahane and Polguère (2001). Moreover, lexicographic definitions start to be partially formalized using the markup type that has been developed in the Definiens project (Barque and Polguère, 2008; Barque et al., 2010), which aims at formalizing lexicographic definitions with genus and specific differences for the TLFi[2].

One exception is the proprietary linguistic processor ETAP-3 that implements a variety of ECD for Natural Language Processing (Apresian et al., 2003; Boguslavsky et al., 2004). Linguistic knowledge is asserted, and linguistic and grammatical rules are directly formalized in first order logic.

Adding to these formalization works, our goal is to propose a formalization from a knowledge engineering perspective, compatible with standard KR formalisms. The term formalization here means not only make non-ambiguous, but also make operational, i.e., such that it supports logical operations (e.g., knowledge manipulation, query, reasoning). We thus adopt a knowledge engineering approach applied to the domain of the MTT.

In this paper we first justify the introduction of the new Unit Graphs formalism (§2), we then detail the conjunctive unit types hierarchy (§3) at the core of this framework, and we finally draw some important implications for the MTT (§4).

2 Choice of a Knowledge Representation Formalism

At first sight, two existing KR formalisms seem interesting for the MTT. Semantic web formalisms (RDF/S, OWL, SPARQL), because the linked data is built on them, and Conceptual Graphs (CGs) formalism (Sowa, 1984; Chein and Mugnier, 2008), as we are to lead logic reasoning on graphs. Both formalisms are based on directed labelled graph structures, and some research has been done towards using them to represent dependency structures and knowledge of the lexicon (OWL in (Lefrançois and Gandon, 2011a; Boguslavsky, 2011), CGs at the conceptual level in (Bohnet and Wanner, 2010)).

Let us first recall that for a specific Lexical Unit L, Mel'čuk, (2004, p.5) distinguishes considering L in language (i.e., in the lexicon), or in speech (i.e., in an utterance). KR formalisms also do this distinction using types. Objects of the represented domain are named instances (or objects, or individuals), and are typed (or classified).

---

[1] The web of data is a W3C initiative, highly active today, http://linkeddata.org

[2] Trésor de la Langue Française informatisé, http://atilf.atilf.fr

2.1   Semantic Web Formalisms

There is a world wide deployment of the semantic web formalisms, and the RDF[3] syntax is the standard for structured data exchange over the web of linked data. The expressivity of RDF would be sufficient to represent the knowledge of the ECD. Yet, the semantics of RDF, in the logical sense, is limited to that of oriented labelled multi-graphs, and we wish also to enable the manipulation and reasoning over linguistic knowledge of the ECD. We thus need to introduce more semantics with RDFS[4] or OWL[5], while keeping the expressivity as low as possible to keep good computational properties. Yet RDFS and OWL only support binary relations, which is not the case of most valency-based predicates. One would need to use reification of n-ary relations[6], but then no semantics is attributed to such relations.

The ULiS project (Lefrançois and Gandon, 2011a, b) nevertheless proposed an architecture to enable such semantics: each lexical unit supports the projection of its lexicographic definition over itself. Lefrançois (2013) proved that this solution leads to an overwhelming computational complexity, i.e., the undecidable first order logic.

One alternative to represent lexicographic definitions of lexical units would be to use two reciprocal CONSTRUCT SPARQL[7] rules. But we then face the problem of rule languages and their compatibility with OWL (c.f., Krisnadhi et al., 2011), that led to no consensus nor standard today. These different problems led us to consider another formalism to represent knowledge of the ECD. We nevertheless want to be able to export these knowledge in RDF to exchange them over the web of linked data.

2.2   The Conceptual Graphs Formalism

The Conceptual Graphs (CGs) formalism (Sowa, 1984; Chein and Mugnier, 2008) has many similarities with the MTT. In their basic version, CGs represent typed instances interconnected by typed n-ary relations. Actually, the main goal of Sowa was natural language processing, and he originally inspired from the same works than MTT founders: Tesnière (1959). Sowa (1989) early suggested to introduce type definition of concepts and relations that do look similar to lexical units definitions in the ECD, and later on, Leclère (1998) also worked on the possibility to reason with type and concept definitions. One more asset of CGs is the fact that there are transformations between CGs and RDF/S (c.f., Corby et al., 2000; Baget et al., 2010). One could use these transformations to rewrite CGs to RDF for publication over the web of linked data. Moreover, one could adapt the architecture described in the ULiS (Lefrançois and Gandon, 2011a,b) project to CGs. Yet it is also not natural to represent the knowledge of the ECD using the CGs. Here are two reasons for that:

---

[3]   RDF - Resource Description Framework, http://w3.org/RDF/

[4]   RDFS - RDF Schema, http://www.w3.org/TR/rdf-schema/

[5]   OWL - Web Ontology Language, http://www.w3.org/TR/owl2-overview/

[6]   N-ary relations on the Semantic Web, http://www.w3.org/TR/swbp-n-aryRelations

[7]   SPARQL, http://www.w3.org/TR/sparql11-overview/

- A semantic unit may be represented as a concept type as it is instantiated in utterance semantic representations. On the other hand, if the associated lexical unit is predicative and has Semantic Actant Slots (SemASlots), then the semantic unit may dually be represented as a n-ary relation, so that its instances link other semantic units. The CGs do not offer a natural representation of this duality. In fact, in CGs, one must alternate concepts and relations, and a semantic representation of an utterance such as the one in figure 1 can't be directly represented by a CG.

- SemASlots of a lexical unit may differ from those of the lexical unit from which its sense derives[8]. Yet in the CGs, the inheritance mechanism of relation types that models the fact that a relation type is more specific than another, is constrained so that two relations with different arities must be incomparable. One thus cannot use this natural inheritance mechanism to model the specialization of senses.
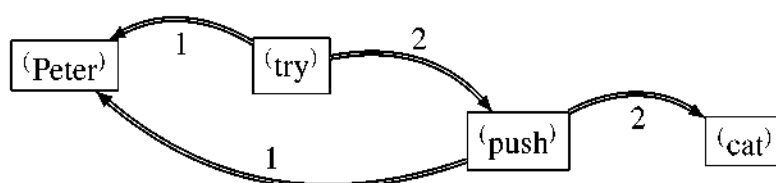

Figure 1: Illustration of the duality concept / relation of semantic units in the MTT, semantic representation of utterance "Peter tries to push the cat".

2.3 The New Unit Graphs Formalism

To sum up, neither the semantic web formalisms nor the CGs formalism allow for a natural representation of a hierarchy of unit types that may have actant slots, which is the basic knowledge of the ECD. As the CGs formalism is the closest from the MTT, we decide to use it as a starting point for designing a new graph-based formalism adapted to the representation of the knowledge of the ECD. We will also define transformations to the RDF syntax for sharing knowledge and publishing over the web of data. As we are to represent linguistic units of different nature (e.g., semantic units, lexical units, grammatical units, words), we choose to use the term unit in a generic manner and name the result of this adaptation Unit Graphs (UGs) framework.

3   The Unit Types Hierarchy

In this section we study how we shall revisit the CGs formalism so as to make it adapted to represent a hierarchy of unit types that may have actant slots. First of all, in the Unit Graphs (UGs) mathematical framework, the objects of the represented domain are named units, and are typed. Parallel with existing KR formalisms and Mel'čuk (2004, p.5), we thus establish a distinction between:

---

[8]   For instance, semantic unit ⟨rain⟩ is more specific than ⟨fall⟩ but the meaning of what falls and where it falls from is fixed to ⟨water drops⟩ and ⟨sky/cloud⟩ (Mel'čuk, 2004, p.14).

- Unit types (e.g., semantic unit type, lexical unit type), described in the ECD;

- Units (e.g., semantic unit, lexical unit), represented in the Unit Graphs (UGs).

Unit types will specify through actant slots how their instances (i.e., units) shall be linked to other units in a UG. Unit types and their actantial structure are described in a structure called hierarchy and denoted *T*.

Definition 1. A hierarchy of unit types is a tuple

$$\mathcal{T} \overset{\text{def}}{=} (T_D, \boldsymbol{S}_\mathcal{T}, \boldsymbol{\gamma}, \boldsymbol{\gamma_1}, \boldsymbol{\gamma_0}, C_A, \{\boldsymbol{\varsigma}_t\}_{t \in \boldsymbol{T}}, \perp_A^{\sqcap})$$

that enables to construct a coherent pre-ordered set of unit types with an actantial structure, i.e., actant slots that may be obligatory, optional or prohibited. Actant slots are signed, their signatures characterise the type of units that fill these slots.

3.1 Primitive Unit Types and Actant Slots

First, *T* contains a set of declared *Primitive Unit Types (PUTs)* denoted T_D. This set contains linguistic PUTs of different nature (e.g., semantic, lexical, grammatical). So that Actant Slots (ASlots) are named, *T* contains a set of binary relation symbols called Actant Symbols (ASymbols)*,* and denoted S_T. S_T contains numbers for the semantic unit types, and other classical symbols for the other levels under consideration (e.g, roman numerals I to VI for the MTT's Deep Syntactic level).

Then, no matter whether it is semantic, lexical or grammatical, a PUT $t \in \boldsymbol{T}$ has a set (that may be empty) of *Actant Slots (ASlots)* whose symbols are chosen in the set of ASymbols. Some ASlots may be obligatory, other optional (Mel'čuk, 2004, p.24), and we postulate that some may be prohibited too. For instance the Lexical Unit Type (LexUT) TO EAT has at least one obligatory semantic ASlot which is for the animal that eats, and an optional semantic ASlot which is for the container the animal eats in. If one specializes the meaning of TO EAT to define a new LexUT, we identify three basic cases that may happen:

- An optional ASlot may become obligatory;

- An optional ASlot may become prohibited, e.g., the container for TO GRAZE;

- A new ASlot may be introduced.

In order to represent these different types of ASlots and so that their presence in the hierarchy of Unit Types is coherent, we introduce three bijective mappings over the set of ASymbols:

- γ assigns to every $s \in S_T$ its radix unit type *γ(s)* that introduces an Actant Slot (ASlot) of symbol *s.* We denote $\boldsymbol{\Gamma}$ the range of γ, i.e., the set of *radices.*

- γ_1 assigns to every $s \in S_T$ its obligat unit type $\gamma_1(s)$ that makes the ASlot of symbol *s* obligatory. We denote $\boldsymbol{\Gamma}_1$ the range of γ_1, i.e., the set of *obligant.*

- γ_0 assigns to every $s \in S_T$ its prohibet unit type $\gamma_0(s)$ that makes the ASlot of symbol s prohibited. We denote $\boldsymbol{\Gamma_0}$ the range of γ_0, i.e., the set of *prohibent*.

The set *of Primitive Unit Types (PUTs)* is denoted T and defined as the disjoint union of the set of declared PUTs T_D, radices $\boldsymbol{\Gamma}$, obligant $\boldsymbol{\Gamma_1}$, prohibent $\boldsymbol{\Gamma_0}$, plus the *prime universal PUT* $\top$

$$\boldsymbol{T} \stackrel{\text{def}}{=} T_D \uplus \boldsymbol{\Gamma} \uplus \boldsymbol{\Gamma_1} \uplus \boldsymbol{\Gamma_0} \uplus \{\bot\} \uplus \{\top\} \tag{1}$$

and the *prime absurd PUT* $\bot$ (eq. 1).

We then introduce an inheritance mechanism for the PUTs, in the form of a specialization pre-order[9] $\lesssim$ over the set T. $t_1 \lesssim t_2$ models the fact that the PUT t_1 is more specific than the PUT t_2. $\lesssim$ is defined as the smallest pre-order such that: i) it includes the set $C_A \subseteq T^2$ of asserted PUTs comparisons (eq. 2), ii) $\top$ (resp. $\bot$) is maximal (resp. minimal) (eq. 3), and iii) for all ASymbol the radix is greater than the obligat and the prohibet (eq. 4).

$$(t_2, t_1) \in C_A \Rightarrow t_1 \lesssim t_2 \tag{2}$$

$$\forall t \in \boldsymbol{T}, \bot \lesssim t \lesssim \top \tag{3}$$

$$\forall s \in \boldsymbol{S_T}, \boldsymbol{\gamma_1}(s) \lesssim \boldsymbol{\gamma}(s) \text{ and } \boldsymbol{\gamma_0}(s) \lesssim \boldsymbol{\gamma}(s) \tag{4}$$

As every ASlot has a symbol, the set of ASlots of a PUT $t \in T$ is defined as the set of their symbols $\boldsymbol{\alpha}(t) \subseteq S_T$. Formally, the set $\boldsymbol{\alpha}(t)$ is the set of ASymbols whose radix is more general or equivalent to t (eq. 5), and thus every PUT that is more specific than the radix of an ASymbol $s \in S_T$ inherits an ASlot with symbol s. Similarly, the set of obligatory (resp. prohibited) ASlots of a PUT t is denoted $\boldsymbol{\alpha_1}(t)$ (resp. $\boldsymbol{\alpha_0}(t)$) and is defined as the set of ASymbols whose obligat (resp. prohibet) is more general or equivalent to t (eq. 6-7). The set of optional ASlots of a PUT t is denoted $\boldsymbol{\alpha_?}(t)$ and is the set of ASlots that are neither obligatory nor prohibited (eq. 8). The number of ASlots of a Primitive Unit Type (PUT) is denoted its *valency*.

$$\boldsymbol{\alpha}(t) \stackrel{\text{def}}{=} \{s \in \boldsymbol{S_T} \mid t \lesssim \boldsymbol{\gamma}(s)\} \tag{5}$$

$$\boldsymbol{\alpha_1}(t) \stackrel{\text{def}}{=} \{s \in \boldsymbol{S_T} \mid t \lesssim \boldsymbol{\gamma_1}(s)\} \tag{6}$$

$$\boldsymbol{\alpha_0}(t) \stackrel{\text{def}}{=} \{s \in \boldsymbol{S_T} \mid t \lesssim \boldsymbol{\gamma_0}(s)\} \tag{7}$$

$$\boldsymbol{\alpha_?}(t) \stackrel{\text{def}}{=} \boldsymbol{\alpha}(t) - \boldsymbol{\alpha_1}(t) - \boldsymbol{\alpha_0}(t) \tag{8}$$

In the lexicographic definitions, the type of the potential fillers of a SemASlot is sometimes written before the name of the variable. In the unit types hierarchy, *signatures* of PUTs give means to represent this information explicitly. More generally, not any unit may fill a specific ASlot of a PUT. For instance, only semantic units may fill ASlots of a semantic unit, and only units of type ⟨animal⟩ may fill ASlot 1 of Semantic Unit Type (SemUT) ⟨to eat⟩.

Formally, the set of signatures of PUTs is denoted $\{\varsigma t\}_{t \in T}$ and is a set of functions from $\boldsymbol{\alpha}$ to $T^{\cap}$. For all PUT t, ς_t is a function that associates to every ASlot s of t a set of PUTs $\varsigma t(s)$ that

---

[9]   A pre-order is a reflexive and transitive binary relation

characterise the type of the units that may fill this slot. For instance, the signature of ⟨to eat⟩ for its ASlot 1 is noted $\varsigma_{⟨to\ eat⟩}(1) = \{⟨animal⟩\}$. Signatures of a PUT t_1 may only be more specific than those of each of its ancestors t_2: if $t_1 \overset{\sim}{\lesssim} t_2$ and s is a common ASlot of t_1 and t_2, the signature of t_1 for s must be more specific or equivalent than that of t_2. For instance, the signature of ⟨savour⟩ for 1, i.e., $\{⟨person⟩\}$, is more specific than that of ⟨to eat⟩ which is $\{⟨animal⟩\}$.

The actantial structure of a PUT t is thus defined as the sets of its obligatory, prohibited and optional ASlots, and their signatures. It is inherited and possibly specialized by every descendent of t.

3.2 Hierarchy of Unit Types

A unit type may consist of several conjoint PUTs. In particular, it may be a lexical PUT and multiple grammatical PUTs, like {*def, plur,* ANIMAL} (⟨the animals⟩). To represent this, we introduce the set $T^\cap$ of possible Conjunctive Unit Types (CUTs) over T as the powerset[10] of T, i.e., $T^\cap = 2^T$. The definition of the actancial structure of PUTs is naturally extended to CUTs as follows:

$$\boldsymbol{\alpha}^\cap(t^\cap) \overset{\text{def}}{=} \bigcup_{t \in t^\cap} \boldsymbol{\alpha}(t) \tag{9}$$

$$\boldsymbol{\alpha}_1^\cap(t^\cap) \overset{\text{def}}{=} \bigcup_{t \in t^\cap} \boldsymbol{\alpha}_1(t) \tag{10}$$

$$\boldsymbol{\alpha}_0^\cap(t^\cap) \overset{\text{def}}{=} \bigcup_{t \in t^\cap} \boldsymbol{\alpha}_0(t) \tag{11}$$

$$\boldsymbol{\alpha}_?^\cap(t^\cap) \overset{\text{def}}{=} \boldsymbol{\alpha}^\cap(t^\cap) - \boldsymbol{\alpha}_1^\cap(t^\cap) - \boldsymbol{\alpha}_0^\cap(t^\cap) \tag{12}$$

$$\varsigma_{t^\cap}^\cap(s) \overset{\text{def}}{=} \bigcup_{t \in t^\cap | s \in \boldsymbol{\alpha}(t)} \varsigma_t(s) \tag{13}$$

Some PUTs are incompatible. For instance, no unit may be of both grammatical unit types *def* and *indef.* To represent this, T contains a set of declared absurd CUTs, denoted $\perp_A^\cap$, with $\perp_A^\cap \subseteq T^\cap$.

Finally, the pre-order $\overset{\sim}{\lesssim}$ over T is extended to a pre-order $\overset{\cap}{\lesssim}$ over $T^\cap$ (c.f., Lefrançois and Gandon, 2013). $\overset{\cap}{\lesssim}$ is computed as the smallest pre-order such that: i) it contains the natural extension of a pre-order over a set to a pre-order over its powerset (eq. 14), ii) $\top$ and $\emptyset$ are both maximal elements (eq. 15), iii) every CUT declared absurd is minimal (eq. 16), iv) the conjunction of $\boldsymbol{\gamma}_1(s)$ and $\boldsymbol{\gamma}_0(s)$ is minimal for all $s \in S_T$ (eq. 17), and v) if a CUT has a signature that is minimal, then it is minimal (eq. 18). The bottom of the pre-ordered set $T^\cap$ is the set of absurd CUTs, i.e., the unit types that cannot be instantiated.

---

[10] The powerset of X is the set of all subsets of X: 2^X

$$\forall t_2 \in t_2^\cap, \exists t_1 \in t_1^\cap, t_1 \lesssim t_2 \Rightarrow t_1^\cap \overset{\cap}{\lesssim} t_2^\cap \tag{14}$$

$$\varnothing \overset{\cap}{\lesssim} \{\top\} \tag{15}$$

$$\forall t^\cap \in \perp_A^\sqcap, t^\cap \overset{\cap}{\lesssim} \{\perp\} \tag{16}$$

$$\forall s \in \boldsymbol{S}_\mathcal{T}, \{\boldsymbol{\gamma_1}(s), \boldsymbol{\gamma_0}(s)\} \overset{\cap}{\lesssim} \{\perp\} \tag{17}$$

$$\exists s \in \boldsymbol{\alpha}^\cap(t^\cap), \boldsymbol{\varsigma}_{t^\cap}^\cap(s) \overset{\cap}{\lesssim} \{\perp\} \Rightarrow t^\cap \overset{\cap}{\lesssim} \{\perp\} \tag{18}$$

Lefrançois and Gandon (2013) proved that in the hierarchy of unit types, if $t_1^\cap \overset{\cap}{\lesssim} t_2^\cap$ then the actantial structure of $t^\cap{}_1$ is more specific than that of $t^\cap{}_2$, except for some degenerated cases (i.e., the void unit type Ø, and the absurd unit types). Thus as one goes down the hierarchy of unit types, an ASlot with symbol *s* is introduced by the radix *{γ(s)}* and first defines an optional ASlot for any unit type $t^\cap$ more specific than *{γ(s)}*, as long as $t^\cap$ is not more specific than the obligat *{γ₁(s)}* (resp. the prohibet *{γ₀(s)}*) of *s*. If that happens, the ASlot becomes obligatory (resp. prohibited). Moreover, the signature of an ASlot may only be more specific than that of its parents.

Any unit type that possesses ASlots thus represents a linguistic predicate as defined by Mel'čuk (2004, p.8), and unit nodes having that type must (resp1. may, resp2 must not) be linked to its obligatory (resp1. optional, resp2. prohibited) actants in a UG.

4 Implications for the Different Levels of Representation

As semantic ASymbols are numbers, the pre-order over semantic unit types cannot represent a specialization of meanings. Let us give an example to justify this.

The French semantic unit type ⟨outil⟩ (⟨tool⟩) has an ASlot 1 that corresponds to the person X that uses the tool, and a split ASlot 2 that corresponds either to the activity Y_1 or to the profession Y_2 for which the tool is designed[11]. Now ⟨ciseaux⟩ (⟨scissors⟩) has a stricter meaning than ⟨outil⟩, and also a ASlot 2 that now corresponds to the object Y that it is intended to cut. Thus ⟨ciseaux⟩ cannot be lower than ⟨outil⟩ in the hierarchy of semantic unit types because this would imply that an object is some kind of activity or profession.

We hence propose to introduce a deeper level of representation where one may describe meanings: the deep semantic level. We thus establish a distinction between deep and surface semantic unit types. Let us precise their definition and their actantial structure.

Definition 2 (Surface Semantic Unit Types and their ASlots). To every meaningful Lexical Unit Type (LexUT) L is associated a Surface Semantic Unit Type that is denoted ⟨L⟩. The ASlots of ⟨L⟩ correspond to the SemASlots of L as defined in (Mel'čuk, 2004, p.39) and are numbered.

Definition 3 (Deep Semantic Unit Types and their ASlots). To every meaningful LexUT L is associated a Deep Semantic Unit Type (DSemUT) that is denoted ⟨L⟩. The set of deep semantic ASymbols are semantic roles (e.g., agent, experiencer, object). The set of ASlots of

---

[11] See (Mel'čuk, 2004, p.43) for more details on split ASlots

a DSemUT corresponds to obligatory or optional participants of the linguistic situation denoted by L that are: a) SemASlots of L, or b) SemASlots of a LexUT whose meaning is more generic than that of L.

For instance figure 2 illustrates the actantial structure of ⌐outil¬ (⌐An artefact designed for a person X to use it for an activity Y_1 (or for a profession Y_2)¬) and ⌐ciseaux¬, which derives from ⌐tool¬. ⌐tool¬ has two obligatory actant slots *possessor* and *activity*, and an optional ASlot *profession*. ⌐ciseaux¬ inherits the ASlots of ⌐tool¬, and restricts the signature of *activity* to be ⌐cut¬. As CISEAUX also introduces a SemASlot which is the object to be cut, ⌐ciseaux¬ also introduces a new ASlot *objectToBeCut*.

⌐outil¬
⇒ **possessor :** ⌐person¬
⇒ **activity :** ⌐activity¬
(⇒) **profession :** ⌐profession¬

⌐ciseaux¬
⇒ **possessor :** ⌐person¬
⇒ **activity :** ⌐cut¬
(⇒) **profession :** ⌐profession¬
⇒ **objectToBeCut :** ⌐object¬

Figure 2: Actantial structures of ⌐outil¬ and ⌐ciseaux¬.

One may need to introduce a new ASymbol every time a SemASlot that conveys a new meaning is introduced. The set of semantic roles thus cannot be bound to a small set of universal semantic roles.

5 Conclusion

In this paper we showed that both semantic web and conceptual graphs formalisms are not adapted to represent knowledge of the ECD while ensuring good computational properties. We hence justified the introduction of the new Unit Graphs (UGs) graph-based knowledge representation formalism.

The Unit Types hierarchy is the core structure of the UGs. It consists in a minimal set of mathematical objects that allows to construct a pre-ordered set of unit types with actantial structures. The actantial structure of a unit type is composed of actant slots that may be optional, obligatory, or signed, and that are signed. Moreover, a unit type inherits and possibly specialize the actantial structure of its parents.

The so-defined Unit Types hierarchy has strong implications for the MTT. In fact, the preorder over unit types cannot correspond to a meaning-specialization relation for semantic unit types as defined in the MTT. We therefore introduced a deep-semantic representation level, and defined the deep and surface semantic unit types and their actantial structure.

Current directions of work include the definition of UGs and their semantics, in the logical sense; the definition of lexicographic definitions that must be at the deep semantic level; and the representation of the semantic derivation part of lexical functions, which constrains the actantial structure and the definitions of deep semantic units.

Bibliography

Apresian, J., Boguslavsky, I., Iomdin, L., Lazursky, A., Sannikov, V., Sizov, V., and Tsinman, L. (2003). ETAP-3 linguistic processor: A full-fledged NLP implementation of the MTT. In *First International Conference on Meaning–Text Theory (MTT'2003),* pages 279–288.

Baget, J.-F., Croitoru, M., Leclère, M., Mugnier, M.-L., and Gutierrez, A. (2010). Translations between RDF(S) and conceptual graphs. *ICCS'10: 18th International Conference on Conceptual Structures – From Information to Intelligence*, pages 28–41.

Barque, L., Nasr, A., and Polguère, A. (2010). From the Definitions of the 'Trésor de la Langue Française' To a Semantic Database of the French Language. In Fryske Akademy, editor, *Proceedings of the XIV Euralex International Congress, Fryske Akademy*, pages 245–252, Leeuwarden, Pays-Bas. Anne Dykstra et Tanneke Schoonheim, dir.

Barque, L. and Polguère, A. (2008). Enrichissement formel des définitions du *Trésor de la Langue Française informatisé (TLFi) dans une perspective lexicographique. Lexique*, 22.

Boguslavsky, I. (2011). Semantic Analysis Based on Linguistic and Ontological Resources. In Boguslavsky, I. and Wanner, L., editors, *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011)*, pages 25–36, Barcelona, Spain. INALCO.

Boguslavsky, I., Iomdin, L., and Sizov, V. (2004). Multilinguality in ETAP-3: reuse of lexical resources. In Sérasset, G., editor, *Proc. COLING 2004 Multilingual Linguistic Ressources,* pages 1–8, Geneva, Switzerland. COLING.

Bohnet, B. and Wanner, L. (2010). Open source graph transducer interpreter and grammar development environment. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 19–21, Valletta, Malta. European Language Resources Association (ELRA).

Chein, M. and Mugnier, M.-L. (2008). *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*. Springer-Verlag New York Incorporated.

Corby, O., Dieng, R., and Hébert, C. (2000). A Conceptual Graph Model for {W3C} Resource Description Framework. In Ganter, B. and Mineau, G. W., editors, *Conceptual Structures: Logical, Linguistic, and Computational Issues*, number 1867 in Lecture Notes in Computer Science, pages 468–482. Springer Berlin Heidelberg.

Kahane, S. and Polguère, A. (2001). Formal foundation of lexical functions. In *Proceedings of ACL/EACL 2001 Workshop on Collocation*, pages 8–15.

Krisnadhi, A., Maier, F., and Hitzler, P. (2011). OWL *and Rules. Reasoning Web. Semantic Technologies for the Web of Data*, pages 382–415.

Leclère, M. (1998). Raisonner avec des définitions de types dans le modèle des graphes conceptuels. Revue d'intelligence artificielle, 12(2):243–278.

Lefrançois, M. (2013). Représentation des connaissances du DEC: Concepts fondamentaux du formalisme des Graphes d'Unités. In *Actes de la 15ème Rencontre des Étudiants*

Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL'2013), pages 164–177, Les Sables d'Olonne, France.

Lefrançois, M. and Gandon, F. (2011a). ILexicOn: Toward an ECD-Compliant Interlingual Lexical Ontology Described with Semantic Web Formalisms. In Boguslavsky, I. and Wanner, L., editors, *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011)*, pages 155–164, Barcelona, Spain. INALCO.

Lefrançois, M. and Gandon, F. (2011b). ULiS: An Expert System on Linguistics to Support Multilingual Management of Interlingual Semantic Web Knowledge bases. In Elena Montiel-Ponsoda John McCrae, P. B. and Cimiano, P., editors, *Proc. of the 9th International Conference on Terminology and Artificial Intelligence (TIA 2011)*, volume 775, pages 108–114, Paris, France. CEUR-WS.

Lefrançois, M. and Gandon, F. (2013). *The Unit Graphs Mathematical Framework*. Research Report RR-8212, Inria.

Lux-Pogodalla, V. and Polguère, A. (2011). Construction of a French Lexical Network: Methodological Issues. In *Proceedings of the International Workshop on Lexical Resources*, Ljubljana.

Mel'čuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In Wanner, L., editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–102. Benjamins Academic Publishers, Amsterdam/Philadelphia.

Mel'čuk, I. (2004). *Actants in Semantics and Syntax I: Actants in Semantics*. Linguistics, 42(1):247–291.

Mel'čuk, I. (2006). *Explanatory combinatorial dictionary. Open Problems in Linguistics and Lexicography*, pages 225–355.

Polguère, A. (2009). *Lexical systems: graph models of natural language lexicons. Language resources and evaluation*, 43(1):41–55.

Sowa, J. (1989). Using a lexicon of canonical graphs in a semantic interpreter. In *Relational models of the lexicon*, pages 113–137. Cambridge University Press New York, NY, USA.

Sowa, J. F. (1984). *Conceptual structures: information processing in mind and machine. System programming series.* Addison-Wesley Pub., Reading, MA.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. C. Klincksieck (Colombes, Impr. ITE).

Application of the Unit Graphs Framework to Lexicographic Definitions in the RELIEF Project

Maxime Lefrançois, Romain Gugert, Fabien Gandon, Alain Giboin

WIMMICS, Inria, I3S, CNRS, UNSA
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis, France
{Maxime.Lefrancois, Romain.Gugert, Fabien.Gandon, Alain.Giboin}@inria.fr

Abstract

The Unit Graphs (UGs) framework is a graph-based knowledge representation (KR) formalism that is designed to allow for the representation, manipulation, query, and reasoning over linguistic knowledge of the Explanatory Combinatorial Dictionary of the Meaning-Text Theory. One crucial advantage of this new formalism over other KR formalisms is that it is designed to represent valency-based predicates, and lexicographic definitions of lexical units in the form of semantic graphs. The goal of this paper is twofold. It both introduces the core of the UGs framework and illustrates how it may be used to represent lexicographic definitions in the RELIEF lexicographic edition project.

Keywords

Linguistic Knowledge Representation, Meaning-Text Theory, Unit Graphs, Explanatory Combinatorial Dictionary, Lexicographic Definitions, Deep Semantic Representation level

1 Introduction

We are interested in the benefits of using a graph-based Knowledge Representation (KR) formalism to enable the formalization (from a knowledge engineering perspective), of linguistic knowledge of the Explanatory Combinatorial Dictionary (ECD) (Mel'čuk, 2006).

In this paper we focus on lexicographic definitions in the RELIEF lexicographic edition project (Lux-Pogodalla and Polguère, 2011), which aims at representing a lexical system graph named RLF (Polguère, 2009) where lexical units are interlinked by paradigmatic and syntagmatic links of lexical functions (Mel'čuk, 1996). The RELIEF is already based on different formalization works to represent lexicographic definitions, namely: a hierarchy of semantic labels (Polguère, 2011), the markup type that has been developed in the Definiens project (Barque and Polguère, 2008; Barque et al., 2010) to specify genus and specific differences, and the disambiguation of meaningful words in the definition.

Maxime Lefrançois, Romain Gugert, Fabien Gandon, Alain Giboin

Adding to these formalization works, our goal is to propose a formalization from a knowledge engineering perspective, compatible with standard KR formalisms. The term formalization here means not only make non-ambiguous, but also make operational, i.e., such that it is adapted to logical operations (e.g., knowledge manipulation, query, reasoning). We thus adopt a knowledge engineering approach applied to the domain of the Meaning-Text Theory (MTT).

At first sight, two existing KR formalisms seem interesting for this job: semantic web formalisms (e.g., RDF[1], RDFS[2], OWL[3] SPARQL[4]), and Conceptual Graphs (CGs) (Sowa, 1984; Chein and Mugnier, 2008). Both of them are based on directed labelled graph structures, and some research has been done towards using them to represent dependency structures and knowledge of the ECD (OWL in (Lefrançois and Gandon, 2011; Boguslavsky, 2011), CGs at the conceptual level in (Bohnet and Wanner, 2010)). Yet Lefrançois (2013) showed that neither of these KR formalisms can represent valency-based predicates, therefore lexicographic definitions.

These issues led to the introduction of the new Unit Graphs (UGs) KR formalism, which is a graph-based KR formalism originally designed to formalize hierarchies of unit types that have an actantial structure. Term unit is used in a generic manner and may refer to linguistic units of different nature (e.g., semantic units, lexical units, grammatical units, words).

Apart from introducing the UGs framework and implications for the MTT, this paper details an application scenario for the edition of lexicographic definitions in the RELIEF project. This paper first describes the current scenario of lexicographic definition edition in the RELIEF project (§2), then successively overviews three important aspects of the UGs formalism, and the added value for the RELIEF project:

- The core of the UGs framework which is the hierarchy of unit types. We will justify the introduction of a deep semantic representation level for the MTT. At this level one may refine the semantic labels hierarchy so that every semantic label (= deep semantic unit types in this paper) is assigned an actantial structure (§3).

- UGs and unit types definition, which enable the formal definition of lexical units in the form of an equivalence between two deep-semantic UGs (§4).

- Rules, which enable the specification of the correspondence between deep and surface semantic actant slots (§5).

---

[1] RDF - Resource Description Framework, c.f., http://w3.org/RDF/

[2] RDFS - RDF Schema, c.f., http://www.w3.org/TR/rdf-schema/

[3] OWL - Web Ontology Language, c.f., http://www.w3.org/TR/owl2-overview/

[4] SPARQL, c.f., http://www.w3.org/TR/sparql11-overview/

2   Current Scenario

The lexicographic edition software developed in the RELIEF project is named MVSDicet. Let us sketch a scenario where Alain, the leader of the project, assigns the task of defining French lexical unit PEIGNE$_{2A}$, which is defined in (Mel'čuk et al., 1999) by:

PEIGNE$_{2A}$: $^($comb$^)$≡$^($Weaving tool that a person X uses to untangle object Y$^)$.

(1)   Sophie first seeks for a semantic label in the hierarchy of semantic labels (Polguère, 2011). She chooses $^/$outil$^\backslash$ ($^($tool$^)$).

(2)   Sophie determines that PEIGNE$_{2A}$ has two obligatory semantic actants: a person X, and an object Y. She then seeks for a fitting propositional form in a hierarchy that only Alain develops. She may choose: de X ~ [pour Y] (~ of X [for Y]).

(3)   Sophie then writes the lexicographic definition markuped with genus and specific differences as in the Definiens project (Barque and Polguère, 2008; Barque et al., 2010). Finally, for each of the meaningful words of the lexicographic definition, Sophie specifies to what lexical unit in the RLF it refers to.
<CC label="outil">outil de tissage</CC>
<PC role="utilisation">que X utilise pour peigner#2 Y</PC>

3   Refinement of the Semantic Labels Hierarchy

First, for a specific Lexical Unit L, Mel'čuk (2004, p.5) distinguishes considering L in language (i.e., in the lexicon), or in speech (i.e., in an utterance). KR formalisms and the UGs formalism also do this distinction using types. In this paper and in the UGs formalism, there is thus a clear distinction between *units* (e.g., semantic unit, lexical unit), which will be represented in the UGs, and their *types* (e.g., semantic unit type, lexical unit type), which are described in the ECD.

The core of the UGs framework is a structure called *hierarchy of unit types* and noted *T,* where unit types and their actantial structure are described. This structure is thoroughly described in (Lefrançois and Gandon, 2013a) and studied in (Lefrançois and Gandon, 2013b).

Whether they are semantic, lexical or grammatical, unit types are assigned a set of *Actant Slots (ASlots),* and every ASlot has a so-called Actant Symbol (ASymbol) which is chosen in a set denoted S_T. S_T contains numbers for the semantic unit types, and other "classical" symbols for the other levels under consideration (e.g, Roman numerals I to VI for the Deep Syntactic actants). The set of ASlots of a unit type *t* is represented by the set $\alpha(t)$ of ASymbols these ASlots have. Moreover,

- some ASlots are obligatory, they form the set $\alpha_1(t)$ of Obligatory Actant Slots (OblASlots)*;*

- other are prohibited, they form the set $\alpha_0(t)$ *of* Prohibited Actant Slots (ProASlots)*;*

- the ASlots that are neither obligatory nor prohibited are said to be optional, they form the set $\alpha_?(t)$ of Optional Actant Slots (OptASlots)*.*

Finally, every unit type $t \in T$ has a signature function q_t that assigns to every ASlot of t a unit type, which characterizes units that fill such a slot.

The set of unit types is then pre-ordered[5] by a specialization relation $<$, and for mathematical reasons as one goes down the hierarchy of unit types the actantial structure may only become more and more specific: (i) some ASlot may appear, be optional a moment, and at some points become obligatory or prohibited; (ii) the signatures may only become more specific.

As semantic ASymbols are numbers, the pre-order over semantic unit types cannot represent a specialization of meanings (Lefrançois and Gandon, 2013a). Let us give an example to justify this.

The French semantic unit type ⟨outil⟩ (⟨tool⟩) has an ASlot 1 that corresponds to the person X that uses the tool, and a split ASlot 2 that corresponds either to the activity Y1 or to the profession Y_2 for which the tool is designed[6]. Now ⟨peigne2$_a$⟩ (⟨comb⟩) has a stricter meaning than ⟨outil⟩, and also an ASlot 2 that now corresponds to the object Y that it is intended to untangle. Thus ⟨peigne2$_a$⟩ cannot be lower than ⟨outil⟩ in the hierarchy of semantic unit types because this would imply that an object is some kind of activity or profession.

We hence propose to introduce a deeper level of representation where one may describe meanings: the *deep semantic level*. We thus establish a distinction between surface and semantic unit types. Let us precise their definition and their actantial structure.

Definition 1 (Surface Semantic Unit Types and their ASlots). To every meaningful Lexical Unit Type (LexUT) L is associated a *Surface Semantic Unit Type (SSemUT)* that is denoted ⟨L⟩. The ASlots of ⟨L⟩ correspond to the Semantic Actant Slots (SemASlots) of L as defined in (Melčuk, 2004, p.39), and are numbered.

Definition 2 (Deep Semantic Unit Types and their ASlots). To every meaningful LexUT L is associated a Deep Semantic Unit Type (DSemUT) that is denoted $^{/}$L$^{\backslash}$ The set of deep semantic ASymbols are semantic roles (e.g., *agent, experiencer, object)*. The set of ASlots of a DSemUT corresponds to obligatory or optional participants of the linguistic situation denoted by L that are: a) SemASlots of L, or b) SemASlots of a LexUT whose meaning is more generic than that of L.

Actually, one may need to introduce a new ASymbol every time a SemASlot that conveys a new meaning is introduced. The set of semantic roles thus cannot be bound to a small set of universal semantic roles.

In the RELIEF project, the set of semantic labels are pre-ordered with respect to the specialization of meanings, as is the hierarchy of DSemUT in the UGs framework. We thus propose to identify semantic labels and DSemUTs, and to augment them with actantial structures. One major implication is that one need one DSemUT per meaningful LexUT.

Let us sketch the extension of the scenario described in section 2. Sophie wants to define the French LexUT PEIGNE$_{2A}$. She thus needs to characterize its associated DSemUT $^{/}$peigne$_{2a}$$^{\backslash}$.

---

[5]   A pre-order is a reflexive and transitive binary relation.

[6]   See (Mel'čuk, 2004, p.43) for details on split ASlots.

She first opens a new tab in which ′peigne$_{2a}$\ appears in a void box as illustrated in figure 1a. Sophie needs to choose the nearest parent in the hierarchy of DSemUTs. She clicks on the question mark and the current hierarchy of DSemUTs appears like in figure 1b. She chooses ′tool\ The box that was void now contains the inherited actantial structure of ′tool\ as illustrated in figure 1c. ′tool\ has three arbitrarily symbolized ASlots:

- *possessor* for variable X is obligatory and has signature ′person\;

- *activity* for variable Y_1 is obligatory and has signature ′activity\;

- *profession* for variable Y_2 is optional and has signature ′profession\.

Now Sophie may restrict the actantial structure of ′peigne$_{2a}$\ .

1. ′peigne$_{2a}$\ is designed to untangle, so Sophie clicks on ′activity\ and chooses ′untangle\ in the hierarchy of DSemUTs.

2. ′peigne$_{2a}$\ is designed for the weaver profession, so Sophie clicks on ′profession\ and chooses ′weaver\ in the hierarchy of DSemUTs.

3. the ASlot *profession* is obligatory for ′peigne$_{2a}$\, so Sophie clicks on symbol (⇒)*,* which becomes ⇒ .

4. ′peigne$_{2a}$\ introduces a new obligatory ASlot *object* for variable Y with signature /object\. So Sophie clicks on New actant slot, and fills a form where she defines a new ASymbol *object,* specifies that this ASlot is obligatory, and specifies the signature: ′object\

Thus the description of the actantial structure of ′peigne$_{2a}$\ looks like in figure 1d.

Let us go back to how lexicographic definitions are currently defined in the RELIEF project. The DSemUT ′tool\ has no actantial structure for the moment. Yet,

- the central component of the definition (element CC) specifies the profession for which ′peigne$_{2a}$\ is designed: ′weaver\.

- in the peripherical component (element PC), a human reader immediately understands that the activity (PC role utilisation here) for which ′peigne$_{2a}$\ is designed is: ′peigner$_2$\ (′untangle\).

- ′peigner$_2$\ has a SemASlot Y that does not correspond to a participant of ′tool\, but that is related to a participant of ′untangle\.

Providing ′peigne$_{2a}$\ with an actantial structure that specializes that of ′tool\ enables to explicit some of this knowledge, and to give a partial but formal lexicographic definition to ′peigne$_{2a}$\. To complete the formalization of the lexicographic definition of ′peigne$_{2a}$\, one need for instance to represent the fact that the SemASlot X of ′peigne$_{2a}$\ corresponds to the SemASlot X of ′untangle\ for instance.

$$\triangledown\ /\text{entity}^\backslash\ [(\Rightarrow)1\ ]$$
$$\triangledown\ /\text{physical object}^\backslash\ [(\Rightarrow)1\ ]$$
$$\triangledown\ /\text{artefact}^\backslash\ [(\Rightarrow)1]$$
$$\triangledown\ /\text{instrument}^\backslash\ [\Rightarrow 2]$$
$$\triangleright\ /\text{tool}^\backslash\ [(\Rightarrow)1, \Rightarrow 2]$$

$/\text{peigne}_{2a}{}^\backslash \lesssim ?$
⊞ New actant slot

(a) Starting point to define PEIGNE$_{2A}$

(b) List view of the hierarchy of DSemUTs

$/\text{peigne}_{2a}{}^\backslash \lesssim /\text{tool}^\backslash$
$\Rightarrow$ **possessor :** $/\text{person}^\backslash$
$\Rightarrow$ **activity :** $/\text{activity}^\backslash$
$(\Rightarrow)$ **profession :** $/\text{profession}^\backslash$
⊞ New actant slot

(c) Inherited actantial structure of $/\text{tool}^\backslash$

$/\text{peigne}_{2a}{}^\backslash \lesssim /\text{tool}^\backslash$
$\Rightarrow$ **possessor :** $/\text{person}^\backslash$
$\Rightarrow$ **activity :** $/\text{untangle}^\backslash$
$\Rightarrow$ **profession :** $/\text{weaver}^\backslash$
$\Rightarrow$ **object :** $/\text{object}^\backslash$
⊞ New actant slot

(d) Actantial structure of $/\text{peigne}_{2a}{}^\backslash$

Figure 1: Definition of the actantial structure of $/\text{peigne}_{2a}{}^\backslash$.

4 Definition of Unit Types and Lexicographic Definitions

Now the actantial structure as defined in previous section is not sufficient to represent the lexicographic definition. For instance, $/\text{untangle}^\backslash$ has an agent ASlot, an this agent must correspond to the possessor of $/\text{peigne}_{2a}{}^\backslash$. One thus need UGs to fully represent the definition of $/\text{peigne}_{2a}{}^\backslash$. Let us first introduce the definition of UGs.

UGs include actantial relations, which are considered of type predicate-argument and are described in the hierarchy of unit types that we introduced in section 3. Now UGs also include circumstantial relations which are considered of type instance-instance. Example of such relations are the deep syntactic representation relations ATTR, COORD, APPEND of the MTT, but we may also use such relations to represent the link between a lexical unit and its associated surface semantic unit for instance. Circumstantial relations are labelled by symbols chosen in a set of so-called Circumstantial Symbols (CSymbols), denoted *Sc,* and their categories and usage are described in a hierarchy denoted *C*.

UGs are defined over a so-called support, $S =^{def} (T,C,M)$ where *T* is a hierarchy of unit types, *C* is a hierarchy of CSymbols, and *M* is a set of *unit identifiers*. To make a long story short, UGs have an underlying oriented labelled graph structure. Nodes are called *unit nodes* and are labelled by a unit type and one or more unit identifier. Every arc is labelled and represents an actantial (resp. circumstantial) relation if its symbol belongs to the set of ASymbols (resp. CSymbols). Finally some unit nodes may be asserted to be equivalent, i.e., to represent the same unit. Lexicographic definitions are to be represented at the deep semantic level, as an equivalence between two deep semantic UGs.

Definition 3 (Definition of a unit type, Lexicographic definition of a LexUT). Let $^/L^\backslash$ be the DSemUT associated with lexical unit L. The lexicographic definition of L corresponds to the definition of $^/L^\backslash$, which is a triple $D_{/L\backslash} =^{\text{def}} (D^-_{/L\backslash}, D^+_{/L\backslash}, \kappa)$, where:

1. $D^-_{/L\backslash}$ represents only a central unit node typed with $^/L^\backslash$ and some other unit nodes that fill some of the ASlots of $^/L^\backslash$;

2. $D^+_{/L\backslash}$ is a UG called the *expansion* of $^/L^\backslash$ with no circumstantial triple in these two λ-UG because circumstantials must not be part of the lexicographic definition of a LexUT.

3. κ is a mapping from the unit nodes of $D^-_{/L\backslash}$ to some unit nodes of $D^+_{/L\backslash}$.

Let us sketch how Sophie may define the lexicographic definition of PEIGNE$_{2A}$, i.e, the definition of $^/$peigne2$_a^\backslash$.

The starting point is the box that represents the actantial structure of /peigne2$_a$ as illustrated in figure 1d.

1. Sophie first drags and drops some ASlots outside the box. This enables to make explicit for instance that $^/$untangle$^\backslash$ has two obligatory ASlot. The result of this process is illustrated in figure 2a.

2. Sophie may then drag the ASlot *agent* of $^/$untangle$^\backslash$ and drop it over the box of $^/$person$^\backslash$. This merges participants as illustrated in figure 2b.

3. The *object* of $^/$peigne$_{2a}^\backslash$ and the *fibres* of $^/$untangle$^\backslash$ must be linked by a meronymy relation. For the sake of illustration, we assume there exists a DSemUT $^/$partOf$^\backslash$ that carries this meaning. Sophie clicks on a "add a unit node" button, and seeks for $^/$partOf$^\backslash$ in the hierarchy of DSemUTs. A unit node typed $^/$partOf$^\backslash$ is then added as in figure 2c.

4. Sophie drags the *whole* of $^/$partOf$^\backslash$ and drops it over the *object* of $^/$peigne$_{2a}^\backslash$; and drags the *part* of $^/$partOf$^\backslash$ and drops it over the *fibres* of $^/$untangle$^\backslash$ The result of this process is illustrated on figure 2d.
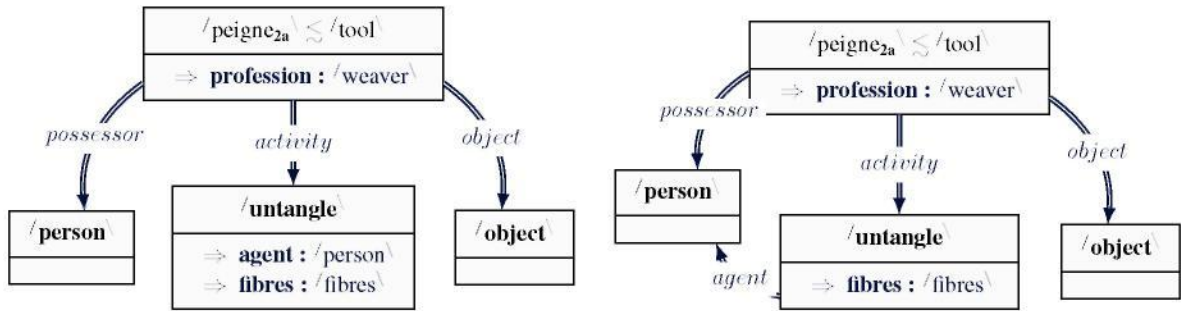
(a) Interesting participants of the definition of $^/$peigne$_{2a}^\backslash$ may be given a node by drag and drop.

(b) One may merge participants using drag and drop.
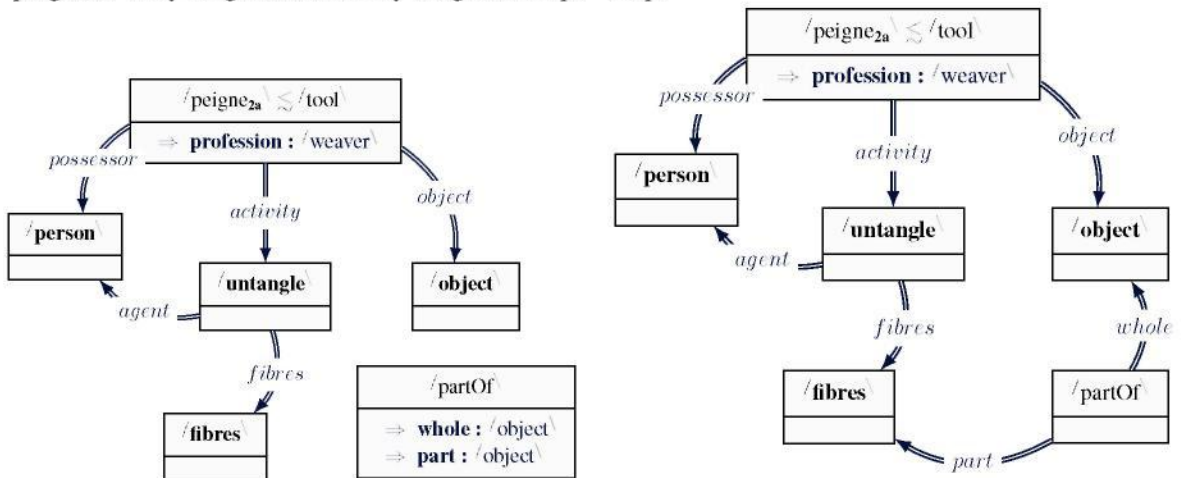
(c) One may add nodes in the definition.

(d) Complete definition of $^/$peigne$_{2a}^\backslash$.

Figure 2: Different steps in the definition of the Deep Semantic Unit Type $^/$peigne$_{2a}^\backslash$.

From this graph one may automatically build the definition $D_{/peigne2a\backslash} = (D^-_{/peigne2a\backslash}, D^+_{/peigne2a\backslash}, \kappa)$ of $^/$peigne$_{2a}^\backslash$ such as defined in definition 3. This definition is illustrated in figure 3
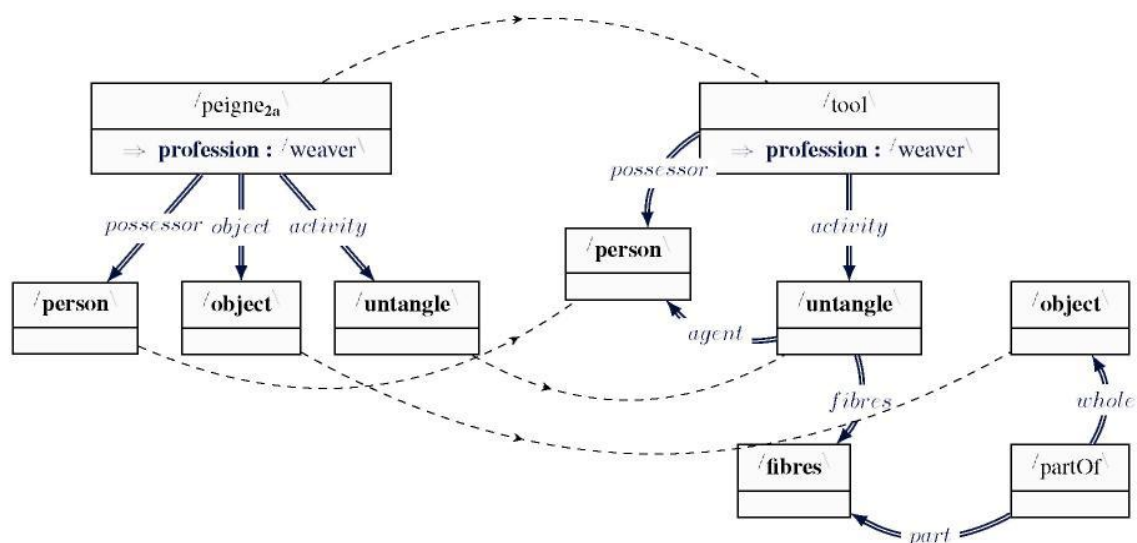
Figure 3: Illustration of the definition $D_{/\text{peigne}_{2a}\backslash} = (D^-_{/\text{peigne}_{2a}\backslash}, D^+_{/\text{peigne}_{2a}\backslash}, \kappa)$ of $/\text{peigne}_{2a}\backslash$. $D^-_{/\text{peigne}_{2a}\backslash}$ is on the left, the expansion $D^+_{/\text{peigne}_{2a}\backslash}$ on the right, and the dashed links represent the mapping κ.
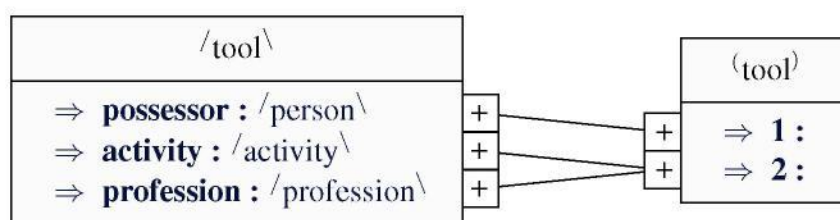
5   Rules and Deep-Surface Semantic ASlots Correspondence

In the UGs formalism, a rule is composed of two UGs: a hypothesis H and a conclusion C, and a partial function from the unit nodes of H to the unit nodes of G. If the hypothesis H projects on to a UG G (the rule is applicable), then one may add C to G accordingly (apply the rule).

To one definition may thus correspond two reciprocal rules: one that adds $D^+_{/L\backslash}$ to a graph where $D^-_{/L\backslash}$ projects, and one that adds $D^-_{/L\backslash}$ to a graph where $D^+_{/L\backslash}$ projects. If there is the defined PUT in a UG then one may infer its definition, and vice versa.

Rules also enable to represent correspondences between representations of two adjacent levels, and some shall be automatically generated from the government pattern. In this section we will define the correspondence between ASlots of a DSemUT and ASlots of a SSemUT.

Suppose Sophie now wants to represent the correspondence between the deep and surface semantic actant slots for TOOL and PEIGNE$_{2A}$. Sophie opens a new dedicated tab for each of these tasks. The content on the tab is: on the left a box for the DSemUT with its actantial structure, and on the right a box for the SSemUT with its actantial structure. A button is situated in front of each ASlot as illustrated in figures 4a and 4b, and Sophie just needs to drag and drop one of these buttons to the other, so as to link deep semantic ASlots with surface semantic ASlots. Every ASlot of a SSemUT must be linked to at least one ASlot of a DSemUT, several in case of split SemASlots.

(a) Illustration of the deep-surface semantic ASlots correspondence for TOOL.
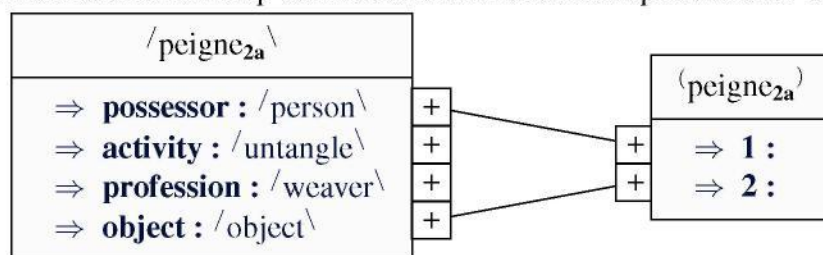
(b) Illustration of the deep-surface semantic ASlots correspondence for PEIGNE$_{2A}$.

Figure 4: Illustration of the correspondence between the actantial structure of a Surface Semantic Unit Type, and the actantial structure of its associated Deep Semantic Unit Type.

6 Conclusion

We thus illustrated how the UGs framework may be used to edit lexicographic definitions in the RELIEF project.

We overviewed the hierarchy of unit types that enables to describe unit types with their actantial structure: optional, obligatory and prohibited Actant Slots (ASlots) and their signature. The pre-order over unit types is such that the actantial structure may only become more and more specific as one goes down the hierarchy of unit types. We then justified the introduction of a new representation level for the MTT: the deep semantic representation level. The deep semantic unit type $^{/}\mathrm{L}^{\backslash}$ associated with a LexUT L has ASlots that are symbolized by semantic roles, and that correspond to participants of the linguistic situation denoted by L which are SemASlots of L or of LexUTs whose meaning is less specific than L. We detailed an application scenario in the context of the RELIEF project: the semantic labels are deep semantic unit types and one may specify their actantial structure.

A UGs is a set of unit nodes that are typed and interlinked through actantial and cirumstantial relations. We introduced the lexicographic definition of LexUTs as definitions of their associated DSemUT. We detailed an application scenario in the context of the RELIEF project: a lexicographer may manipulate nodes so as to little by little construct a deep semantic graph that represents the decomposition of the deep semantic unit type associated with the defined LexUT.

Finally rules enable to specify correspondences between ASlots of corresponding unit types at adjacent representation levels. We illustrated our approach with a scenario at the deep-surface semantic level interface, and showed how split ASlots shall be dealt with.

There are several research directions that we currently investigate:

1. Many rules may be needed to represent correspondences between the deep semantic and the surface semantic representation levels in case some SemASlot are optional or split. More research is needed to represent these cases and to generalize the definition of rules so as these cases may be factorized. Same goes for definitions of DSemUTs that have optional ASlots.

2. We developed a prototype web application and produced a demonstration available online: http://wimmics.inria.fr/doc/video/UnitGraphs/editor1.html. We currently lead an ergonomic study in partnership with actors of the RELIEF project in order to enhance the workflow of our prototype.

Bibliography

Barque, L., Nasr, A., and Polguère, A. (2010). From the Definitions of the 'Trésor de la Langue Française' To a Semantic Database of the French Language. In Fryske Akademy, editor, *Proceedings of the XIV Euralex International Congress,* Fryske Akademy, pages 245– – 252, Leeuwarden, Pays-Bas. Anne Dykstra et Tanneke Schoonheim, dir.

Barque, L. and Polguère, A. (2008). *Enrichissement formel des définitions du Trésor de la Langue Française informatisé (TLFi) dans une perspective lexicographique*. Lexique, 22.

Boguslavsky, I. (2011). Semantic Analysis Based on Linguistic and Ontological Resources. In Boguslavsky, I. and Wanner, L., editors, *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011)*, pages 25–36, Barcelona, Spain. INALCO.

Bohnet, B. and Wanner, L. (2010). Open source graph transducer interpreter and grammar development environment. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 19–21, Valletta, Malta. European Language Resources Association (ELRA).

Chein, M. and Mugnier, M.-L. (2008). *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs.* Springer-Verlag New York Incorporated.

Lefrançois, M. (2013). Représentation des connaissances du DEC: Concepts fondamentaux du formalisme des Graphes d'Unités. In *Actes de la 15ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL'2013)*, pages 164– – 177, Les Sables d'Olonne, France.

Lefrançois, M. and Gandon, F. (2011). ILexicOn: Toward an ECD-Compliant Interlingual Lexical Ontology Described with Semantic Web Formalisms. In Boguslavsky, I. and Wanner, L., editors, *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011),* pages 155–164, Barcelona, Spain. INALCO.

Lefrançois, M. and Gandon, F. (2013a). The Unit Graphs Framework: A graph-based Knowledge Representation Formalism designed for the Meaning -Text Theory. In *Proceedings of the 6th International Conference on Meaning-Text Theory (MTT'2013)*, Prague, Czech Republic.

Lefrançois, M. and Gandon, F. (2013b). *The Unit Graphs Mathematical Framework. Research Report* RR-8212, Inria.

Lux-Pogodalla, V. and Polguère, A. (2011). Construction of a French Lexical Network: Methodological Issues. In *Proceedings of the International Workshop on Lexical Resources*, Ljubljana.

Mel'čuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In Wanner, L., editor, Lexical Functions in Lexicography and Natural Language Processing, pages 37–102. Benjamins Academic Publishers, Amsterdam/Philadelphia.

Mel'čuk, I. (2004). Actants in Semantics and Syntax I: Actants in Semantics. Linguistics, 42(1):247–291.

Mel'čuk, I. (2006). Explanatory combinatorial dictionary. Open Problems in Linguistics and Lexicography, pages 225–355.

Mel'čuk, I., Arbatchewsky-Jumarie, N., Iordanskaja, L., Mantha, S., and Polguère, A. (1999). Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexicosémantiques IV. Les Presses de l'Université de Montréal, Montréal, Canada.

Polguère, A. (2009). Lexical systems: graph models of natural language lexicons. Language resources and evaluation, 43(1):41–55.

Polguère, A. (2011). Classification sémantique des lexies fondée sur le paraphrasage. Cahiers de lexicologie, 98:197–211.

Sowa, J. F. (1984). Conceptual structures: information processing in mind and machine. System programming series. Addison-Wesley Pub., Reading, MA.

Means of Argumentation and the Meaning of Conjunctions (Casual Meaning of Russian Conjunction *a to*)[1]

Irina Levontina

Russian Language Institute (Vinogradov Institute)
Russian Academy of Sciences
Moscow
irina.levontina@mail.ru

Keywords

Semantics, conjunctions, polysemy, causation, implicature, Russian

1   Introduction

The conjunction *a to*, just as *a ne to*, has many times been the subject of linguistic description [Санников 1989; Белошапкова 1970; Подлесская 2000; Собинникова 1969; Колосова 1980; Инькова-Манзотти 2000; Израэли 2000, Урысон 2008, 2010]. Researchers were mostly interested in semantic differences between these conjunctions and the conditions of their interchangeability. Attention was also given to the structure of these conjunctions' polysemy, especially in the case of *a to*. Of special importance are the works of E.V. Uryson, who made an attempt to elucidate the semantic contribution of the elements *a, to, ne* in the semantics of *a to* and *a ne to.*

1.1

The goal of this work is more modest. It seems that one of the more interesting lexemes of the conjunction *a to* has not yet been adequately described. It's the meaning is usually described as causal [Белошапкова 1970; Санников 1989; Урысон 2008]:

(1)   *Sxodi v buločnuju, a to xleba net*
      'Will you go to the bakery, (because) we've got no bread?'

(2)   *Pojdem domoj, a to zavtra rano vstavat'*
      'Let us go home, (because) we have to rise early tomorrow'

---

(3) *Net li u tebya soli, a to u menya končilas'?*
'Have you got any salt, (because) I'm out of mine'

It could be argued that the absence of bread is presented here as the reason for the need to go to the bakery, and tomorrow's early start is the reason for the decision to go home (the salt example, though, is more difficult to interpret that way). However, it is obvious that the mere idea of a cause is not enough. It is not possible to say:

(4) **Na ulice luži, a to šel dožd'*
'There are puddles in the street, because it rained'

(5) **U nego kašel', a to on promočil nogi*
'He is coughing, because he got his feet wet'

If the meaning of the conjunction were limited to causality, these phrases would have been possible.

According to A. Israeli, the meaning of *a to* here is as follows: 'p *a to* q': 'q explains why the speaker thinks p or states p; q is the reason for p'. Israeli states that this is an "exclusively illocutive *a to*, as opposed to other cases". Indeed, in most cases with the causal *a to* in the second part of the sentence, the speaker is justifying the speech act made in the first part, or the thought contained therein:

(6) *D'ad', – skazal Genka, – mne domoj nado. A to pozdo. Otec budet rugat'.* [Juriy Koval'. Groza nad kartofel'nym polem (1974)]
"Uncle", said Genka, "I've got to go home. Because it is late. Father will scold me'

The speaker explains why he needs to go home.

(7) *Ty na noč' okno ne zakryvaj, a to dušno.* [Ju. O. Dombrovskij. Ledi Makbet (1970)]
'Do not close the window for the night, because it is stuffy'

Here the speaker explains his request as well.

The fact that the conjunction *a to* is often used in contexts where it connects to the illocutive component of the first clause was already noted in [Падучева 1985: 46]; one of the examples was

(8) *Gde Ivan, a to im načal'stvo interesovalos'*
'Where is Ivan, (because) the bosses have been looking for him'

It was noted in the same paper that the first clause is therefore often a question or an imperative.

The above-mentioned example – *Net li u tebja soli, a to u menja konchilas'?* – is very typical. In this case it is quite obvious that *a to* introduces the commentary to the speaker's question. See the next example, where the speaker also explains why this specific question is being asked:

(9) *Kvartira svoja? Vanna jest? Gut. A to tut obščaja tol'ko. Budete vozit' ee k sebe myt'sja. Ona myt'sja lubit.* [Tatjana Tolstaya. Reka Okkervil' (1983)]

'You have your own apartment? A bathroom? Good. Because here it is shared. You will take her to your place to take baths. She loves taking baths'

1.2

There seem to be certain examples contradicting this interpretation:

(10) *Vozle ee krovati ležali tomiki Axmatovoj, Pasternaka, Baratynskogo... Kogda tetka umerla, biblioteku srazu že rasprodali. Predvaritel'no brat i ego žena vyrvali listy s avtografami. A to neudobno.* [Sergey Dovlatov. Naši (1983)]
'Volumes by Akhmatova, Pasternak, Baratynsky were lying by her bed... When the aunt died, her library was immediately sold. The brother and his wife had torn out signed title pages before selling the books, because it would have been awkward.'

(11) *Govorjat, čto xozyaeva "Pivovaren Ivana Taranova" iščut ploščadku pobliže k centru – a to daleko vozit' "PIT" v Moskvu iz Orenburga.* [Evgenij Tolstyx. Pivka dlya ryvka (2003) // "Soveršenno sekretno", 2003.09.01]
'They say that the owners of "Ivan Taranov's Brewery" are looking for a site closer to the center, because it is a long haul taking "PIT" from Orenburg to Moscow'

(12) *Odin iz dilerov BMW dal takoe objasnenie: "My prosto rešili davat' čestnije ceny, a to vse ravno prixodit čelovek v salon i vidit sovsem ne to, čto v reklame"* [Xasan Ganiev. Novosti (2002) // "Avtopilot", 2002.09.15]
'One of the BMW dealers explained it as follows: "We simply decided to offer fair prices, because the customer would come to the salon and see something completely different than the commercials".'

Even in these examples the non-trivial speaker does not disappear completely. One could argue that these are examples of free indirect speech. It is evident from the example with the signed books. If the phrase went like (13), the evaluation of the situation as 'awkward' would be construed as coming from the narrator:

(13) *Predvaritel'no brat i ego žena vyrvali listy s avtografami. Potomu čto <poskol'ku, tak kak> inače neudobno prodavat' knigi.*
'The brother and his wife had torn out the signed title pages before selling the books, for the reason that otherwise it would have been awkward to sell the books.'

Dovlatov, however, clearly means that the late woman's relatives, from the speaker's point of view, do not understand her at all; one example is that they destroy the most valuable part of the books, the authors' autographs. *Awkward* is the relatives' evaluation: for them Akhmatova's autograph is not a cultural treasure but merely a sign that the book used to belong to someone else. Let us stress again: the examples above show that it is often not the justification of the speech act, but rather the foundations for actions undertaken that are being announced by the speaker. An analogy could be in place here.

The word *pochemu* 'why' as a separate utterance is often used in Russian if the asker enquires about the reason of the situation discussed, but not the reason for the speech act itself. Thus, it would be normal to say (14 a), but (14 b) would sound strange:

(14) a. *Sol' otsyrela. – Počemu?*
'The salt got wet. – Why?'
b. *U tebya sol' jest'? – Počemu? [počemu ty sprašivaeš?]*
'Do you have any salt? – Why? [why are you asking?]'

Saying "Why are you asking" here is possible, but the isolated *Pochemu?* would have been incorrect. At the same time the English *Why?* or the German *Warum?* could be freely used in such contexts. In Russian, a different phrase should be used, namely "*A čto*?"

(14) c. *U tebya sol' est'? – A čto?*
'Do you have any salt? – Why?'

A čto, in its turn, is not used in questions pertaining to the objective cause. This shows that other Russian means of expressing causality could be sensitive to the difference between the cause for a situation and the justification of a speech act.

2.1

However, Alina Israeli's interpretation does not explain all the cases of limitations on the use of *a to* in question. Consider the following example:

(15) a *Xorošo, čto ty priexal, a to ja tak skučala* [Aleksej Varlamov. Kupavna // "Novyj Mir", 2000]
'It is good that you came, because I missed you so much'

This phrase is completely natural and quite common. The speaker makes an appraisive statement and goes on to justify the appraisal. Now let us try to modify the statement:

(15) b **Žal' čto on uexal, a to s nim bylo tak veselo*
'Pity he left, (because) it was so much fun with him'

It is obvious that this is not a good sentence, yet at first glance the semantic relations between the parts are exactly the same as in the above-listed quite natural examples: in the second part the speaker justifies the statement made in the first part. Sentences given below are also dubious or impossible:

(16) **Ne nado vyryvat' listy s avtografami. A to s nimi knigi ešče dorože.*
'Do not tear out the autographed pages. (Because) the books are even more expensive with them'

(17) **Iščut ploščadku pobliže k centru, a to eto gorazdo udobnee.*
'They are looking for a site closer to the center, (because) it is much more convenient'

(18) **Ne zakryvaj okno, a to vozdux takoj čudesnyj.*
'Do not close the window, because the air is so nice'

The original sentences given above were quite normal. It is obvious that the reason for that lies in the field of semantics, and we cannot consider our description of the meaning of *a to*

complete before our explanation gets to the root of why such examples are incorrect. Let us review a couple of sentences in more detail.

(19) a. *Xorošo, čto ty priexal, a to ja tak skučala* [Aleksej Varlamov. Kupavna // "Novyj Mir", 2000]
'It is good that you came, because I missed you so much'
 b **Žal' čto on uexal, a to s nim bylo tak veselo*
'Pity he left, (because) it was so much fun with him'

It could be noted that in the correct example the second part, roughly speaking, is about something negative, or at least about a situation that promises some unwanted consequences. The speaker explains the statement made in the first part with exactly the wish to avoid that unwanted result. The same is true of other correct examples:

a to opozdaem	'because we will be late'
a to dušno	'because it is stuffy'
a to zavtra rano vstavat'	'because we have to rise early tomorrow'

This could be called 'negative' motivation. If we see 'positive' motivation, *a to* becomes misplaced:

*a to tak udobnee	'because it is more convenient'
*a to tak vygodnee	'because it is more profitable'
*a to tak prošče	'because it is easier'
*a to tak veselee	'because it is more fun'

Let us review two similar pairs of examples:

(20) a. *Možete perezvonit'? A to mne očen' dorogo*. [Andrei Volos. Nedvižimost' (2000) // Novyj Mir, 2001]
'Could you call me back? Because it is very expensive for me'
b. **Davajte ja perezvonu. A to mne eto besplatno.*
'Let me call you back, because it is free for me'

(21) a. *Slušaj, ty ne možeš najti mne rabotu? Ubirat'sja v kvartire. U kakix-nibud' novyx russkix, tol'ko ne opasnyx. A to mne v teatrax ničego počti ne playat.* [Katya Metelitsa. Fruska (1997) // Stolitsa, 1997/06/17]
'Look, could you find me a job? Cleaning the apartments. At some New Russians, just not dangerous ones. Because I am hardly paid anything at the theaters.'
b. **Slušaj, ty ne možeš najti mne rabotu? Ubirat'sja v kvartire. U kakix-nibud' novyx russkix, tol'ko ne opasnyx. A to za uborku očen' xorošo platyat.*
'Look, could you find me a job? Cleaning the apartments. At some New Russians, just not dangerous ones. Because they pay very well for cleaning work.'

As we see here, in the correct sentences with *a to* the speaker refers to a negative argument, in the incorrect sentences — to a positive one.

2.2

It should be noted that the circumstance the speaker refers to can be neutral or positive by itself, but in the context of a specific situation it invokes certain complications. It is quite possible to say:

(22) a. *Daj mne putevoditel', a to mne na sledujuščej nedele v Veneciju exat'.*
'Lend me your guide book, because I am off to Venice next week'

In this example the 'motivational' parts could be construed as (22 b), but not (22 c):

(22) b. *a to moj očen' tyaželyj* <bestolkovyj>
'because mine is too heavy <quite useless>'
c. * *a to on u tebya očen' tolkovyj*
'because yours is very useful'

In the correct sentences it is implied that a trip without a guidebook or with a different guidebook would be worse. In the incorrect sentences, the motivation is often positive. Consider also the following example:

(23) *Kvitancija, tovarišč direktor, v mojem pasporte pod obložkoj, na remont velosipeda, už pyat' dnej propuščeno, a to zavtra op'at' vyxodnoj.* [Ju. O. Dombrovskiy. Fakul'tet nenužnyx vešej, čast' 1 (1978)]
'The receipt, comrade director, is under the cover of my passport, for bike repair, it has been due for five days, because it is a holiday again tomorrow'

Generally speaking, having a holiday tomorrow is good rather than bad. However, what it means here is that the repair shop will be closed on a holiday, and it would be impossible to take back the bike. This undesirable circumstance is the one that the speaker is referring to, motivating the need to go to the repair shop as soon as possible.

It should be noted that *a to* is prospective: this conjunction does not just presuppose an unpleasant situation, but rather the possibility of such a situation arising or continuing in the future. For that reason, it is correct to say (24 a), but it would be incorrect to say (24 b):

(24) a. *Dolgo nam ešče ždat'? A to ja uže načinaju somnevat'sja, čto xoču pokupat' eto plat'e.*
'Are we in for a long time? Because I'm starting to have doubts about this dress.'
b. * *Idem otsjuda. A to ja peredumala pokupat' eto plat'e.*
'Let's leave. (Because) I changed my mind about buying this dress.'

2.3

The idea of something unwanted somewhat aligns the 'causal' *a to* with one of the other lexemes of this conjunction, which has the meaning of threat/warning:

(25) *Ne kriči, a to ujdu.*
'Do not yell, or I will leave'

(26) *Zastegnis', a to prostudišsja.*
'Button up, or you will get cold'

It is important in this case that both the conjunction *a to* and the conjunction *a ne to* can be used. They would be synonymous in these cases with words like *inače* (otherwise), *ili* (or), *v protivnom slučae* (alternatively) — with subtle differences between all of them, of course. Let us introduce a number of examples, where the conjunction *a to* functions in the context of threat/warning:

(27) *Davaj čerteži, a to ja nikuda ne poedu! – k etomu svodilos' trebovanie Andreja.*
[Anatoly Azolsky, Lopushok // Noviy Mir, 1998]
'Give me the blueprints, or I won't go anywhere! – that was what Andrei's demand was down to'

(28) *Tol'ko ne davaj mamke. A to ona vse porvet.* [Boris Ekimov. Pinočet (1999)]
'Only don't give it to Mom. She'll tear it up'

(29) *"A možno ja čut'-čut' gostincev poprobuju?" – "Možno", usmexnulsya Ded Moroz. – "Tolko ne uvlekajsya, a to život zabolit".* [Jury Makarov. Pro Zajca // Murzilka, 2001]
'May I taste some of these gifts? – "Go ahead", Father Frost chuckled, "Just do not go overboard, because you will have stomach ache"

(30) *Poproščajtes' i idite k vyxodu. A ešče lučše – begite, a to, ne roven čas, zaboleete. Na vas smotret' xolodno.* [Vera Belousova. Vtoroy Vystrel (2000)]
'Say your goodbyes and go to the exit. Or better run, because, God forbid, you will get sick. It makes me cold just looking at you'.

2.4

In all these cases *a to* can be replaced with *a ne to*. Now let us consider the following example:

(31) a. *Idi v dom, a to xolodno, zamerzneš*
'Come inside, because it is cold, you will freeze'

This is a completely natural example, and the use of *a to* is quite appropriate. Now let us try to break down the phrase in two:

(31) b. *Idi v dom, a to <*a ne to> xolodno*
'Come inside, it is cold'
c. *Idi v dom, a to <a ne to> zamerzneš*
'Come inside, you will freeze'

It is now obvious that in the first case our "causal" meaning is represented, and in the second case the meaning of warning; just as we should expect, in the second case the replacement with *a ne to* is possible, but in the first case it is impossible. However, this juxtaposition makes it clear that *a to* expresses a similar idea in both cases. In the first part it is said that specific action should be taken, in the second part a certain negative outcome is outlined that would occur if the action is not taken. But in the case of *a to zamerznesh' / because you'll freeze* the negative consequence that should be avoided is directly named, while in the other case the thought about its possibility is a pragmatic implicature (it is announced that it's cold, and the listener, naturally, realizes that he could freeze).

Similarly:

(32) a. *Nado ostanovku delat', – skazal on. – Čaj nado pit', a to golova čego-to bolit...*
[Jury Koval'. Labaz (1972)]
"We should stop", he said. "We should drink tea, because I've got a headache..."

If the end of the sentence were

(32) b. *a to golova zabolit*
'because I'll have a headache'
c. *a to golova ne projdet*
'because the headache won't go'

it would have been the warning, and *a to* could have been replaced with *a ne to* or *inače / otherwise.* In this case, it is impossible. The idea about the unpleasant consequence in the form of continuing headache is not expressed directly but represented as an implicature.

(33) *Pročti kusoček, a? A to u menja ruki v kraske. – Kakoj kusoček? – sprosil Andrej. – Luboj. – Togda, – skazal Andrej, – ja s togo mesta načnu, gde sam čitaju.* [Viktor Pelevin. Želtaya strela (1993)]
'Would you read a passage? My hands are covered in paint. – "Which passage?" Andrej asked. – "Any." – "In that case," said Andrej, "I will start where I'm reading myself".'

This implicature is also quite obvious: the book can be ruined (because the speaker's hands are covered in paint, and he would spoil the book if he takes it). Of course, **A ne to u menja ruki v kraske* would be impossible.

2.5

Thus, logically, and probably historically as well, the 'causal' meaning of *a to* is the development of its 'threat/warning' meaning. In this paper we do not intend to trace the formation history of this meaning; we would only like to point out that it is not an innovation. Consider one example:

(34) *Bylo uže dvadcat' minut tret'ego, a učitelja istorii ne bylo ešče ni slyšno, ni vidno daže na ulice, <...> – Kažetsja, Lebedev nynče ne pridet, – skazal Volodya, otryvayas' na minutku ot knigi Smaragdova, po kotoroj on gotovil urok. – Daj bog, daj bog... a to ja rovno ničego ne znaу... odnako, kažetsya, vot on idet, – pribavil ja pečal'nym golosom.* [L. N. Tolstoy. Otročestvo (1854)]
'Lebedeff does not appear to be coming today, said Volodya, tearing himself for a moment from Smaragdoff's book, in which he was preparing his lesson. "God grant, God grant he may not! But I know nothing. But he seems to be coming yonder," I added in a sorrowful voice.' [Translation by Isabel F. Hapgood]

3.1

Now let us consider the place of the causal *a to* among other means of expressing causality. Reviewing the examples above, one could see that in most cases it is not very easy to replace the 'causal' *a to* with classical causal conjunctions *potomu čto, tak kak, poskol'ku, ibo*:

(35) *Sxodi v buločnuju, a to <poskol'ku, potomu čto> xleba net.*
    'Would you go to the bakery, because <inasmuch as, for> we've got no bread'

(36) *Pojdem domoj, a to <tak kak> zavtra rano vstavat'*
    'Let us go home, because <since> we have to rise early tomorrow'

(37) *Net li u tebya soli, a to <$^?$poskol'ku> u menja končilas'?*
    'Have you got any salt, because <$^?$as far as> I am out of mine?'

Even in cases where causal conjunctions are possible, the resulting phrases are not identical to the phrases with *a to.* Cause-and-effect relationship is expressed in them much more explicitly and forcefully. Of course, there is a cause-and-effect idea in *a to*, but its function is different than in purely causal conjunctions.

3.2

It would seem that to a certain degree *a to* might be compared to another interesting word, *ved'*. It is usually thought that *ved'* expresses a referral to the common knowledge shared between speaker and listener. This does not explain, though, many aspects of *ved'*, especially those that are called *ved' prozrenija* (revelational *ved'*) (*A ved' eto Petja! / Why, this is Petya!*) in the works of S. V. Kodzasov and K. Bonno [Bonno, Kodzasov 1998]. Let us also review one of the most frequent contexts of clarification. Somebody speaks about their meeting with a German and clarifies:

(38) *A ja ved' v škole nemeckij učil.*
    'Actually, I've studied German at school'

The addressee might have been unaware of the fact. This statement could have also been made in a conversation with a random interlocutor. The basis for *ved'* here is different: 'I think this should be taken into account for the correct understanding of the situation' [see Levontina 2005]. It seems that *a to* is similar to *ведь* in this sense: both words introduce a message that, from the speaker's point of view, should help the addressee understand the speaker's statement or why the speaker did what is conveyed in the statement. It is not coincidental that in many of our examples *a to* could be replaced with *ved'*:

(39) *Pojdem domoj, a to <ved'> zavtra rano vstavat'.*
    'Let us go home, because we have to rise early tomorrow'

(40) *Idi v dom, a to zamerzneš <zamerzneš ved'>.*
    'Come inside, or you will freeze'

However, *ved'* is freely used in the cases where the speaker quotes a 'positive' argument, and the use of *a to* becomes impossible:

(41) a. ** Davajte ja perezvonju. A to mne eto besplatno*.
    b. *Davajte ja perezvonju. Mne ved' eto besplatno*.
    'Let me call you back, because it's free for me'

3.3

We were trying to demonstrate that the following concepts are contained within the causal (broadly understood) *a to:* (1) *a to* appears in the part of the statement where the speaker clarifies why the statement was made or why the speaker did what is conveyed in the statement. The clarification is made through (2) pointing out a certain circumstance which causes the undesirable consequences that could have followed in case the speaker had not done what he had done.

Bibliography

Beloshapkova V.A. Predlozheniya al'ternativnoi motivacii v sovremennom russkom yazyke // *Issledovaniya po sovremennomu russkomu yazyku*. M.: Izdatel'stvo Moskovskogo universiteta, 1970. P. 13-29.

Boguslavskaya O. Yu., Levontina I. B. Smysly 'prichina' i 'cel' v estestvennom yazyke // *Voprosy Yazykoznanija* 2004, № 2.

Bonno K., Kodzasov S. V. Semanticheskoe var'irovanie diskursivnyh slov i ego vliyanie na linearizaciyu i intonirovanie (na primere chastic *zhe* i *ved'*) // *Diskursivnye slova russkogo yazyka: opyt kontekstno-semanticheskogo opisaniya*. M., 1998.

Israeli A. *The meaning and polysemy of the alternative conjunction **a to***. Manuscript.

Inkova-Manzotti O. Encore sur la conjonction Russe *A TO*. http://www.academia.edu/1381390/Encore_sur_la_conjonction_russe_a_to

Kolosova T.A. O signalah nerazvernutosti nekotoryh implicitnyh slozhnyh predlozhenii // *Sintaksis predlozheniya*. Kalinin. 1980.

Levontina I. B. Chastica *ved'*: zagadki sochetaemosti (On the co-occurrence of one Russian particle) // *Trudy mezhdunarodnogo seminara Dialog 2005 po komp'yuternoi lingvistike i ee prilozheniyam. Tom 1*. M., "Nauka", 2005.

Paducheva E.V. *Vyskazyvanie i ego sootnesennost' s deistvitel'nost'yu*. M., Nauka, 1985.

Podlesskaya V.I. INACHE, A TO, A NE TO: rezumptivnye soyuzy kak sposob vyrazheniya otricatel'nogo usloviya // *Slozhnoe predlozhenie: tradicionnye voprosy teorii i opisaniya i novye aspekty ego izucheniya*. Vyp. 1. M., 2000.

Sannikov V.Z. *Russkie sochinitel'nye konstrukcii*. M.: "Nauka",1989.

Sobinnikova V.I. Slozhnye predlozheniya s soyuzom *a to* v russkih i ukrainskih govorah // *Materialy po russko-slavyanskomu yazykoznaniyu. T. 4*. Voronezh. 1969.

Uryson E.V. *Soyuzy a to i a ne to: pochemu v nekotoryh kontekstah oni sinonimichny* http://www.dialog-21.ru/digests/dialog2008/materials/html/82.htm

Uryson E.V. Sostavnye soyuzy A TO i A NE TO: vozmozhnosti semanticheskogo kompozicional'nogo analiza// *VoprosyYazykoznanija*, 2010, No. 1. P. 61-73.

Pairing Semantic and Communicative Structures for Paraphrase Generation in a Meaning-Text Linguistic Model

Jasmina Milićević

Dalhousie University, 6135 University Avenue
Halifax, Nova Scotia, B3H 4P9, Canada
jmilicev@dal.ca

Abstract

The paper takes a formal look at the interaction of propositional semantic information and communicative information in the generation of paraphrases within a Meaning-Text linguistic model by examining well-formedness constraints on the pairings of a specific Semantic Structure with a number of Semantic-Communicative Structures that can logically be superimposed on it.

Keywords

Paraphrasing, "Semantic Structure ~ Semantic-Communicative Structure" Pairings, Text generation

1   The Problem Stated

In this paper I will look into ways in which the propositional semantic information and the communicative semantic information can be "paired up" during the construction of Semantic Representations of sentences. The propositional meaning of a sentence S is a description of the state of affairs being mentioned in S (who did what to whom, when, where, why, etc); within Meaning-Text framework (Mel'čuk, 1974, 2012; Kahane, 2003), it is modeled by the Semantic Structure (SemS), see below. The communicative aspect of S's meaning is a presentation of S's semantic content in terms of its salience in a particular communication act (what is selected by the Speaker as topic, focus, etc., of the exchange), modeled by the Semantic-Communicative Structure (Sem-CommS).[1]

A given SemS can in principle be paired up (alternatively) with several different Sem-CommSs; varying the Sem-CommS for a given SemS is a major source of paraphrases, in

---

[1]   A full-fledged Semantic Representation additionally contains a Rhetorical Structure and a Referential Structure, not considered here.

Meaning-Text-based linguistic models (Iordanskaja *et al*., 1991; Iordanskaja, 1992) and elsewhere (Elhadad *et al*., 1996).

At least the following three questions arise in connection with "SemS ~ SemCommS$_i$" pairings: 1) well-formedness constraints for such pairings; 2) constraints on the realization of a given well-formed pairing (i.e., on the choice of lexical units and syntactic constructions to verbalize it); 3) constraints on mutual substitutability of sentences realizing different pairings (of the same initial SemS). For lack of space, I will limit myself mostly to the discussion of the first question, which will be addressed by examining pairings of one specific SemS with a number of Sem-CommSs that can logically be superimposed on it. I will draw on the theory of communicative structure (*aka* information structure) expounded in Mel'čuk (2001), building upon Milićević (2002, 2007: 105-115 and 231-244), where these questions were raised, as well as Mel'čuk (2012: 386-391).

2 A Formal Look into Combinability of Semantic and Communicative Information: A Case Study

2.1 Data and Framework

2.1.1 The Initial SemS

The SemS used in the study, henceforth, *Initial SemS* (Fig. 1), underlies a family of paraphrases whose representative is sentence (1):

(1) *The media harshly criticized the Government's poor management of the situation*.

A SemS is a connected, oriented and (fully) labeled graph; graph nodes are labeled with names of semantemes (= lexical meanings of the language in question) and its arcs by distinctive numbers (1-6) indicating semantic dependency relations between a (quasi-) predicative semanteme and its arguments, or semantic actants (= SemAs).
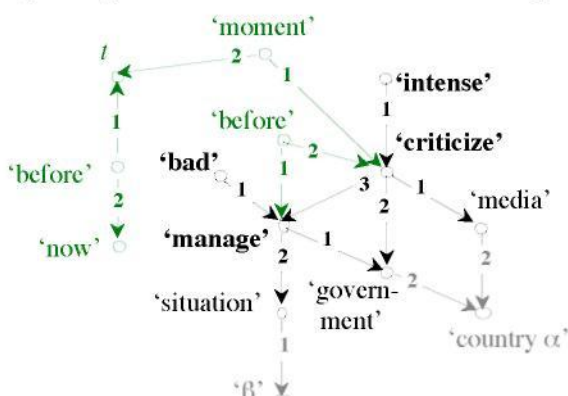

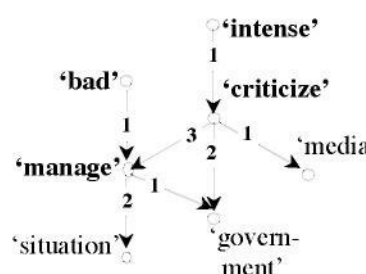Figure 1: Initial SemS   Figure 2: Simplified Initial SemS

The *Initial SemS* is organized around two (genuine) predicates, 'criticize' and 'manage'; the latter is an actant of the former, which means that the situation "managing" is embedded in

the situation "criticizing". The predicates have a shared actant: the quasi-predicate 'government' is the SemA 2 of 'criticize' and the SemA 1 of 'manage'. (This fact, as we shall see, is relevant for "SemS ~ Sem-CommS" pairing well-formedness and realization constraints.) The intensifying predicate 'intense' bears on 'criticize'; the qualifying predicate 'bad', bears on 'manage'.[2] A literal reading of the *Initial SemS*, taking 'criticize' as the entry node, is indicated in (2); semanteme configurations that do not have an explicit expression in the target sentences are bracketed:

(2) 'The media [of the country α] intensely criticized the Government [of α] for the bad management of the situation [involving β] by the Government [of α]'.

For ease of reading, a simplified *Initial SemS* (Fig. 2), from which the inflectional meanings and the meanings not explicitly expressed were removed, will be used in the paper.

2.1.2   *The Communicative Marking Used*

The Sem-CommS is a division of the SemS into communicative areas, such that each area 1) is marked by values of some communicative oppositions, and 2) possesses a communicatively dominant node (= CDN).

Out of the eight communicative oppositions proposed in Mel'čuk (2001), this study used only Thematicity and only two of its values: Theme (≈ Topic) and Rheme (≈ Comment); the third value, Specifier (≈ additional characterization of the Theme and/or the Rheme) was not considered for simplicity's sake, even though in principle it was possible.[3] All-rhematic configurations were not considered, either.

REMARK Another Sem-Comm opposition to some extent relevant here is Giveness, with the values Given (≈ information deemed to be referentially identifiable by the Addressee) and New (≈ information non referentially identifiable); it plays a part in the realization of articles and the pronominalization/ellipsis. By default, the Theme is Given and the Rheme is New.

Underlying questions [= UQ] were used to facilitate the readings of "SemS ~ SemCommS$_i$" pairings, providing a minimal context for their realizations. An UQ of type (3a) identifies (the part of the sentence expressing) the Theme, and one of type (3b) − the Rheme (or rhematic focus):

(3) a. *What about X?*   *[X]$_{Theme}$ did P*.
    b. *Who did P?*   *[X]$_{Rheme}$ (did)*   ⟨*[It was X]$_{Rheme}$ (who did it)*⟩ .

The CDN of a Sem-Comm area is the node to which the entire area can be semantically reduced − a sort of a minimal paraphrase of this area. CDN selection heavily influences the expression of a given semantic configuration; thus, the configuration 'intense' -1→ 'criticize' will be expressed as *harsh criticism*   ⟨*criticize harshly*⟩   if 'criticize' is selected as its CDN

---

[2]   Being a genuine or quasi-predicate (or else a semantic name) determines in part the communicative potential of a semanteme; cf. **2.2** below. On the opposition "(genuine) predicate ~ quasi-predicate", see Mel'čuk & Polguère (2007).

[3]   Since in the *Initial SemS* the situation "criticizing" is not an integral part of the situation "manage", the former could have been considered a Communicative Specifier in at least some of the pairings having the latter in the role of the "main fact", i.e. in the Rheme.

and as *harshness of [the] criticism* if the CDN is 'intense'. CDNs in representations (and their realizations in sentences) will be underscored.

2.1.3   The "SemS ~ SemCommS$_i$" Pairings Considered

Three major well-formedness requirements for "SemS ~ Sem-Comm$_i$" have been proposed so far in the Meaning-Text literature (not all of them are equally strict or as universally applicable, but we can ignore these details here):

1.   CDNs of the Theme and the Rheme should be linked via direct semantic dependency; since the Rheme is supposed to state something about the Theme, non-connectedness of their respective CDNs is an indicator of semantic incoherence.

2.   A Theme/Rheme area should be continuous – that is, all its nodes should be linked via semantic dependencies equally included in the area; in a discontinuous Theme/Rheme more than one node can potentially function as CDN, which makes it impossible to get one definite reading for the corresponding pairing.

3.   A monoactantial predicative semanteme cannot be the only element of the Theme unless it is focalized; stating something about such a "stranded" predicate is either uninformative or semantically bizarre.

It is assumed that a "SemS ~ Sem-Comm" pairing featuring either one of these irregularities will lack acceptable (semantically coherent and/or grammatically correct) realizations and/or an appropriate UQ.

The study considered the pairings of the *initial SemS* with 26 Sem-CommSs (consisting only of Theme ~ Rheme and with only one choice of CDN per area), looking to corroborate the above constraints and possibly find some new ones.[4] Table 1 shows the partial makeup of the pairings considered and the distribution of well-formed vs. ill-formed pairings, with an indication of the cause of deficiency; the two asterisked well-formed pairings did not have acceptable realizations for lack of appropriate lexical expression means.

---

[4]   I did not consider all logically possible, i.e. mechanically obtainable pairings since many of them would be blatantly ill-formed. The total number of pairings of a 7-node SemS like ours with the minimal Sem-CommSs as described above is 2^7-1, i.e., 127 (128 minus 1 solely Thematic configuration – which does not underlie a "normal" sentence and therefore is not our legitimate target).

CDN of the Rheme is	Well-formed P.	Ill-formed P.		Total
a genuine predicate:				
'criticize'	4	2	[discontinuous theme]	6
'manage'	5	2	[discontinuous theme]	7
'intense'	1	1	[discontinuous rheme]	2
'bad'	1	1	[discontinuous rheme]	2
an actant of a genuine predicate:				
ASém$_{1/2}$ of 'criticize'	2	1	[discontinuous rheme]	3
ASém$_{1/2}$ of 'manage'	2**	1+1	[discontinuous rheme; stranded predicate]	4
				1
ASém$_1$ of 'intense'	0	1	[stranded predicate]	1
ASém$_1$ of 'bad'	0	1	[stranded predicate]	
	15	11		26

Table 1: Well-formed vs. Ill-formed "Sem ~ Sem-CommS$_i$" Pairings

In addition to these, 4 pairings were considered in which the semanteme 'government' 'appeared in an overlapping segment, i.e. was simultaneously in the Theme and the Rheme (which it could as the shared actant of 'criticize' and 'manage'). These pairings were checked for well-formedness and equivalence with their counterparts without overlaps (having the exact same configuration except for the overlap), included in the 26 pairings above.

2.1.4 Lexical and Syntactic Variation in the Realizations of "SemS ~ Sem-CommS$_i$" Pairings

Lexical variation found in the sentences realizing the pairings is summarized in Table 2; lexical units and collocations involved are described in terms of lexical functions (Wanner, ed. 1996).

Semantemes	Realizations
'criticize'	CRITICIZE; CRITICISM$_{=S0(CRITICIZE)}$: *level*=OpeRh$_{(CRITICISM)}$, *attract*=Caus2/3Func1$_{CRITICISM}$) ~ *targets*$_{=Func2CRITICISM)}$ [N$_{Y/Z}$]; CRITICAL$_{=A1(CRITICIZE)}$: *be*$_{= Oper1(CRITICAL)}$ ~ [of N$_{Y/Z}$]
'intense'	harshly ⟨severely⟩ $_{=Magn(CRITICIZE)}$; harsh ⟨severe⟩ $_{=Magn(CRITICISM)}$; highly=$_{Magn(CRITICAL)}$
'criticize' ←1- 'intense'	BLAST$_{(v)}$ ⟨SLAM⟩ $_{=//Magn(CRITICIZE)}$: FLAK$_{=//Magn(CRITICISM)}$: *draw*$_{= Oper2((FLAK)}$ ~
'manage'	MANAGE; MANAGEMENT$_{=S0(MANAGE)}$
'bad'	*poorly*$_{=AntiBon(MANAGE)}$; *poor* ⟨sloppy⟩ $_{=AntiBon(MANAGEMENT)}$
'manage' ←1-'bad'	MISMANAGE $_{=//AntiBon(MANAGE)}$; MISMANAGEMENT $_{=//AntiBon(MANAGEMENT)}$

Table 2: Lexical Variation in the Realization "SemS ~ Sem-CommS$_i$" Pairings

Syntactic variation is illustrated in Table 3. For an easier comparison, lexical variation is kept at a minimum; only realizations of well-formed pairings are included. It was due primarily to

communicative factors. The choice of the main predication follows largely from the choice of the CDN of the Rheme (e.g. 1-2 *vs*. 3 *vs*. 4, etc.), while the voice of the main verb depends on the choice of the CDN of the Theme (e.g. 1 *vs*. 2, etc.). Secondary predication is expressed by weakly subordinated relative clauses of two different types, which reflects the distribution of the corresponding semantic material into different Sem-Comm areas: adnominal relatives are a part of the expression of the Theme (e.g. 7-9), and sentential relatives a part of the expression of the Rheme (e.g. 10-11). Syntactic variation was also due to the alternative Government Pattern realizations, possible with some lexical items; thus, the verb CRITICIZE allows for a two- or three-actantial realization (e.g. 1-3 *vs*. 4, etc.).

	Examples of realization
Strong subordination	1-[*The G.'s poor <u>management</u> of the S.*]$_{\text{Th}}$ [*<u>was</u> harshly <u>criticized</u> in the M.*]$_{\text{Rh}}$ 2-[*The <u>M.</u>*]$_{\text{Th}}$ [*harshly <u>criticized</u> the G.'s poor management of the S.*]$_{\text{Rh}}$ 3-[*The M's harsh <u>criticism</u>*]$_{\text{Th}}$ [*targeted the G.'s poor <u>management</u> of the S.*]$_{\text{Rh}}$ 4-[*The M.'s <u>criticism</u> of the G. for its poor management of the S.*]$_{\text{Th}}$ [*was <u>harsh</u>.*]$_{\text{Rh}}$ 5-[*The harsh <u>criticism</u> of the G.s' poor management of the S.*]$_{\text{Th}}$ [*came form the <u>M.</u>*]$_{\text{Rh}}$ 6-[*The M.'s harsh <u>criticism</u> of the poor management of the S.*]$_{\text{Th}}$ [*targeted the <u>G.</u>.*]$_{\text{Rh}}$
Weak subordination	7-[*The <u>G.</u>, **who managed the S. poorly,***]$_{\text{Th}}$ [*<u>was</u> harshly <u>criticized</u> in the M.*]$_{\text{Rh}}$ 8-[*The <u>G.</u>, **(who was) harshly <u>criticized</u> by the M.,***]$_{\text{Th}}$ [*<u>managed</u> the S. poorly.*]$_{\text{Rh}}$ 9-[*The G.'s <u>management</u> of the S., **(which was) harshly criticized by the M.,***]$_{\text{Th}}$ [*was <u>poor</u>.*]$_{\text{Rh}}$ 10-[*The <u>G.</u>*]$_{\text{Th}}$ [*<u>was</u> harshly <u>criticized</u> by the M., **(which was) due** to it's poor management of the S.*]$_{\text{Rh}}$ 11-[*The <u>G.</u>*]$_{\text{Th}}$ [*<u>managed</u> the S. poorly, **which attracted** harsh criticism from the M.*]$_{\text{Rh}}$

Table 3: Syntactic Variation in the Realizations of "SemS ~ Sem-CommS$_i$" Pairings

REMARK Sentences such as *The G. managed the S. poorly **and** (it) was harshly criticized for it by the M.* ⟨*The M. harshly criticized the G.; the latter managed the S. poorly*⟩ , which make use of coordination, were not considered: their SemS differs slightly from our *Initial SemS* and reflects a different sentence planning strategy (different chunking). The two SemSs are quasi-equivalent and could be linked by means of a semantic paraphrasing rule (in this connection, see Milićević 2010).

2.2 Findings

In this subsection I will discuss and illustrate some of the results presented in Table 1 and the problem of pairings with overlapping segments.

REMARKS 1) In the sentences realizing the pairings shown below, the material presenting old information (with respect to the corresponding UQs) is pronominalized and the redundant material earmarked for ellipsis is indicated by a strikethrough. This makes the sentences sound more natural (as if they appeared in a running text) and also helps evaluate the substitutability of realizations from different pairings. 2) The original pairing numbers from the study are in square brackets.

2.2.1 Ill-formed Pairings: Discontinuity in a Sem-Comm Area

– A Discontinuous Theme
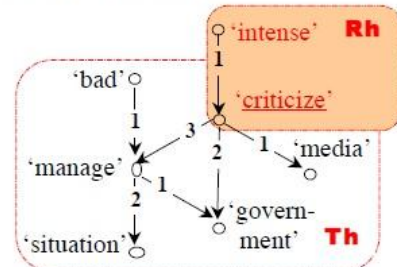

Pairing 1[6] — No viable UQ
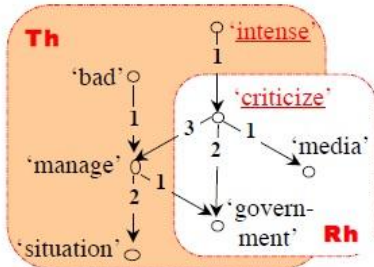
No realization

Pairing 2[5] — UQ: *What about the M. and the G.'s bad management of the S.?* / *What about the G.'s bad management of the S. and the M.?*

[*They*₌ The M.]Th [*harshly criticized it*₌ the G. the bad management ...]Rh
[*It*₌ The G. or The management]Th [*was harshly criticized by them*₌ The M.]Rh

The discontinuity in Pairing 1 is more severe: contrary to Pairing 2, its Theme nodes are not even indirectly connected (through a common governor in a different Sem-Comm area). As a result, Pairing 1 has no viable UQs or realizations, while Pairing 2 has both, albeit deficient. There are two UQs, each having a coordinate structure, since there are two candidates for the CDN of the Theme (in fact three, but two is bad enough and illustrates the point) whose relative salience is unclear; some sort of supplementary communicative marking — Subthemes? — is needed to resolve this. In the realizations, the elements implementing the Theme are partially integrated into the prosodic phrase formed by the implementation of the Rheme.

An UQ of the type *How did the M. react to the G.'s bad management of the S.?* would be more natural; however, since this is a WH-question, it entails the focalization of the Rheme, which, combined with the presence of a modifier, results in an "information overload" (emphatic prosody is marked by ALL CAPS): [?]*They harshly CRITICIZED it* ⟨[?]*They HARSHLY CRITICIZED it.*⟩
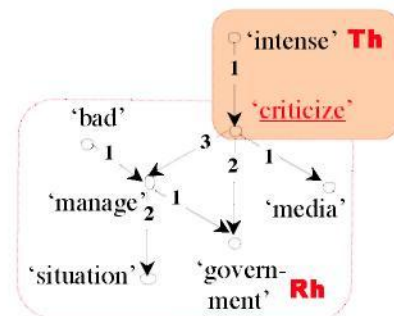
—A Discontinuous Rheme

Pairing 3[17] — UQ: *What about the G.'s management of the S.?*
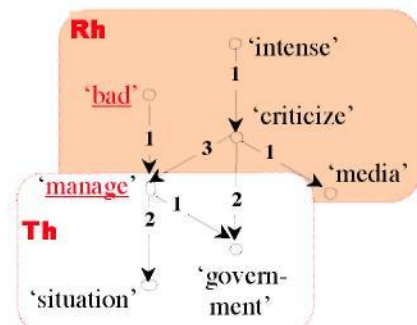
[*It*₌ The G.'s management of the S.]Th
[*was poor **and** (it) attracted harsh M.'s criticism*]Rh

Pairing 4[20] — UQ: *What about intense criticism?*

[*It*₌ Harsh criticism]Th
[*came from the M. **and** targeted the G.'s poor ...*]Rh

The only way to "repair" the discontinuity in Pairing 3 is to force a coordinate realization of the predicates 'bad' and 'criticize', but this realization actually has a different SemS from the one we see in this pairing (cf. REMARK after Table 3).

Pairing 4 (mirror-image of Pairing 2) has a somewhat unnatural UQ (perhaps acceptable as a request to repeat what was said); here too, realizations are possible only with added coordination. A multiple WH underlying question (*Who harshly criticized whom for what?*) would sound more natural, but (besides adding focalization) we still could not avoid "doctoring" the initial SSem to get acceptable realizations; cf. a two-sentence answer: *The media (did). They blasted the G. for the way it managed the/some S.*

2.2.2 Ill-formed Pairings: "Stranded" Predicate

The problem with Pairing 5 is of the type already known: a qualifying predicate does not make for a good Theme by itself. This kind of irregularity is irreparable; thus, the realizations one can get here at the cost of added focalization and some rather complex semantic processing (cf. the link between *way/how* and [*manage*] *poorly*), are too clumsy to be worth the trouble: *Poor is the <u>way</u> in which the/some S. was managed by the G., who got blasted for it by the M.* 〈*Poorly is <u>how</u> the G. managed the/some S., for which it got blasted by the M.*〉
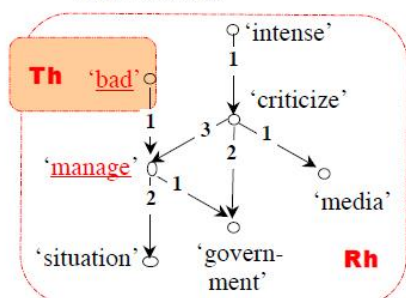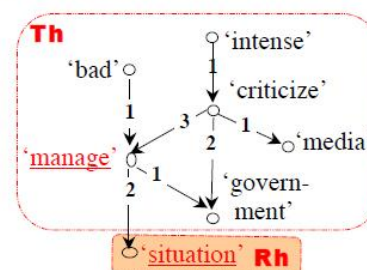
Pairing 6 has as the only node of the Rheme a taxonomic semanteme 'situation' — i.e., a semanteme of a very general meaning — not informative enough to be the CDN of the Rheme. The problem would not arise if 'situation' were qualified in some way (*the current* 〈*the aforementioned*〉 ~) or if it had its actantial slot saturated (*human rights* 〈*financial*〉 ~). This type of irregularity — stranded predicate in the Rheme — had not been observed previously. It seems that we are dealing here with a different type of predicate; 'the situation' would have been just fine as the unique node of the Theme (if marked as Given, i.e., carrying the default Giveness value for Themes), cf.: [*The <u>situation</u>*]_{Th} [*is <u>critical</u>* 〈*<u>unbearable</u>, <u>under control</u>*〉]_{Rh}.

2.2.3 Well-Formed Pairings Inexpressible Due to Lexical Lacunae

Pairing 7[21]
UQ: *What about the bad management of the S.?*

Pairing 8[22]
UQ: *What about the bad management of the S., criticized…?*
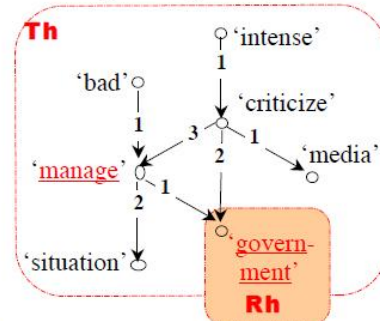
In both cases, the realizations are bad because the corresponding lexical material is missing: in English, there is no support verb linking the lexeme MANAGEMENT (or MISMANAGEMENT) in the subject position, which it occupies as the implementation of the CDN of the Theme, with the expression of its first actant (in lexical functional terms, there is no $Func_1$, not even $Oper_1$ to use in the passive). We could try *The bad management is to be blamed on* ⟨*is attributable to, is a doing of*⟩ *the Government*, but these realizations add meaning not specified in the *Initial SemS*. This SemS ~ Sem-Comm configuration is not bad as such; with different lexical material, it would yield quite normal sentences: *This ACTION **was taken*** ⟨*The MISTAKE **was made**, The FRAUD **was perpetrated***⟩ *by the Government*.[5] See also Pairing 10 below. This case may at the first glance seem similar to that illustrated by Pairing 6, but actually it is not: the other problem is deeper.

2.2.4 Status of Pairings with Overlapping Segments

The question of overlaps between Sem-Comm areas arises (in our case) because of the presence in the SemS of an actant "shared" by two predicates, which also happens to be triggering pronominalization and ellipsis operations, performed at a later stage of sentence synthesis. This needs to be taken into account when evaluating realizations from pairings featuring overlapping Sem-Comm in an attempt to assess their well-formedness: pairings that at a first glance seem ill-formed because of redundancies in their realizations may actually be all right if we remember that the rules of pronominalization and ellipsis have yet to be applied. Another thing to keep in mind is that there is a procedure for "reading" the semantic network and processing semantic dependencies and that this procedure should be able to eliminate some redundancies by precluding, when necessary, multiple readings of a network node.

---

[5] An analogous case was noted in Milićević (2002: 41): mental state predicates do not seem to allow for a nominal expression from a configuration similar to that of Pairings 7/8 (*The CERTAINTY*mental state *of Ulysses' return is* *of Penelope* ⟨*Penelope's*⟩ vs. *The COMMUNICATION*act *of Ulysses' return comes from Penelope*).

Pairing 9[1]

UQ: *What about the G.'s bad management of the S.?*
No overlaps
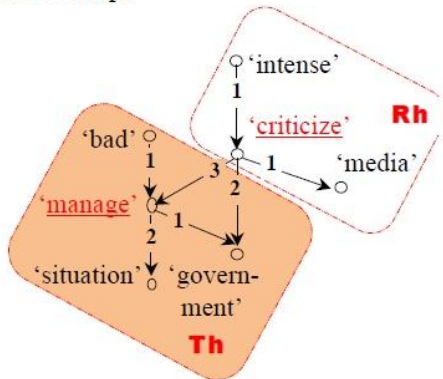
Pairing 9′[1′]

UQ: *What about the G.'s bad management of the S.?*
'government' in an overlapping area

$[It_{=The\ poor\ \underline{management}\ of\ the\ S\ by\ the\ G.}]_{Th}$
$[attracted\ harsh\ M.'s\ criticism.]_{Rh}$
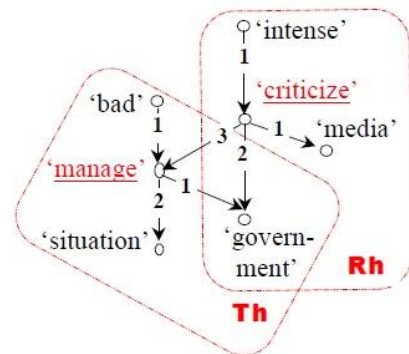
$[It_{=The\ poor\ \underline{management}\ of\ the\ S\ \text{by the G.}}]_{Th}$
$[attracted\ harsh\ M.'s\ criticism\ \textbf{of the G}.]_{Rh}$
$[It_{=The\ poor\ \underline{management}\ of\ the\ S\ \text{by the G.}}]_{Th}$
$[attracted\ harsh\ M.'s\ criticism\ \textbf{of the G}.]_{Rh}$

If we assume that in Pairing 9 the arc 'criticize'-2→'government' is not read (because it crosses the Rh/Th boundary), and that in Pairing 9′ it is, then they yield slightly different realizations, provided, of course, that ellipses are performed in the latter case. But I am not sure that this assumption is legitimate. In terms of processing costs, there is no clear advantage one way or another.

Pairing 10[19]

UQ: *What about the M.'s criticism of the bad…?*
No overlaps
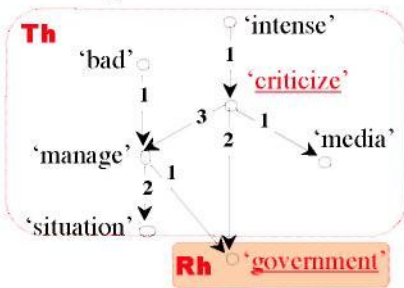
Pairing 10′[19′]

UQ: *What about the M.'s criticism of the G.'s bad…?*
'government' in an overlapping area

$[It_{\text{The M.'s harsh criticism of the poor management of the S.}}]_{Th}$
$[targeted\ \langle was\ leveled\ at\rangle\ \textbf{the G}.]_{Rh}$
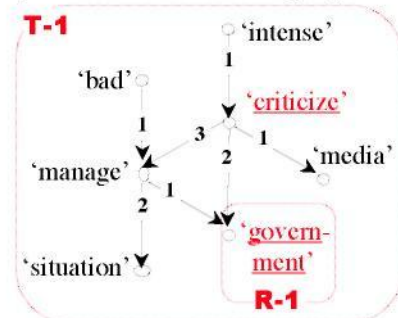
$[It_{\text{The M.'s harsh criticism \textbf{of the G.} for the G.'s poor management of the S.}}]_{Th}$
$[targeted\ \langle was\ leveled\ at\rangle\ \textbf{the \underline{G}}.]_{Rh}$

In order to get non-redundant realizations here, either the arc 2 from 'criticize' should not be read twice and the arc 1 from 'manage' should not be read at all (Pairing 10) or else two ellipses should be performed (Pairing 10). We see more redundancies with the realizations of these pairings due to the sentence-final position of the element expressing the CDN of the Rheme.

However, the realization of Pairing 10′ remains redundant with respect to the UQ; which indicates that 'government' should not appear in the Theme so as not to be mentioned in the UQ — in other words, Pairing 10′ is at best dubious. This of course does not mean that all overlaps are bad. By all means, the problem needs to be looked into in more detail.

3 Summary and Conclusion

Three major factors were found to be relevant for well-formedness of "SemS ~ Sem-Comm" pairings and acceptability of their realizations.

(1) Properties of semantemes. Depending on their logical nature (genuine vs. quasi-predicate), semantic class (qualifying vs. action vs. mental-state predicate, etc.) and level of generality of meaning (taxonomic vs. "ordinary" semanteme), some semantemes are more or less fit to appear in a specific communicative configuration (e.g., the "stranded predicate" phenomenon, discussed above) or in a specific communicative role (e.g., never as the Theme or always as the Rheme focus [Mel'čuk, 2012: 390]).

(2) Connectedness of semantemes within a Sem-Comm area. Discontinuities within the Theme or the Rheme are unacceptable to different degrees and with varying "repair" possibilities. In particular, discontinuous rhemes can be almost "normal" (a sentence as a reply to a multiple WH-question has such a rheme [Mel'čuk 2012: 314]); some languages accept discontinuous themes (e.g., French, in so-called *segmented sentences*). However, some sort of supplementary communicative marking is necessary in all of these cases.

(3) Level of tension between semantic and communicative dominance. In principle, the two dominance types can go in the opposite directions, but a bigger discrepancy between them (items that are highly prominent communicatively are deeply embedded propositionally) will result in a more convoluted (= syntactically and lexically involved) expression or may preclude it altogether.

Verbalizing some problematic pairings was possible at a cost of "forced" communicative marking (focalization in several cases), restructuring (use of coordination in case of some pairings with discontinuous rhemes) or addition of meaning (use of material accessed via semantic decomposition, such as taxonomic semantemes 'way'/'how' in the case of Pairing 5). Actually, considering the speed with which we have to build our SemRs, this may well be what "regular" speakers do: construct less than perfect "SemS ~ Sem-CommS" pairings and then repair them by adding or subtracting material during realization.

A well-formed "SemS ~ Sem-CommS" pairing may be non-realizable because the appropriate expression means are unavailable in the language considered; we saw two examples of such pairings above (Pairings 7 and 8). This problem needs to be dealt with in terms of constraints on lexicalization (and arborization) rules. A different type of lexicalization problem that was highlighted is that of pronominalization and ellipsis (partially) driven by communicative factors.

Finally, mutual substitutability of realizations from different "SemS ~ Sem-CommS" pairings (especially if we consider their variants with pronominalization and ellipsis) seems to be severely limited. This is not unexpected since we are dealing here with paraphrases in the broad sense only, as opposed to paraphrases in the narrow sense, which share both the SemS and the Sems-CommS (Milićević 2007: 73), but a separate study is needed to determine what this distinction really entails.

Acknowledgements

Bibliography

Elhadad, M., McKeown, K. & Robin, J. 1996. Floating Constraints in Lexical Choice. *Computational Linguistics* 22; 1-45.

Kahane, S. 2003. The Meaning-Text Theory. In: Agel, V. *et al*., eds. *Dependency and Valency. An International Handbook of Contemporary Research*, vol. 1. Belin/New York: De Gruyter; 546-570.

Iordanskaja, L., Kittredge, R. & Polguère, A. 1991. Lexical Selection and Paraphrase in a Meaning-Text Generation Model. In: Paris, C. *et al*., eds. *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Boston: Kluwer; 293-312.

Iordanskaja, L. 1992. Communicative Structure and its Use during Text Generation. *IFID* 17/2:15-27.

Mel'čuk, I. 2012. *Semantics. From Meaning to Text*. Amsterdam/Philadelphia: Benjamins.

Mel'čuk, I. 2001. *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*. Amsterdam/ Philadelphia: Benjamins.

Mel'čuk, I. 1974. *Opyt teorii lingvističeskix modelej Smysl ~ Tekst*. Moskva: Nauka.

Mel'čuk, I. & Polguère, A. 2007. Prédicats et quasi-prédicats sémantiques dans une perspective lexicographique. *Revue de Linguistique et de Didactique des Langues* (LIDIL) 37: 99-114.

Milićević, J. 2007. *La paraphrase. Modélisation de la paraphrase langagière*. Bern: Peter Lang.

Milićević, J. 2002. Communicative Structure of Sentences as a Means of Controlling the Generation of Paraphrases. In: Gelbukh, A., ed. *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics (CICLing'2000)*. Mexico-City: Instituto politecnico nacional; 36-48.

Milićević, J. 2010. Extraction de paraphrases sémantiques et lexico-syntaxiques de corpus parallèles bilingues. In: L'anglais, Ph. *et al*., eds. *Actes de la 17e conférence Traitement Automatique des Langues Naturelles* (TALN 17), Montréal: Université de Montréal; pp. 10.

Wanner, L., ed. 1996. *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins.

Indefiniteness and Weak Definiteness in Russian

Elena V. Paducheva

Russian Academy of Sciences, Institute of scientific and technical information
Moscow
elena.paducheva@yandex.ru

Abstract

The paper is concerned with the semantics of indefiniteness in Russian, including WEAK DEFINITENESS (= HALF-DEFINITENESS, = WEAK INDEFINITENESS), and in the first place with the Russian adverb *odnaždy* 'once upon a time'. Investigations using the English term and concept of weak definiteness are also taken into consideration. It is suggested that the Russian term "weak definiteness", which is applied to the class of pronouns wholly absent in English, should rather be translated into English as "weak **in**definiteness", the English term "weak definiteness" being applied to phenomena of slightly different nature.

Keywords

Indefiniteness, non-referentiality, weak definiteness, discourse, register of interpretation

1 Russian *odnaždy* and the Concept of Weak Definiteness

Until recently the adverb *odnaždy* had never been an object of close analysis. The fundamental step towards apprehension of its semantics and combinability was made in the paper (Иорданская, Мельчук 2013), where the attention was paid to the fact that *odnaždy* is related – not only formally but also semantically, – to the word *odin* 'one' in the meaning of weak definiteness (Падучева 1985: 212-215). In Russian the two terms, WEAK DEFINITENESS and WEAK INDEFINITENESS, are used as synonyms, which is easy to explain: Ju.I.Levin, who was the first to describe the corresponding class of pronouns explicitly, see (Левин 1973), calls them HALF-INDEFINITE. Russian weak definiteness is, in fact, half-definiteness; namely, it is definiteness for the speaker and indefiniteness for the addressee. The term used in (Иорданская, Мельчук 2013), a paper written in Russian, is *slabaja neopredelennost'* "weak indefiniteness"; I prefer the Russian term *slabaja opredelennost'* "weak definiteness" because the speaker has the priority over the addressee. In any case, Russian weakly definite pronouns constitute a subclass of indefinite pronouns.

Weak definiteness is the pearl of the Russian linguistic model of the world. In general, the sphere of indefiniteness is developed in Russian as in no other language, see (Падучева 1996).

(Wierzbicka 1992) bears the responsibility for the idea that the degree of **development** of a certain semantic sphere can be different in different languages; see (Зализняк, Левонтина, Шмелев 2003) on Russian linguistic model of the world where several such well-developed semantic spheres are considered. Thus, the problem goes far beyond a single word and a single language.

The observation made in (Иорданская, Мельчук 2013) concerning weak definiteness of *odnaždy* transfers this word from the category of semantic isolates to an entity that occupies a well-defined place in the system of referential oppositions in Russian.

The term "weak definiteness" is widely used outside Slavic linguistics. Therefore, I have to begin with its use in the English-speaking linguistic community.

2   Definiteness in English

According to generally accepted definitions, DEFINITENESS in English is a morphosyntactic, i.e. grammatical category. For example, NPs with the determiner *the* are definite, and NPs with the determiner *a* are indefinite. In possessive NPs definiteness is not expressed unambiguously, cf. *my father* [definite] and *John's daughter* [not definite if John has more than one daughter]. A reliable syntactic correlate of definiteness is inability to appear in the existential *there* construction – as a rule, definite NPs cannot appear in this construction:

a. There is a student in the garden.
b. *There is *him /*John /*the student /*every student in the garden.

Now, in Russian, though it lacks articles, semantic opposition "definiteness *vs.* indefiniteness" is also relevant, though it reveals itself differently; suffice it to mention the Genitive of Negation construction, which is essentially linked with indefiniteness of the NP.

3   Uses of the Term "Weak Definiteness" in English

Different uses of the term "weak definiteness" exist in English-speaking linguistic tradition. I illustrated these uses by examples from different languages.

3.1   Weak Definiteness as Generic Reference

The most obvious case is constituted by such examples (from Carlson, Sussman 2005) as (1), (2):

(1)   a. *Could you please open **the window*** [in a room with several windows];
        b. *<Where is Mary?> She went to **the drugstore*** [there are several drugstores in the vicinity];

Similar class of NP uses can be found in Russian.

(2)   a. *Tebe nado pojti **k vraču***
        'you ought to go *to the doctor*';

   b. *Otec pošel **v bank***
   'dad went *to the bank*'.

In (Aguilar-Guevara & Zwart 2010) this phenomenon is interpreted as REFERENCE TO KINDS. These NPs are called GENERIC; they were extensively studied in the 80's (see Шмелев 1984, Падучева 1985: 97ff).

Generic NPs are characterized by several specific features of linguistic behavior:

   A.   If the action is repeated the object referred to does not need to be the same:

   (3)   a. *Mary xodit **v cerkov'** po voskresen'jam*
      'Mary goes to *the church* on Sundays';
      b. *My ljubim xodit' **v restoran***
      'nous aimons aller *au restaurant*'
      'we like going to the restaurant'

   B.   For different subjects the object referred to does not need to be the same:

   (4)   *Ted i Bob v odin i tot že den' popali **v bol'nicu***
      'Ted and Bob got to *the hospital* on the same day'.

   C. "Sentences with weak definites usually carry more information than what seems to be conveyed by the straightforward composition of its constituents. This **meaning enrichment** is stereotypical in the sense that it invokes the most common circumstances under which the referred event could happen, e.g. the most common circumstances under which one goes to the store is to go shopping" (Aguilar-Guevara, Zwart 2010).

In French (I am grateful to Isabelle Valloton for the examples) this enrichment is revealed in the choice of articles. In *tu devrais aller chez l'oculiste* 'you should go to an eye doctor' it does not matter how many eye doctors there are in the town or whether the person has a permanent eye doctor: the speaker wants someone to go to a representative of a class and get cured. But in the context of the word *consulter* the articles convey their regular definiteness opposition: *tu devrais consulter un oculiste* 'you should consult an eye doctor' means that there is a choice; *tu devrais consulter l'oculiste* means «the eye doctor I know or the one spoken about or the only one in town». The same goes for German: *zum Arzt gehen,* but *einen Arzt aufsuchen.* Cf Cieschinger 2011 on contraction in German.

   D. Possibilities for expressing anaphora are different in case of generic reference. Cf. example (5a), which is the Russian translation of (1a), and its continuation (5b).

   (5)   a. *Vy ne mogli by otkryt' **okno**?*
      'would you mind opening the window' [there are several windows in the room];
      b. *Imejte v vidu, čto **oni** otkryvajutsja s trudom*
      'mind that *they* are hard to open' [only plural of the pronoun can be used].

Many interesting issues were brought to light in this connection, cf. deviant anaphora in examples (6), (7) from (Падучева 1985) which demonstrate that co-reference between the generic and the referential term is impossible:

(6) *Na mne byl **frak, bez kotorogo** nikomu ne sovetuju vyxodit', daže na oxotu*
'I was wearing a *tail-coat, without which* I wouldn't advise anyone to go out, even on a hunt'. (Turgenev)

(7) *Lisica nikogda ne videla **l'va.** I vstretiv **ego,** ona ispugalas'*
'the fox never saw the lion. And having met *him* once she got scared' (Aesop's fables, translated by M.Gasparov)

Thus, weak definiteness in Russian examples (2) − (5) can also be looked upon as a case of reference to kinds (= generic reference).

3.2   Weak Definiteness as Reference to Relational Nouns

Another class of weak definites is constituted by what is called POSSESSIVE WEAK DEFINITES in (Barker 2010). It was proposed that a definite determiner in a possessive NP indicates that the head noun is a FUNCTIONAL concept. Thus, in example (1) we dealt with SORTAL NOUNS, while in possessives the head of the NPs under consideration is a RELATIONAL NOUN. In (Poesio 1994) a class of systematic exceptions to this generalization was noticed − no uniqueness implication in (8):

(8) a. *I hope the cafe is located on **the corner of a busy intersection**.*
b. *Then Superman smashed into **the side of a Marlboro-emblazoned truck**.*

For Russian, with no grammatical definiteness markers, there is nothing special in these uses of relational nouns. Indeed, *Putin's father* is an identifying description while *Putin's daughter* is not, there is an ambiguity, but this is not a question of grammar.

In (Barker 2010) no specific explanation for English possessive weak definite NPs is given. I would suggest that *the*-construction, which is prototypically used with functional relational nouns, in examples such as (8) has an EXPANSIVE use. Take the case of the Russian Genitive of Negation. Prototypically, it is used in the context of non-existence (Babby 1980); but it can also be used expansively in the context of ABSENCE IN THE FIELD OF VISION of the observer (Падучева 2011). The same expansion can be demonstrated on the verbs *vozniknut'* 'to appear' and *isčeznut'* 'to disappear': they can refer to arrival / disappearance in the world and, metaphorically, in the field of vision of the observer.

4   Russian Analogues of the English Notion of Weak Definiteness

Here are some issues in Russian, presumably connected with the English idea of weak definiteness or, more generally, weak reference.

4.1   Attributive Definite Descriptions

See example from (Летучий 2008), with the Genitive of Negation in the context of a definite description:

(9) *52-letnij Filipp Shut utverždaet, čto ne ubival **materi***
'Filipp Shut claims that he did not kill his *mother*'.

Prototypically, the target of the Genitive of Negation is a non-referential NP (Babby 1980). Still the Genitive sounds perfectly natural in (9), although *mat' Filippa Shuta* is, by itself, a perfectly definite description, with conditions of existence and uniqueness satisfied, so that the description unambiguously identifies the object. Thus, according to modern norms, Genitive of Negation should be infelicitous here. I explain this Genitive by the fact that the definite description has an ATTRIBUTIVE use here (Donnellan 1966). When a description is used referentially the speaker **has a particular object in mind** and makes a statement about this object. In case of attributive use the statement concerns **whoever or whatever** satisfies the description. Thus, definite reference cannot be restricted to existence and uniqueness conditions. A definite description does not guarantee genuine definiteness.

The notion of attributive reference gives a better approximation to the nature of the phenomenon than that of weak reference (though attributive reference is a specific kind of weak reference). In fact, there is a limitation on the context of use for utterances with attributively referential NPs: they can only have NON-DIRECT EVIDENTIALITY. It is no wonder: if the speaker does not know the OBJECT, but only its description, their statements about it can only be inferential based on the object's nature.

4.2   Distributive Definiteness

Consider an example from (Падучева 1985: 157):

> (10) *Vpišem v každuju iz okružnostej **mnogougol'nik i budem** udvaivat' čislo **ego** storon*
> 'let us inscribe *a polygon* in each of the circles and then double the number of *its* sides repeatedly'.

Clearly, there are several polygons in the described situation. Nevertheless, the anaphoric pronoun is in the singular. It is the context of distributivity that justifies a singular NP denoting plurality: the phenomenon of DISTRIBUTIVE DEFINITENESS occurs in the scope of a distributivity operator, cf. (Шмелев 1996: 85ff). In (11), on the other hand, there is no distributivity, and the hospital is the same for all the injured:

> (11) *Postradavšix uvezli v **bol'nicu**, gde im byla okazana pomošč*
> 'the injured were taken to *the hospital* where they received medical attention'.

By all means, sentence (10) does not exemplify the reference to kinds that we saw in ex. (1), (2).

Rules for expression of distributive definiteness are different in different languages; cf. the difference in grammatical number in Russian and English:

> (12) a. *Sovetniki nadeli **na nos** očki*
> 'the counselors put spectacles *on the nose';*
> b. *The counselors put spectacles **on their noses**.*

Note that distributive definiteness should be distinguished from SLOPPY IDENTITY (as in Karttunen's famous example with giving *one's paycheck to* one's wife or one's mistress).

4.3   Unidentified Relationship Between NPs; Partial Co-Extensiveness

In example (13), from (Bulygina 1990), the problem is the relationship between the proper name *Ippolotov* in the first sentence and the implied subject of the verb meaning 'kiss' in the second one (note that in Russian what we have is not a passive construction but a construction with zero subject).

> (13) *Proščajas' **Ippolotov** poceloval ej ruku. Vpervye v žizni ej **celovali** ruku*
> 'When parting *Ippolotov* kissed her hand. It was for the first time in her life that she *was kissed* on the hand'

In fact, there should be some relationship between the two NPs – otherwise the text would not have been coherent. But it is neither co-reference nor sloppy identity. To express the same idea with a non-zero argument we have to use a pronoun. The zero agent cannot be represented by an existentially quantified term: existential quantification is expressed by *-nibud'* pronouns, which are appropriate only in the scope of some VERIDICALITY CANCELLATION operator (another term is SUSPENDED ASSERTION, see (Weinreich 1963), translated into Russian by I.A.Melchuk as *snjataja utverditel'nost'*, see this term used in (Падучева 1985: 33, 94, 95, 215ff); so they are not appropriate in this context. A *-to* pronoun can be inserted, but it means that the agent is unknown, which is not the case:

> (14) *<...> Vpervye v žizni **kto by to ni bylo (*kto-nibud') (≠ kto-to)** celoval ej ruku*
> 'It was for the first time in her life that somebody kissed her hand'.

Pronouns with *by to ni bylo* constitute a special series, which is not included in (Haspelmath 1997) list or in our indefinite pronoun type nomenclature. They are semantically opposed to indefinites with *-nibud'*.

Cf co-reference relationships in the example (15) (the beginning of Pasternak's "Doctor Živago"):

> (15) *Šli i šli i peli "Večnuyu pamjat", i kogda ostanavlivalis', kazalos', čto ee po zalažennomu prodolžayt pet' nogi, lošadi, dunovenija vetra. Proxožie propuskali šestvie, sčitali venki, krestilis'. Ljubopytnyje vxodili v processiju, sprašivali: "Kogo **xoronjat**?" Im otvečali: "Živago"*

In the indefinite-personal clause, the first part of the sentence, zero subject of the indefinite-personal clause of the first part of the sentence *Šli i šli i peli* introduces an indefinite plural agent, and the same group of people is the implied subject of *ostanavlivalis';* as for *kazalos',* the implied subject can be the same, though the external observer is not improbable. The subject of *xoronjat* is uniquely recovered – it is not mentioned because it is evident, the sentence as a whole should be interpreted as elliptical rather than indefinite-personal. The only problem arises in connection with the implied subject of *otvečali:* it is not those who *šli i peli,* but **some of them**. The relationship is not that of co-reference but PARTIAL CO-EXTENSIVENESS. If we treat *by to ni bylo* pronouns as markers of universality then the relationship between the NPs *Ippolotov* and *kto by to ni bylo* in (14) can also be called partial co-extensiveness.

Thus, the notion of weak definiteness (and weak referentiality) in its English applications sheds light on some interesting phenomena in Russian. It is still a problem whether the same term should be applied to all of them or, rather, different phenomena are to be distinguished.

5 Indefiniteness in Russian

Now we must look at Russian indefiniteness from the Russian standpoint. Having no grammatically obligatory definiteness, Russian has a highly elaborated system of indefinite pronouns, which demonstrate what it means not to be definite. The following types of indefinite pronouns exist in Russian:

- INDEFINITE NON-REFERENTIAL pronouns; they are marked by *-nibud';* in English terminology this opposition corresponds to SPECIFIC VS. NON-SPECIFIC INDEFINITENESS;

- INDEFINITE REFERENTIAL pronouns; they are marked by *-to,* the Russian term – MESTOIMENIJA NEIZVESTNOSTI);

- HALF-INDEFINITE, or WEAKLY DEFINITE, i.e. INDEFINITE for the addressee but DEFINITE for the speaker; they are marked by the prefix *koe-* and convey the idea that the object is definite for the speaker but is supposed to be unknown to the addressee, as in *Ja tebe koe-čto (*čto-to) prines* 'I have brought something [you do not know what] to you' The pronoun *one* also conveys this idea fairly well: *On ženat na odnoj poljačke* 'He is married to one Pole'*;* the same semantics of weak definiteness is present in *nekotorye* 'some'. All these pronouns are called WEAKLY DEFINITE in (Падучева 1985: 90ff and 225ff).

In English, specific indefiniteness and non-specific indefiniteness are distinguished on the semantic basis only, while in Russian this semantic opposition is expressed formally. As for weak definiteness, it is, perhaps, hinted upon in the semantics of the pronoun *certain.*

There are some other indefinite pronouns in Russian that belong to none of these three types – *nekto, nekij,* both hardly translatable; *tot ili inoj* 'this or that' and others.

6 Russian *odnaždy* and its Semantics of Russian Weak Definiteness

Now let us turn to Russian *odnaždy* and its semantics of weak definiteness as it was defined in Russian contexts.

In fact, *odnaždy* is to be compared with *odin.* As well as *odin, odnaždy* is used not only as a reference marker but also as a numeral – *odnaždy* (Q) can mean «<Q took place> once».

In (Иорданская, Мельчук 2013) the following explication is given for *odnaždy: odnaždy* (Q) means that the temporal localization of Q, as well as Q itself is non-identifiable for the addressee. Non-identifiability of Q is further reduced to three properties of its referent:

- OPTIONALITY (for example, suicide differs in this respect from death, which is obligatory),

- REPEATABILITY (for example, heart attack differs in this respect from death, which is non-repeatable),

- ORDINARINESS (for example, visiting cinema differs in this respect from heart attack). It is argued that if situation Q lacks at least one of these properties *odnaždy* cannot be used in the corresponding sentence.

This definition can be conceived as an attempt to reduce the pragmatic aspects of weak definiteness, namely, definiteness for the speaker and indefiniteness for the addressee, to «objective» semantic properties of the situation the description of which constitutes the scope of *odnaždy*. Still, as we saw in section 4.1, even definiteness of the NP cannot be reduced to the fact that this NP is a definite description of the situation.

Thus, presumably, such a sophisticated concept as weak definiteness cannot be reduced to the conditions satisfied by the description of the situation constituting its scope.

The paper purports to demonstrate that *odnaždy* is an egocentrical word (Падучева 2011): it has pragmatic valencies that can be filled only by the participants of the speech act, the speaker and the addressee. Thus, its semantics and syntactics (i.e. combinability) can only be described with reference to the REGISTER OF INTERPRETATION: in different registers its semantics may be different.

Though a remarkable guess of L.N.Iordanskaja and I.A.Melčuk (namely, that *odnaždy* resembles *odin* because both belong to the sphere of weak definiteness) is correct, these two words differ in several important aspects. The word *odin* primarily belongs to dialogues, while the natural place for *odnaždy* is in the narrative. Moreover, *odnaždy* has a DISCOURSIVE meaning: its meaning potential is fully realized only in the text, for it has a steady cataphoric function. In the narrative context epistemic priority of the speaker over the reader is included in the semantics of the weakly definite *odnaždy,* and this reveals itself in its INTRODUCTIVE function, namely, in syntactically obligatory cataphora. It is obligatoriness of cataphoric reference, disappearing under some conditions, that determines the possibility or impossibility of using *odnaždy* in one or another context.

The paper (Иорданская, Мельчук 2013) is an important contribution to the study of the semantics of Russian weak definiteness; it amply demonstrates the contexts where *odnaždy* either cannot be used or, which is more interesting, loses its semantics of weak definiteness and becomes a marker of plain indefiniteness or even non-referentiality.

As for the terms, though in the context of Russian *slabaja opredelennost'* is a good choice, its English translation should rather be *weak **in**definiteness,* because the term *weak definiteness* has acquired an autonomous reputable tradition in the English-speaking linguistics.

Bibliography

Aguilar-Guevara, Ana & Joost Zwart. 2010. Weak definite and reference to kinds. *Proceedings of SALT*, vol. 20, 1-15.

Barker, Chris. 2005. Possessive weak definite. *Possession and Beyond: Semantics and Syntax*. 89-113.

Carlson, Gregory & Rachel Sussman. 2005. *Seemingly indefinite definite. Linguistic evidence: empirical, theoretical and computational perspectives*. 26-30

Cieschinger, Maria. 2011. Definite Reference. *Talk at the "Definiteness and Reference" seminar.* University of Osnabrück.

Donnellan, Keith S. 1966. Reference and definite descriptions // *The philosophical review*, v. 75, No. 3, 281-304.

Haspelmath M. 1997. *Indefinite Pronouns*. Oxford: Clarendon Press.

Poesio, Maximo. 1994. Weak definite. *Proceedings of the fourth Conference of Semantics and Linguistic Theory. SALT.*

Weinreich U. 1963. On the semantic structure of language // *Universals of Language* / Ed. J. Greenberg. Cambridge (Mass.): MIT Press, 1963. P. 114— 171. Reprinted in: U. Weinreich. *On semantics*. [Philadelphia:] Univ. of Pennsylvania Press, 1980. P. 37—96. Russian translation by I.A.Melchuk: Вейнрейх У. О семантической структуре языка // *Новое в зарубежной лингвистике. Вып. 5. Языковые универсалии*. М.: Прогресс: 163-249.

Wierzbicka A. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. N. Y.: Oxford Univ. Press.

Зализняк Анна А., Левонтина И.Б., Шмелев А.Д. 2005. *Ключевые идеи русской языковой картины мира*. М.: ЯРК.

Иорданская Л.Н., Мельчук И.А. 2013. Наречие *однажды*: неопределенный временной спецификатор. *Вопросы языкознания*, № 1, 22-37.

Булыгина Т.В. 1990. *Я, ты* и другие в русской грамматике // RES PHILOLOGICA. *Филологические исследования: Памяти акад. Г.В. Степанова* 1919–1986. М.; Л.

Левин Ю.И. 1973. О семантике местоимений // *Проблемы грамматического моделирования.* М.: Наука, 108-121.

Летучий А.Б. 2008. Генитив, классы объектов и пресуппозиции // *Исследования по теории грамматики*. Вып. 5. М.: Пробел-2000, 78-96.

Падучева Е.В. 1985. *Высказывание и его соотнесенность с действительностью*. М.: Наука, 1985. 6-th edition – Moscow: URSS, 2010.

Падучева Е.В. 1996. Неопределенность как семантическая доминанта русской языковой картины мира. *Problemi di morphosintassi delle lingue slave, v.5. Determinatezza e indeterminatezza nelle lingue slave.* Padova: Unipress, 163-186. http://lexicograph.ruslang.ru/TextPdf1/dominanta1_1996.pdf

Падучева Е.В. 2011. Эгоцентрические валентности и деконструкция говорящего. *Вопросы языкознания* № 3, 3-18. http://lexicograph.ruslang.ru/TextPdf1/egocentricals.pdf

Шмелёв А.Д. 1984. К типологии генерических именных групп // *Лингвистические исследования: Типология. Диалектология. Этимология. Компаративистика*. Ч.2. М.

Шмелёв А.Д. 1996. *Референциальные механизмы русского языка*. Тампере: University of Tampere Press.

The Role of Grammatical Constraints in Lexical Component in Functional Generative Description

Jarmila Panevová, Magda Ševčíková

Charles University in Prague, Faculty of Mathematics and Physics
Malostranské nám. 25, 118 00 Prague
{panevova,sevcikova}@ufal.mff.cuni.cz

Abstract

The present paper deals with selected morphological and syntactic features of Czech verbs. Working within the framework of Functional Generative Description (FGD), we demonstrate which features of lexical entries are required by the syntactic component of the description. In addition to the passive voice, traditionally described as diathesis, we briefly describe other kinds of proposed diatheses (resultative-1, resultative-2, and recipient). The constraints for their application will be present as features in the corresponding lexical entry; they will be a part of verbal paradigm in formal morphology. Regular operations within hierarchy of valency participants and their surface-syntactic positions are introduced into the grammatical component. Reciprocalization is characterized as a kind of shifting of valency complementations into the surface-syntactic position. We also specify the requirements of the verbs governing the infinitive and content clauses and point out to the interplay between the governing verb and modality of the content clause.

Keywords

Deep-syntactic structure, passive voice, diatheses, resultative, recipient diathesis, reciprocalization, modality of the dependent content clauses.

1   Introduction

Recent linguistic models are based on the division of labor between the lexical and the grammatical component. Though both components have been considered indispensable, individual linguistic approaches usually declare one of them as more central and important; the respective component is then elaborated more extensively as for the scope and depth of the issues involved (cf., for instance, the prevailing concern with grammatical issues in Chomsky's generative approach).

Since its original proposal (Sgall, 1967), the Praguian Functional Generative Description (FGD) adopts both the lexical and grammatical module; nevertheless, the main focus has been laid on the grammatical, esp. syntactic issues (Sgall et al., 1986). During the elaboration of the theory and most importantly during the application of the theory to the building up of the Prague Dependency Treebank (Hajič et al., 2001; Hajič et al., 2006), the lexicon has turned out to be of crucial importance. FGD is a multi-level description of language, where synonymous sentences are represented by the same representation on its uppermost level (tectogrammatics). On the contrary, ambiguous sentences have different representations on the tectogrammatical level, while they differ on some of the lower levels of representation.

In the present paper we want to demonstrate several issues where the grammatical component strongly requires an introduction of particular features and data in the lexicon. The aim of these constraints is to block the generation of ill-formed sentences and to contribute to the theoretical description of the syntactic and morphological properties of the verbs. The structure of the lexical entry has been studied in connection with the treatment of valency, coreference between valency complementations of the governing and the embedded verb (Section 2). In Section 3 several examples of lexical entries and the syntactic rules cooperating with them are given.

2 Reflections of the Grammatical Constraints in the Lexical Component of FGD

2.1 Valency Frames of Verbs and Diatheses

In FGD, a lexical entry of a verb contains a valency frame, consisting of inner participants (Actor, Patient, Addressee, Origin, Effect) and those free modifications that are determined as semantically obligatory for the respective verb by the so-called dialogue test (e.g. the modification of manner for the verb *chovat se* 'to behave'; Panevová, 1974/75). The inner participants are classified as (semantically) obligatory or optional with respect to the verb, for each of the inner participants its (morphological) form is further specified. The following features related to the valency and to the properties of the inner participants and obligatory modifications as well as their flexibility to express particular grammatical diatheses are to be specified in the lexical entry of the verb.

A. Surface **deletability** of a valency complementation which does not lead to ungrammaticality and which is not a textual ellipsis, see ex. (1).

B. Differences in valency behavior between **aspectual counterparts** (ex. (2a) vs. (2b)).

C. **Lexicalization** of some meanings of the verb that influence the valency, see ex. (3), where the Patient s*luchátko* 'receiver' is implied.

D. Possibility of **generalization (Gen)** of an inner participant, see ex. (4), − which means that everything written by the author is witty.

E. Though the **passive diathesis** is productive enough in Czech, there are some verbs (intransitive stative verbs, reflexives and some other), which cannot be passivized (e.g.

běžet 'to run', *spát* 'to sleep', *stát* 'to stand', *bát se* 'to be afraid', *plakat* 'to cry', *lhát* 'to lie'). There are also some transitives without passivization (e.g. *mít* 'to have'; *pít* 'to drink' in imperfective aspect). The feature "*-pass*" will be assigned to them in the lexical entry. Moreover, the verbs participating in the passive diathesis differ in which participant is shifted to the subject position of the passive construction: Patient is involved with the verb *přeložit* 'to translate' (ex. (5)), Addressee is converted into the subject with *informovat* 'to inform' (ex. (6)), the shift of an Effect into the subject position can be seen with *psát/napsat* 'to write' (ex. (7)).

F.   The **resultative diathesis** is a less productive, though still grammaticalized category of Czech verbs (Mathesius, 1925). It has two variants: the objective resultative (*res1*) consisting of the auxiliary *být* 'to be' and a passive participle, and the possessive resultative (*res2*) with the auxiliary *mít* 'to have'. The differences between these two types (syntactic and semantic) are described in Panevová (2011). Some of them could be understood from Sect. 3 (Ex. 3.1 and 3.2). The possible participation of the verb *otevřít* 'to open' in the possessive resultative (*obchod má otevřeno od 8 hodin* 'the shop is opened since 8 o'clock') will be included in its lexical entry as the feature "*+res2*". The resultative is prototypically used with perfective aspect of transitive verbs, while imperfective verbs with the stative meaning usually do not participate in this category (**je/bylo spáno* 'it-is/was sleeping', **je/má leženo* 'it-is/he-has laid', **je/má chlubeno* 'it-is/he-has boasted'). However, there are exceptions of the resultative combined with imperfective verbs (*je/má chráněno* 'it-is/he-has saved') and with intransitive verbs (*je/má namířeno* 'it-is/he-has aimed', *je/má našlápnuto* 'it-is/he-has trodden on'). Thus the possibility of the lexical item to form this category is to be marked in the lexicon by the feature "*+res1*" (for the objective resultative), see (7), or by "*+res2*" (for the possessive resultative), see (8), (9), (10).

G.   The number of verbs participating in **the recipient diathesis** is more limited than the number of verbs with the resultative diatheses; however, the recipient forms are constituted paradigmatically with the verbs which enter the semantic groups listed in Daneš (1985) and in Panevová et al. (ms.), see examples (11), (12).

(1) *Moji přátelé   právě   přijeli*.
     (my friend-NOM-PL   just   arrive-PST-PFV)
     'My friends have just arrived.'

(2a) *Včera   Jan   četl   až do půlnoci.*
     (yesterday John-NOM   read-PST-IPFV   till midnight)
     'Yesterday, John read till midnight.'

(2b) **Včera   Jan   přečetl   až do půlnoci.*
     (yesterday John   read-PST-PVF   till midnight)

(3) *Jan rychle   zavěsil.*
     (John quickly   hang-PST-PVF)
     'John has hung up quickly.'

(4) *Autor M.K.   píše   vtipně.*
(author M.K.   write-PRS-IPFV witty)
'Author M.K. writes with wit.'

(5) *Tento román   byl   překladatelem   přeložen*

(this novel_**Patient-**NOM   be-AUX-PST   translator-INS   translate-
PTCP-PASS
neadekvátně.
inadequately)
'This novel was translated by the translator inadequately.'

(6) *Turisté   byli   informováni   průvodcem*
(turist_**Addressee-**NOM   be-AUX-PST   inform-PTCP-PASS guide-INS
o   historii   zámku.
about history   castle)
'The tourists were informed by the guide about the history of the castle.'

(7) *O tom hrozném neštěstí   byla   novináři*
(about this terrible accident_**Effect** be-AUX-PST journalist-PL-INS
napsána   jen stručná zmínka.
write-PTCP-PASS –F only short remark-NOM-F)
'Only a short remark was written by journalists about this terrible accident.'

(8) *Na neděli   už   je   uvařeno.*
(*for Saturday   already   be-AUX-PRS   cook-PTCP-PASS-SG-N*)
'It is already cooked for Saturday.'

(9) Matka   už   má   na neděli   oběd
(Mother   already   have-AUX-PRS for Saturday   lunch-ACC-M
uvařen.
cook-PTCP-PASS-M)
'Mother already has cooked a lunch for Saturday.'

(10) *Jan   má   posun   zkoušky   schválen*
(John have-AUX-PRS shift-ACC-M   exam-GEN   confirm-PTCP-PASS-M
děkanem.
dean-INS)
'John has the shift of his exam confirmed by the dean.'

(11) *Očividně dostal   dávno   odpuštěno. (SYN2005)*
(Obviously get-AUX-PST-M   long time ago excuse-PTCP-PASS-N)
'He obviously has got to be excused long time ago.'

(12) *V domově budou   mít   obyvatelé*
 (in hostel   be-AUX-FUT   have-AUX-INF   inhabitant-PL
zajištěno   nejen ubytování,   ale i stravu. (SYN2006PUB)
arrange-PTCP-PASS-N   not only accomodation   but also food)

'In the hostel the inhabitants will have not only accommodation, but also food arranged.'

2.2 Valency Frames and Coreference

Verbs that take an infinitive construction as a complementation in a special valency position require the coreference between the participant triggered in the valency frame and the implied subject of the infinitive. Such requirements must be reflected in its valency frame. The member of valency frame controlling the (unexpressed) subject of its infinitive complement (as its antecedent) is marked in the lexical entry by the upper index "*-er*" (controller): *bát se* 'to be afraid' *Actor^{-er} (NOM), Patient (GEN/INF/Clause)*; *nařídit* 'to order' *Actor (NOM), Addressee^{-er} (DAT), Patient (Clause/INF)*. We present here only examples of two types of coreference between the controller and its controlee as prototypes of the requirements for the infinitive constructions in valency positions which are to be reflected in the given lexical entry; other types of Czech infinitive constructions with different types of coreference (control) are described in Panevová (1998). In (13), (14), the verb *bát se* 'to be afraid' in one of its meanings requires the identity (coreference) between its Actor and the subject of the embedded infinitive, while in (15), (16) for *nařídit* 'to order' the coreference between its Addressee and the subject of the infinitive is required. The differences between (13) and (14) and between (15) and (16) documents the fact that the role of the controlee is filled by the surface (unexpressed) subject:

(13) *Jan_i se bojí* [Sb_i] *jít do lesa sám.*
(John se-REFL is-afraid go-INF to forest alone)
'John_i is afraid [Sb_i] to go alone to the forest.'

(14) *Jan_i se nebojí [Sb_i] být zařazen*
(John se-REFL is-not-afraid be-AUX include-PTCP-PASS
do družstva pokročilých.
in team advanced]
'John is not afraid to be included in the advanced team.'

(15) *Učitel_i nařídil studentům_j [Sb_j] zorganizovat soutěž v matematice.*
(teacher order-PST student-PL-DAT organize-INF competition in mathematics]
'The teacher_i ordered the students_j [Sb_j] to organize a competition in mathematics.'

(16) *Rodiče nařídili synovi být rychle*
(parents order-PST son-SG-DAT be-AUX-INF quickly
připraven k odjezdu domů.
prepare-PTCP-PASS to leaving home)
'Parents ordered to their son to be quickly prepared for leaving for home.'

2.3 Valency Frames and Reciprocity

Another item which has been included into the lexical entry of Czech verbs in the lexical component of FGD is the information on the ability of valency complementations of the given

verb to enter a reciprocal relation. This ability is marked within the valency frame of the verbs and nouns by the "*Rcp*" index attached to the respective complementations.

The verb *líbat/políbit* 'to kiss', which has the valency frame *Actor^{Rcp} (NOM) Patient^{Rcp} (ACC)*, occurs in its basic (non-reciprocal) usage in the sentence *Pavel políbil Evu* 'Paul kissed Eva'. The reciprocalization (according to the *Rcp* indices) results in the sentence *Pavel a Eva se políbili* 'Paul and Eva kissed each other' (which is to be interpreted that Pavel kissed Eva and at the same time Eva kissed Paul). For the discussion about the boundary between "inherent reciprocals" and reciprocity diathesis see Panevová (1999) and Panevová & Mikulová (2007). The attachment of the *Rcp* index has the following syntactic consequences:

- one of the involved valency slots is omitted,

- the lexeme from the omitted slot becomes a part of a coordinated subject or the subject is in plural,

- the reflexive form of the verb is to be used (if the verb itself is not a reflexive tantum or a derived reflexive, see Panevová, 2008),

- optionally, the lexeme *vzájemně/navzájem* [mutually/one another] etc. can be added into the sentence.

The classification of reciprocalization within the FGD approach is in accordance with Meľčuk's (2006a: 215) arguments why reciprocals should not be classified as a voice.

2.4 Valency Frames and Modality of Dependent Content Clauses

Inner participants of some verbs can be expressed by a dependent (so-called content) clause. At the tectogrammatical level, dependent content clauses are classified as a Patient or an Effect with most verbs, less often as an Actor, and rather rarely as an Addressee or an Origin. The dependency of the content clause on the governing verb is expressed by a subordinating conjunction or by a pronoun (a pronominal adverb/numeral). The choice of the conjunction or pronominal is connected with the semantic properties of the governing verbs and with the modality of the dependent content clause.

A detailed analysis of the PDT 2.0 data has demonstrated that most of the verbs are compatible with a dependent content clause of one modality only (mostly with declarative modality, substantially less frequently an imperative or an interrogative dependent content clauses occur; cf. ex. (17) to (19), respectively). Only with a restricted number of verbs dependent content clauses of other modality types were used, most of them belong to verbs of communication; for instance, *diskutovat* 'to discuss' or *upozornit* 'to point out' (ex. (20a) with a declarative clause and (20b) with an imperative clause). Information on which modality type the verb is compatible with is proposed to be involved in the lexical entry of the respective verb. Three values for the description of the modality of dependent content clauses have been introduced: *declarative*, *imperative* and *interrogative*.

Dependent content clauses that express declarative modality are prototypically introduced by the conjunction *že* 'that', imperative dependent clauses by the conjunctions *aby* and *ať* 'so that', interrogative content clauses by the conjunctions *zda, zdali, jestli, -li* 'whether/if'. The

conjunction listed for imperative clauses as well as the conjunctions of interrogative clauses are considered synonymous.

Description of the relatively transparent relations between the governing verb, the modality of their dependent content clauses and the conjunction used in these clauses is complicated by the fact that there are verbs in Czech with which modality of the dependent content clause (and thus the conjunction) changes depending on the change of grammatical categories of the governing verb (ex. (21a) with the indicative governing verb and (21b) with the conditional).

With verbs like *upozornit* (ex. (20a,b)), several modality values are to be listed in the lexicon since the compatibility of these verbs with dependent content clauses of different modalities is involved in the lexical meaning of the verbs. On the contrary, only the basic modality is to be marked in the lexical entry of the verb *uvítat* in (21a,b) (i.e. *declarative*).

(17) *Na závěr schůzky ředitel dodal, **že** smlouva bude podepsána do týdne.* <declarative>
'In the end of the meeting, the director added **that** the contract will be signed in a week.'

(18) *Učitel nařídil žákům, **ať** zůstanou ve třídě.* <imperative>
'The teacher ordered the pupils **that** they **should** stay in the classroom.'

(19) *Studenti se ptají, **zda** se zítra koná přednáška.* <interrogative>
'The students are asking **whether** the talk is given tomorrow.'

(20a) *Upozornil je, **že** večerní představení začíná o hodinu později.* <declarative>
'He pointed out to them **that** the evening performance begins an hour later.'

(20b) *Upozornil je, **aby** o této skutečnosti nehovořili.* <imperative>
'He pointed out to them **that** they **should** not speak about this fact.'

(21a) *Opozice uvítala, **že** prezident zákon podepsal.* <declarative>
'Opposition welcomed **that** the president had signed the law.'

(21b) *Opozice **by** uvítala, **aby** prezident zákon podepsal.* <imperative>
'Opposition **would** welcome **that** the president would sign the law.'

3 Examples of Lexical Entries and Grammatical Rules Operating on Them

The Czech verb *připravit / připravovat* 'to prepare' has the valency frame (for one of its meanings) and other features analyzed in Sect. 2:

připravovat-IPFV / *připravit*-PFV $^{´+pas, ŕes1, ŕes2}$ *Actor (NOM), Patient ^{Sb}(ACC), Effect^{opt}(k + DAT, na + ACC)*

For the generation of the sentence in Ex. 3.1 and Ex. 3.2 with two types of the *res2* diathesis, the syntactic rules in (Rule I) are applied:

Ex. 3.1

Referent   už   má   připraveny   slidy   na prezentaci.
(speaker   already   have-AUX   prepare-PTCP/PASS   slide-PL-M   for presentation)
'The speaker already has his slides for presentation prepared.'

Ex. 3.2
Pavel   má   od matky   připravenu   večeři.
(Paul   have-AUX   from mother-GEN   prepare-PTCP-PASS-F   dinner-ACC)
'Paul has the dinner prepared by his mother'

Rule I
(i)   Predicate → AUX-mít$_i$ + -n / -t participle V$_j$ (*připraveny* 'prepared')
(ii)   Actor→Sb$_i$ (*referent* [speaker]) / ADV (od + GEN) (*od matky* 'from mother')
(iii)   Addressee → Ø (for the sentence in Ex. 3.1 without an Addressee)
        Addressee → Sb$_i$ (*Pavel* 'Paul') (for the sentence in Ex. 3.2)
(iv)   Patient (N$_j$ - ACC) → Obj$_j$-ACC (*slidy* 'slides' / *večeři* 'dinner')

The valency frame for one of the meanings of the Czech verb *slíbit* [to promise] represents a lexical item with possible reciprocalisation and with several possible diatheses:

slíbit-PFV / *slibovat*-IPFV$^{+pas,+res1,+res2-+recipient}$ 'to promise' Actor^{Rcp}(NOM), Patient^{Sb} (ACC/Clause/INF), Addressee^{Rcp}

The verb *slíbit* is compatible with the passive, both resultative and recipient diathesis. The rules for reciprocalization are described in an informal way in Section 2.3. The sentence in Ex. 3.3 with the reciprocity relation between the Actor and Addressee and Ex. 3.4 in recipient diathesis could be generated by them, the Rule II is applied for the generation of Ex. 3.4.

Ex. 3.3
Pavel a Táňa   si   slíbili   věrnost.
(Paul and Tanya   si-REFL   promise-PST-PL   faithfulness)
'Paul and Tanya promised to be faithful each other.'

Ex. 3.4
Pavel   dostal   za aktivní účast   slíbenu
(Paul   get-AUX   for active participation   promise-PTCP-PASS-F-SG-ACC
od trenéra   odměnu.
from coach   payment)
'Paul was promised to receive a payment for his active participation from the coach'

Rule II
(i)   Predicate → AUX-dostat$_i$ + -n /-t participle V$_j$ –SG-F-ACC (*slíbenu* 'promised')
(ii)   Actor→ADV (od + GEN) (*od trenéra* 'from the coach')
(iii)   Addressee → Sb$_i$ (*Pavel* 'Paul')
(iv)   Patient (N$_j$ - ACC) → Obj$_j$ - ACC(*odměnu* 'payment')

4 Conclusions

In the present paper, we focused on the role of the interplay between the lexical and grammatical component in the language description. Starting from an analysis based both on available theoretical descriptions and real corpus data, we have tried to explain that the analyzed grammatical categories of verb need to be treated adequately in the lexicon entries of the respective verb. In the lexical component of FGD an explicit mark of passivization with verbs has been introduced. The same treatment has been proposed for the resultative and recipient diathesis. In addition to the diathesis information, there are many features to be stored in the lexicon that are interconnected with individual valency complementations of the verb; for instance, surface deletability, possibility of generalization or possibility to be expressed by a dependent content clause with a certain modality must be specified for the respective complementations of individual verbs.

The proposed treatment of the issues discussed should allow for an economic and an effective interconnection of the grammatical and lexical module within the Functional Generative Description and should block ill-formed structures, for instance, at the output of the English-to-Czech machine translation procedure.

We agree with the comparison of FGD and MTT given by Žabokrtský (2005), where the similarities between these two models are described in detail. Multi-level and dependency approach are shared as well as Mel'čuk's idea that the deep-syntactic level unifies the synonymous sentence because synonymy is "one of the underlying intuitive notions for the whole of linguistics" (Mel'čuk, 2012: 48).

Acknowledgements

Bibliography

Daneš, F. 1985. *Věta a text*. Academia: Praha.

Hajič, J. 2001. *Prague Dependency Treebank 1.0*. Philadelphia: Linguistic Data Consortium.

Hajič, J. 2006. *Prague Dependency Treebank 2.0*. Philadelphia: Linguistic Data Consortium.

Mathesius, V. 1925. Slovesné časy typu perfektivního v hovorové češtině. *Naše řeč*, 9:200–202.

Meľčuk, I. 2006a. *Aspects of the Theory of Morphology*. Berlin – New York: de Gruyter.

Meľčuk, I. 2012. *Semantics. From meaning to text*. Amsterdam – Philadelphia: John Benjamins Publishing House.

Panevová, J. 1974/75. On Verbal Frames in Functional Generative Description. Part I. *Prague Bulletin of Mathematical Linguistics,* 22:3–40, Part II, *Prague Bulletin of Mathematical Linguistics*, 23:17–52.

Panevová, J. 1999. Česká reciproční zájmena a slovesná valence. *Slovo a slovesnost*, 60:269–275.

Panevová, J. 2008. Problémy se slovanským reflexivem. In *Česká slavistika. Příspěvky k XIV. mezinárodnímu sjezdu slavistů, Slavia*, 77 :153–163.

Panevová, J. 2011. O rezultativnosti (zejména) v češtině. In: *Gramatika i leksika u slovenskim jezicina.* Novi Sad, Beograd: Matica Srbska, Institut za srpski jezik. 165 – 176

Panevová, J. & M. Mikulová. 2007. On Reciprocity. *The Prague Bulletin of Mathematical Linguistics*, 87:27–40.

Panevová, J. et al. (ms.) *Mluvnice současné češtiny II. Syntax na základě anotovaného korpusu*. Praha: Karolinum.

Sgall, P. 1967. *Generativní popis jayzka a česká deklinace*. Praha: Academia.

Sgall, P. & E. Hajičová & J. Panevová. 1986. *The Meaning of the Setence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel Publ. Company.

Žabokrtský, Z. 2005. Resemblances between Meaning-Text Theory and Functional Generative Description. In *Proceedings of the 2nd International Conference of Meaning-Text Theory*. 549–557. Moskva: Slavic Culture Languages Publishers House.

On a typological parallel in the expression of 'imperativeness', 'optativeness' and 'condition' in Modern Russian and Old Russian

Anna Ptentsova

Moscow State University, department of Philology
anna.ptentsova@gmail.com

Abstract

The article is devoted to a syntactical parallel between Contemporary Russian constructions with *pust'* (*puskaj*) and the Old Russian construction *da + praesens*. Both of them express the same set of meanings: imperativeness, optativeness and condition.

Keywords

Semantics, Syntax; imperativeness, optativeness, condition; Modern Russian, Old Russian.

Semantics

It is known that 'imperativeness', 'optativeness' and 'condition' in Modern Russian can be expressed with syntactic constructions starting with the words *pust'* and *puskaj (let/may/dare)*. Therefore, (Russkaya grammatika, 1980) suggests that Russian imperative mood is expressed by syntactic particles *pust'* and *puskaj* accompanied by a verb in the form of 3 sg. or pl. Sometimes these particles may also be accompanied by a verb in the form of 1 or 2 sg. or pl. (in the sentences containing a conjugated verb). Finally, *pust* and *puskaj* may form the optative mood when used with the particle *by*.

Compare: "The meaning of the imperative combined with the optative ('I wish…') is often manifested in the *pust'* / *puskaj* particle constructions: *Byt' možet, vse eto tak budet,* / *Ja toč'no znat' ne mogu* / *No luč'še pust' eto budet* / *V more, č'em na beregu*; 'Perhaps everything is to be this way, I don't know for sure. But let it be at sea rather than on the shore' (Okudzhava)". Similarly, the particles in question can convey the imperative meaning combined with the meaning of obligation/necessity: *Pust' nikto ne igrajet s ognem!* 'Let no one play with fire!' These particles can also convey the meaning of requiredness combined with assumption/admission: *Puskaj umru s poslednej beloj vjugoj;* 'May I die with the last snowstorm' (Akhmatova)" (Russkaya grammatika, 1980).

We are going to draw some definitions of several meanings of the word **PUST'.** The dictionary entry is taken from the Prospect of the Active Dictionary of Russian. **Pust' 1.1**: *pust' A1* 'The speaker expresses their wish that A1 takes place' (*Pust' vsegda budet solnce!* 'May the sun always shine!'). It is important that in the case of a mathematical problem this lexeme acquires a special meaning: 'let us assume that' (*Pust' prjamaja AB parallel'na prjamoj CD* 'Let us assume that the line AB is parallel to the line CD'). Moreover, **pust' 1.1** can express a threat or a warning when pronounced with some specific intonation (*Pust' tol'ko poprobujet!* 'Dare he try!'). **Pust' 2.1**: *A1, pust' A2* 'The speaker is sure that the situation A1 is taking place or is going to take place even though the situation A2 occurs; the speaker considers that if A2 occurs, then a situation opposite to A1 takes place' (*My ne vprave jego osuždat', pust' on i vinovat;* 'We are not to blame him though he is guilty') (V. Apresjan 2010).

The above cited definitions and examples from the dictionary demonstrate the capacity of the particle *pust'* to convey the meanings of optativeness and condition (there is a shade of concession in the latter case).

The word **PUSKAJ** has a dictionary entry similar to **PUST'**.

Thus the presented material shows that the meanings of imperativeness, optativeness, and condition in Modern Russian can be expressed by means of the same lexemes (or *by* with the lexemes in question).

In this respect it may be of great interest to examine the texts in Old Russian and possibly draw a typological parallel between the expression of meanings mentioned above in Old and Contemporary Russian. The most common old counterpart of constructions with *pust'* and *puskaj* is the combination of the particle да (*da*) with verbs in the present (simple future) tense (the so called да + *praes.* construction). This construction is partly preserved in the modern language, namely in the stylistic usage of the following type: *Da budet svet!* 'May there be light!' *Da zdravstvajet svoboda!* 'Long live liberty!' *Da pokoitsa prakh jego s mirom!* 'May his soul rest in peace!'*[1]*

The main function of the да + *praes.* construction was to indicate imperativeness. Consider the following examples:

(1)

бг҃ъ	да	оүщедрить	ны	и	да	възвратить
God	let/may	lend grace	we	and	let/may	return
N.sg.	particle	praes.3sg.ind.	Acc.	conj.	particle	praes.3sg.ind.
градъ	нашь					

---

[1]   Compare a more detailed definition of the modern *da*: "The lexems **pust' 1** (*May the sun always shine!*), **da 9** (*May there be light!*) and **č'tob 4** <…> are getting closer to the synonyms **tol'ko by**, **liš by** and **chot' by** in the sphere of their uses where they mean wish. Their common feature in this case is certain performativeness: saying them, the speaker is trying to get P <…>. The contexts typical of lexeme **da 9** are those where P is a creator; compare: *May God be with you<…>*. In the case of **da 9** the emergence of P is supported by the very act of speech and thereby depends on the speaker. Compare: <…> *Let it be so! – said she, – I accept you decision as God's mercy* (I.A.Goncharov. The Precipice)" (Apresjan Ju.D. et al., 2004).

city	our
Acc.sg.	Acc.sg.m.

(Uspenskiy Sbornik of XII-XIII cc. The Tale by Prophet Jeremiah about the Capture of Jerusalem; 1в); 'May God lend us His Grace and give us our city back'.

We must note that the да + *praes.* construction is traditionally associated with standard language and is considered borrowed into Russian documents from Old Church Slavonic.

But imperativeness might not have been the only meaning initially attached to the construction in question. In addition, literary works may not have been its only scope of application. Let us start with two extracts from Mstislavova gramota, an Old Russian document of business correspondence (hence a nonliterary document), which was written around 1130, most likely in Kiev (or in Novgorod). This document is a letters patent saying that the Great Prince Mstislav and his son, Prince Vsevolod of Novgorod, pass Bujtsy village into the ownership of Yurjev Monastery along with an autumnal "poliudie" (a kind of a toll) amounting to 25 grivnas, and a silver plate.

The text contains two fragments of identical structure (the so called sanctions) which warn whoever would dare to break the conditions in the gramota. Here are these two fragments:

(2)

да	же	которыи	кнаѕь	по	моѥмь
if	any	prince	after	my
conj (?) / particle + particle (?)		N.sg.	N.sg.	prep.	Loc.sg.m.

кнажени҇и	почьнеть	хотѣти	ѿꙗти	оу
rule	begin	wish	take away	from
Loc.sg.	praes.3sg.ind.	inf.	inf.	prep.

сто҇го	геѡргиꙗ	а	бго҇ъ	боуди	ѕа	тѣмь
Saint	George	then	God	be	for	this
G.sg.	G.sg.	conj.	N.sg.	3sg.imper.	prep.	Loc.sg.n.

҅и	ста҇ꙗ	бц҇а	и	тъ	стꙑи	ге҅ѡрги҅и
and	Saint	Mother of God	and	that	Saint	George
conj.	N.sg.f.	N.sg.	conj.	N.sg.	N.sg.m.	N.sg.

оу	него	то	ѿимаѥть
from	whom	take away
prep.	G.sg.	particle	praes.3sg.ind.

'If any prince after my rule wants to take away [the granted], then be there (i.e. 'punish for it') God, and Holy Mother of God, and Saint George from whom the prince takes away'. (Yurjev Monastery was devoted to Saint George).

(3

да	же	кто	ѕапъртит	или	тоу	дань
if	anybody	spoil	or	that	toll
conj(?)/particle + particle (?)		N.sg.	praes.3	conj.	Acc.sg.f.	Acc.sg.

и	се	блюдо	да	соудить	кмоу	бъ
and	this	plate	let	judge	he	God
conj.	Acc.sg.n.	Acc.sg.	particle	praes.3sg.ind.	Dat.sg.	N.sg.

въ	днь	пришьствиꙗ	свокго	и	тъ	стыи
on	day	advent	his	and	that	Saint
prep.	Acc.sg.	G.sg.	G.sg.n.	conj.	N.sg.m.	N.sg.m.

[г](є'ѡ)ргии
George
N.sg.

'If anybody breaks the regulation connected with that (mentioned above) toll or with this plate, may God and that Saint George judge him on the Day of His Advent'.

There is no doubt that the first part of both statements points out the very condition under which God's wrath is to fall upon the breaker. What linguistic material expresses this idea in these statements? It is traditionally assumed that the condition is introduced by a lexeme **даже** (*dazhe*) which historical dictionaries interpret similarly to "*if*" (Slovar XI – XIV, 1989; Sreznevskij I.I., 1958).

However, it should be noted that there was a considerable number of conditional conjunctions in Old Russian, which were partially different with respect to the type of text in which they were used: – **аще** (a conditional in standard Old Russian), **аче**, **аже**, **оже** (spoken Old Russian), etc. (Lavrov, 1941). In this respect the occurrence of one more conjunction possessing the same semantic meaning should be specially clarified.

The following solution can be introduced in this case: conditionality is likely to have been conveyed by the **да** + *praes*. construction in such contexts, and the particle **же** was inserted into the construction secondarily[2]. Let us try to prove this assumption.

Let us imagine there is no particle **же** in the example (2) – we will get the following:

(2′)	да	который	кнаѕь	по	мокмь	кнажени'и	почьнеть
	let	any	prince	after	my	rule	begin
	particle	N.sg.m.	N.sg.	prep.	Loc.sg.n.	Loc.sg.	praes.3sg.ind.

хотѣти	ѿꙗти
wish	take away
inf.	inf.

Can this phrase, as it is, introduce any condition? Yes, because **да почьнеть хотѣти** corresponds with the Modern Russian *pust' zakhoč'et* (*let them wish*), and this construction, as was stated above, can convey the meaning of conditionality.

Let us compare in more detail Modern Russian language data with the text of Mstislavova Gramota. From the examples we have discussed above we may draw a parallel between the

---

[2] The particle **же** in Old Russian is a first-order enclitic and is therefore placed before all the other enclitics which refer to the predicate and obey Wackernagel's law (Zaliznjak, 2008).

two contexts: the sanctions in Old Russian and the following kind of assumption in Contemporary Russian: *Pust' summa uglov treugol'nika ne ravna 180 gradusam;* 'Let us assume that the sum of the angles in a triangle is not equal to 180 degrees'. Here *pust'* means <u>special</u> 'let us assume'. Similarly, Old Russian examples can be interpreted in such terms, taking stylistic colouring into consideration: the sanctions determine punishments on the basis of the assumption that there will be someone who would go against the will of the author.

But the modern "*pust'* of a warning/threat" is even closer to the Old Russian material under consideration (*Pust' tol'ko poprobuet!* 'Dare he try!' *Pust' tol'ko sunetsa, my jemu pokazhem!* 'Dare he butt in, we will show him!') In fact, the sanctions present a direct threat to a violator.

Thus, the conditions of the Gramota's sanctions may be translated as follows: "Dare (only) some prince after my rule wish to take away [the granted]!" "Dare (only) someone break the regulation!"

However, there is one important difference in terms of syntax: the modern *pust',* being a particle in such contexts, cannot introduce a subordinate predicative unit of a complex sentence. Nevertheless, the semantic similarity of these two constructions is evident.

Let us consider one more similar Old Russian example, namely a piece of scribe Domka's additions on the margins of Novgorod September Menaion (late XI century):

(4)	**да**	**жь**	въ	на	поꙗ	ѡбрꙗще	криво
	if	on	that	sing	find	wrong
	conj.(?)/particle+particle(?)		prep.	Acc.pl.f	part.act.praes.N.sg.m.	praes.3sg.ind	adv.
	а	**вьсе**(…)[3]					
	then	???					
	conj.						

'If [anybody] finds a mistake when celebrating and using them (these books), then…' (f. 176 recto).

In this case the first clause also expresses condition, and both the auxiliary words **да** and **жє** and a verb in the present tense form can be found[4].

In order to prove the assumption that in all the cases conditionality was conveyed (at least initially) without the particle **жє**, solely by means of the **да** + *praes.* construction, it is worth answering the question whether the particle **жє** possesses any integral meaning which this it introduced into this type of contexts.

Let us examine two more Old Russian examples. The first comes from Voproshanie Kyrikovo (VK), a document of the ecclesiastical law, which was created in 1130-1156 by the chronicler and mathematician Kirik of Novgorod. This text is a record of Kirik's conversations with Bishop Niphont of Novgorod and with Metropolitan Kliment Smoljatich and others (Mil'kov

---

[3]   Further the text is broken off.

[4]   The verb in the present tense form presents its dialect variant (the so called "presence form without –ть").

& Simonov, 2011). The text is written in the so-called Hybrid Old Church Slavonic, which represents the features of both standard and spoken language.

(5)

а	намъ	дела	рекше	лихвы	тако	велаше
and	interest	for	that is	interest	so	order
particle	G.pl.	postp.	conj.	G.sg.	adv.	imperf.3sg.

оучити	да	же	не	могоуть	са хабити	то
teach	if		not	able	resist	then
inf.	conj.(?)/particle+particle(?)		particle	praes.3pl.ind.	inf.	conj.

рци	имъ	боудите	мл͡рди	възмѣте	легко
say	they	be	merciful	take	easily
2sg.imper.	Dat.	2pl.imper.	N.pl.m.	2pl.imper.	adv.

(VK, Basic Version, XV c.List, f.17б-18a; cited: (Mil'kov & Simonov, 2011); 'And about interests, that is about likhva, (Niphont) ordered to instruct in this way: <…> If they cannot resist, then tell them: "Be merciful! Take little".'

The second example is taken from the Novgorod First Chronicle, the Synod Copy. It is also written in Hybrid Old Church Slavonic:

(6)

твьрдиславъ	же	позра	на	ст͞оую	софию	и
Tverdislav	≈ and	see	on	Saint	Sophia	and
N.sg.	particle	part.praes.act.N.s	prep.	Acc.sg.f.	Acc.sg.	conj.

рече	да	же	боудy	виноватъ	да	боудоу
say	if		be	guilty	let	be
aor.3sg.	conj.(?)/particle+particle(?)		praes.1sg.ind.	N.sg.m.	particle	praes.1sg.ind.

тоу	мертвъ
here	dead
adv.	N.sg.m.

(NFC, Syn.Copy, f. 90 verso, 1218); 'Tverdislav said before Saint Sophia: "If I happen to be guilty, may I die here".'

It seems that all the Old Russian examples drawn here have one semantic feature in common: all conditions suggest some taboo or a rule being broken. In fact the sanctions of Mstislavova Gramota are addressed to a potential violator of the Prince's will; Domka assumes that he could have made mistakes while rewriting the holy scriptures; a violation of an ethic norm is suggested in the context of Voproshanie Kyrikovo. Practically the same phenomenon is observed in the piece from the Novgorod First Chronicle: from the speaker's point of view, his guilt could have broken the routine course of events.

Thus we may assume with more confidence that the particle же occurring in the constructions of this type carried an adversative meaning. Же used to be independent from the да + *praes.* construction which introduced a condition. This particle merged with the construction later. Before it did, the combination да + же + *praes.* was probably closest in meaning to the modern 'but if P'.

There is one more indirect proof of our hypothesis: in the absolute majority of cases the construction is used with a verb in the present tense. At the same time, other conditional conjunctions collocate easily with other verb tenses. For example:

(7)

аче	кто	оу	половець	оутечашеть	оу	городъ
if	somebody	from	Polovtsian	run away	into	town
conj.	N.sg.	prep.	G.pl.	imperf.3sg.	prep.	Acc.sg.

а	тѣхъ	не	выдавашеть
then	those	not	return back
conj.	Acc.pl.	particle	imperf.3sg.

(Hypatian Codex, 1154); 'In the case when somebody from the Polovtsians runs away to town, [he] never gave them up'.

Having examined the specific usage features of the auxiliary word combination, we should draw a parallel between it and the combination да + ти. Да + ти formed an Old Russian (namely in the Novgorod dialect) conjunction introducing purpose clauses дати (дать) (*dati / dat'*).

The description of this dialect was produced by A. Zalyznyak on the material of birch-bark documents.

Consider the following type of sentences:

(8)

ажето	лодиа	присълана	кыанина	обѣсти	ж
if	boat	send	of the Kievan	tell	it
conj.	N.sg.	part.past,pass.N.sg.f	N.sg.f.	2sg.imperat.	Acc.sg.f.

къназоу	дати	не	боуде	присловъа	ни
prince	in order	not	be	bad name	nor
Dat.sg.	conj.	particle	praes.3sg.ind.	G.sg.	particle

тобѣ	ни	павълови
you	nor	Pavel
Dat.	particle	Dat.

(birch-bark doc. № 745, late XI –first quarter XII c); 'If the Kiev's man boat has [already] been sent, then tell the prince about it, least neither you, nor Pavel were talked about badly' (translated by A.Zaliznjak).

According to A.Zaliznjak, "this conjunction in Old Novgorod dialect emerged as a result of blending of да 'let' with the particle ти"[5]. Back in the XII century other words could still be inserted between these components.

---

[5]   The particle ти in Old Russian is a fourth-order enclitic (Zaliznjak, 2008).

 "The conjunction **дати** (**дать**), – writes A. A. Zaliznjak, – is normally combined with present tense" (as it does in birch-bark doc.745). In (Zaliznjak, 2004) this conjunction is connected directly with the construction **да** + *praes*.

Note that **да** + *praes*. is used in a different way compared to the examples with the particle **же** discussed earlier in the article: the construction expresses purpose (close to optativeness) rather than condition.

Let us consider one more example with the conjunction **дати**. This fragment is taken from the already cited VK (so called Special Version of the XVI century copy).

(9)	**аже**	**са**	**болєнъ**	**покаетⷮ**	**велаше**	**безъ**	**ризⷷ**
	if		sick	confess	order	without	chasuble
	conj.	refl.particle	N.sg.m.	praes.3sg.ind.	imperf.3sg.	prep.	G.pl.
	покаати	**его**	**аще**	**не**	**может**	**до**	**цркⷯви**
	confess	he	if	not	can	to	church
	inf.	Acc.sg.	conj.	particle	praes.3sg.ind.	prep.	G.sg.
	доити	**аще**	**речⷷ**	**жена**	**боудеть**	**то**	**ѡблещи са**
	go	if	say	woman	be	then	dress
	inf.	conj.	aor.3sg.	N.sg.	praes.3sg.ind.	conj.	inf.
	в	**ризы** <…>	**велико**	**бо**	**речⷷ**	**покаание**	**дати**
	in	chasuble	great	for	say	repentance	in order
	prep.	Acc.pl.	N.sg.n.	particle	aor.3sg.	N.sg.	conj.
	не	**просто**		**мⷩнатъ**	**дроузии**		
	not	humble		think	some people		
	particle	N.sg.n.		praes.3pl.ind.	N.pl.m.		

(Mil'kov & Simonov, 2011); 'If the sick repents, (the Lord) ordered to hear his confession without a chasuble if he cannot reach the church. If only, – said (he), – a woman begins to confess, then you must be dressed in the chasuble. For great is, – said he, – repentance, [one should be dressed in] in order some people don't think it humble (that is rustic)'.

This translation (close to the translation in (Mil'kov & Simonov, 2011) suggests that we consider the combination **дати** an integral conjunction. However, taking into consideration that the text was created in the XII century, when **да** and **ти** hadn't blended yet into one conjunction, we can say that the "pure" construction **да** + *praes*. is expressing an imperative meaning, whereas **ти** performs an intensifying function (which was typical for this particle). In this case the translation must run roughly as follows: "For great, – said he, – is repentance. Dare not other people (that is some others) think it humble".

Talking about the imperative meaning of **да** + *praes*. let us turn to the text of Mstislavova Gramota once again and examine the closing part of the second sanction (3):

(3′) **да соудить кмоу бъ въ днь пришьствиꙗ своего и тъ стыи [г](еꙿѡ)ргии.**

In this case the **да** + *praes.* construction is used per se, without any other particles. How should this construction be interpreted? Mstislavova Gramota belongs to the genre of business correspondence that reflects the spoken language and hence does not represent the features of standard Old Russian. Therefore, the chance that this construction came from the literary sphere is quite small (nevertheless, the existence of a certain set form (e.g. **се азъ**; *se azъ*; 'here am I') in standard language should be taken into account).

We can assume that we deal with **да** introducing the main clause, whereas the subordinate is conditional. We may treat the modern *to* and *tak*, the optional parts of the conjunction which introduce purpose clauses, as its counterparts. Such type of uses can be found, for instance, in birch-bark documents.

(10)

не	хоцеши	ли	да	а	боле	не
not	want	if	then	I	more	not
particle	praes.2sg.ind.	conj.	particle	N.sg.	adv.compar.	particle

могоу	водат(и)
can	give
praes.2sg.ind.	inf.

(Tver birch-bark doc № 1, late XII c – first quarter XIII c); 'If you don't want (to take a half of the money), then I cannot give more [anyway]'.

If the above translation is correct, then the corresponding piece from Mstislavova Gramota should be interpreted the following way: 'If somebody spoils that (stated above) toll or this plate, then God and that Saint George will judge him on the day of His Advent'.

Such interpretation is certainly quite possible, but the idea that we encounter the **да** + *praes* construction in its optative meaning again seems more credible. All the above speculations about the origin of the conjunctions **даже** and **дати** as well as the idea about the symmetry of these two sanctions support this conclusion. Let us recall that in the example (2) the 3rd person imperative form is used, but toward God this form should be interpreted as optative:

(2″) а бг҃ъ боуди за тѣмь и ст҃а ҃га бц҃а и тъ ст҃ыи ге҃оргии.

Therefore, it is quite natural to expect optative modality in the second sanction as well.

To conclude, the examined material enables us to draw a rather definite syntactical parallel between Old Russian and Modern Russian: the modern particles *pust' (puskaj)* and the Old Russian particle **да** combined with a verb in the present (simple future) tense express not only imperativeness/optativeness, but condition as well.

It is worth mentioning that the imperative mood (known to express condition in Modern Russian) originates in the Indo-European optative mood. "None of the Slavonic forms descend from the Indo-European imperative mood. The Slavonic imperative mood descends from The Indo-European optative mood. Nothing of this kind could be found in other languages but Slavonic. The Indo-European optative mood <…> means possibility and wish" (Meillet, 2001).

On a typological parallel in the expression of 'imperativeness', 'optativeness' and 'condition' in Modern Russian and Old Russian

This fact may indicate that imperativeness, optativeness and condition used to comprise an integral semantic complex which could be expressed by syntactical constructions and grammatical forms.

Bibliography

Apresjan Ju. D. et al. 2004. *The New Explanatory Dictionary of Synonyms of Russian. Second edition.* Moscow-Vienna: Jazyki slavjanskoj kultury. Wiener Slawistischer Almanach. Sonderband 60. 1162.

Apresjan V. Slovarnye stat'i polej 'sootvetstvie i nesootvetstvie dejstvitel'nosti', 'maloe količestvo i stepen', 'ustupitel'nost' i 'organizacii', 2010 // *Prospekt aktivnogo slovaria russkogo iazyka* [Prospect of the Active dictionary of Russian]. Moscow: Jazyki slavjanskix kultur. 203-206.

Lavrov B. V., 1941. *Uslovnye I ustupitel'nye predlozhenija v drevnerusskom jazyke.* Moscow: Izdatel'stvo AN SSSR. 39-115.

Meillet A., 2001. *Obshcheslavjanskij jazyk.* Moscow: Progress. 264.

Mil'kov V. V., Simonov R. A., 2011. *Kirik Novgorodec: uchenyj I myslitel'*. Moscow: Krugъ. 351.

Russkaya grammatika, 1980. Vol.2: 105-115. Moscow: Nauka.

Slovar drevnerusskogo jazyka (XI – XIV cc), 1989. Moscow: Russkij jazyk. 422.

Sreznevskij I. I. 1958. *Materialy dlja slovarja drevnerusskogo jazyka*. Moscow: Gosudarstvennoe izdatelstvo inostrannyx I nacionalnyx slovarej. 624.

Uspenskij sbornik XII-XIII cc, 1971. Moscow: Nauka.

Zaliznjak A., 2004. *Drevnenovgorodskij dialect.* Moscow: Jazyki slavjanskoj kul'tury. 198-200.

Zaliznjak A., 2008. *Drevnerusskie enklitiki*. Moscow : Jazyki slavjanskix kultur. 28-30, 32.

On the Idiosyncratic Nature of the Semantic Metalanguage of the Active Dictionary of Russian. A View from the German Perspective

Tilmann Reuther

Alpen-Adria-Universität Klagenfurt
Austria
tilmann.reuther@uni-klu.ac.at

Abstract

The semantic metalanguage of the Active Dictionary of Russian is a sublanguage of Modern Standard Russian. Designed to be user-friendly, it carries in its core several idiosyncratic characteristics of Russian which are non-isomorphic to their functional equivalents in other natural languages. The present paper provides an overview of crucial syntactic and lexical problems arising from the search for a German-based equivalent of the Russian-based semantic metalanguage of the Active Dictionary of Russian.

Keywords

Semantic metalanguage, natural language, Russian, German, Active Dictionary of Russian, syntactic structures, lexical elements

1 Introduction

1.1 Towards the Active Dictionary of Russian

In 2010, the core group of the Moscow Semantic School headed by Jurij D. Apresjan published the Prospectus of a new kind of dictionary – the Active dictionary of Russian (Prospekt, 2010). At that time the group already had a rich experience in lexicography due to more than a decade of common work on the New Explanatory Dictionary of Synonyms of Russian (NOSS, 2004). The latter, however, was more a collection of in-depth studies into Russian vocabulary than a dictionary for common use. At present, the Moscow group is endeavouring to profit from the outcomes of previous theoretical and practical work, in order to create a user-friendly, alphabetically-ordered dictionary which will be built on the principles of systemic lexicography (cf. Apresjan 2006), and will meet the needs of a broader public.

In the Active dictionary, every lexeme, which is the basic unit of lexicographic description, will be supplied with the following information: name of the lexeme; morphological markers; syntactic markers; stylistic markers; the standard sentential form (if applicable); the explanation (Russ. tolkovanie) of its meaning in a semantic metalanguage; the governance model; a set of collocations; illustrations of usage from the corpora; a list of synonyms, antonyms, converses, analogues and derived lexemes.

1.2 The Semantic Metalanguage – a Natural Sublanguage

In this paper I will concentrate on the explanations, i.e. on the semantic metalanguage used by the Moscow Semantic School. Being a simplified and standardized sublanguage of Modern Standard Russian (Apresjan 2006, p. 54), the semantic metalanguage carries in its core many idiosyncratic characteristics of Russian which are non-isomorphic to their functional equivalents in other natural languages. Within the framework of the Active dictionary, the semantic metalanguage is declared to provide the basis for user-friendly explanations[1], in other words, explanations are expected to be in reasonably good, readable Russian. The goal of this article is to concentrate on idiosyncrasies of the Russian explanations and to present an overview of crucial syntactic and lexical problems arising from the search for German-based equivalents of the Russian explanations.

2 An Example Containing Several Idiosyncrasies

The following example (1) – the explanation of the lexeme **ždat 1.1** (Engl. (to) wait) taken from (Prospekt, 2010, p. 88) – shows three major syntactic idiosyncrasies of Russian:

a)   the preponed converb construction ('znaja ili sčitaja' / lit. knowing or supposing),

b)   the postponed attributive construction ('sobytie A2, nužnoe čeloveku A1 ili kasajuščeesja ego' / lit.[2] event A2, necessary for person A1 or concerning him), and

c)   the inverse order of subject and predicate ('čto v meste A3 dolžno ili možet proizojti sobytie A2' / lit. that at place A3 must or can happen event A2).

In English, constructions a) and b) are possible, but c) is not. The complete Russian explanation of **ždat 1.1** and its English word-by-word gloss are the following:

(1) **ždat 1.1** (Engl. to wait)
ZNAČENIE. *A1 ždet A2 v A3 v tečenie A4* 'Znaja ili sčitaja, čto v meste A3 dolžno ili možet proizojti sobytie A2, nužnoe čeloveku A1 ili kasajuščeesja ego, A1 v moment ili otrezok vremeni A4 naxoditsja v sostojanii gotovnosti k nemu, vozmožno, naxodjas' v meste A3 i želaja, čtoby ono proizošlo'.

---

[1]   "… dobavljaetsja trebovanie k druželjubnosti tolkovanij. Oni dolžny byt' obščeponjatnymi .., čtoby po tekstu tolkovanija mogla ugadyvat'sja sootvetstvujuščaja leksema" (… the request for user-friendliness of the explanations is added. They must be generally understandable …, so that the corresponding lexeme can be guessed from the text of the explanation. – Translation T.R.) (Prospekt, 2010, p. 88)

[2]   Literal word-by-word glosses are marked as lit. and come without semantic brackets. Nouns are used without definite/indefinite article.

(2) **wait 1.1** *A1 waits for A2 at A3 during A4* <u>lit.</u> Knowing or supposing that at place A3 must or can occur event A2 necessary for person A1 or concerning him, A1 at moment or period of time A4 is in state of readiness for it, possibly being at place A3 and wishing that it occurs.

When turning to German, the syntactic discrepancies become even bigger. While the preponed converb construction a) is possible for German, though stylistically marked, the two constructions b) and c) from above are impossible. Additionally, there are two more syntactic complications. In German

d) the position of adverbials cannot be between subject and predicate ('A1 v moment ili otrezok vremeni A4 naxoditsja v sostojanii gotovnosti' / <u>lit.</u> A1 im Moment oder Zeitabschnitt A4 befindet sich im Zustand der Bereitschaft / <u>recte</u> A1 befindet sich im Moment oder Zeitabschnitt A4 im Zustand der Bereitschaft);

e) the attributive construction cannot be postponed and always has the congruent attribute next to the noun ('sobytie A2, nužnoe čeloveku A1 ili kasajuščeesja ego' / <u>lit.</u> Ereignis A2, nötig für Mensch A1 oder betreffend ihn / <u>recte</u> für den Menschen A1 nötiges oder ihn betreffendes Ereignis).

As a result, the German word-by-word gloss (3) of example (1) is syntactically completely unacceptable:

(3) **warten 1.1** *A1 wartet auf A2 in A3 im Verlauf von A4* <u>lit.</u> Wissend oder meinend, dass an Ort A3 muss oder kann Ereignis A2, nötig für Mensch A1 oder betreffend ihn, sich ereignen, A1 zum Zeitpunkt oder im Zeitabschnitt A4 befindet sich im Zustand der Bereitschaft zu ihm, möglicherweise sich an Ort A3 befindend und wünschend, dass es stattfindet.

Of course, the syntactic deficiencies can be eliminated. A syntactically acceptable German version of explanation (1) is the following:

(4) **warten 1.1** *A1 wartet auf A2 in A3 im Verlauf von A4* 'Wissend oder meinend, dass sich am Ort A3 ein Ereignis A2, welches für den Menschen A1 nötig ist oder ihn betrifft, ereignen muss oder kann, befindet sich A1 zum Zeitpunkt oder im Zeitabschnitt A4 im Zustand der Bereitschaft zu ihm, möglicherweise sich am Ort A3 befindend und wünschend, dass es sich ereignet'.

However, the German explanation (4) still has two lexical deficiencies. The wordings 'ein für den Menschen nötiges Ereignis' (cf. 'sobytie, nužnoe čeloveku') and 'es muss oder kann sich ein Ereignis ereignen' (cf. 'dolžno ili možet proizojti sobytie') are not acceptable from an idiomatic point of view. Possible solutions would be the following: 'ein für den Menschen unerlässliches Ereignis' and 'es muss oder kann ein Ereignis eintreten'.

We will now continue with a closer look at the lexicon of semantic metalanguages, and will afterwards return to syntactical questions.

3 Russian vs German Lexicons for Semantic Metalanguages

3.1 The Main Parts of the Lexicon of a Semantic Metalanguage

In his survey of the lexicon of the semantic metalanguage, Apresjan (2006, p. 52–55) presents three types of sememes which make up the core of the metalanguage:[3]

a) semantic primitives, i.e. meanings *delat'* ('do' / 'machen'), *xotet'* ('want' / 'wollen'), *moč'* ('can' / 'können'), *znat'* ('know' / 'wissen'; 'kennen'), *sčitat'* ('suppose' / 'meinen'), *suščestvovat'* ('exist' / 'existieren'), *naxodit'sja'* ('be in a place' / 'sich befinden'), *vremja* ('time' / 'Zeit'), *uslovie* ('condition' / 'Bedingung'), *situacija* ('situation' / 'Situation'), *xorošij* ('good' / 'gut'), *bol'še* ('more' / 'mehr'), *ne* ('not' / 'nicht') and some others;

b) meanings which are more complex than semantic primitives but can easily be decomposed into semantic primitives, e.g. *gotov* ('ready' / 'bereit'), *dolžen* ('must' / 'müssen'), *načat'sja'* ('begin' / 'beginnen'), *cel'* ('goal' / 'Ziel'), *moment* ('moment' / 'Zeitpunkt') and some others;

c) many intermediate meanings which are part of the explanations of whole groups of lexical units, e.g. *trebovat'* ('urge' / 'verlangen').

These sememes are all words of Modern Standard Russian (English, German) taken in one of their meanings. They are simplified and standardized in exactly this sense.

3.2 Equivalence of Sememes in Different Semantic Metalanguages

A study of Russian sememes and their possible equivalents in other languages raises two sorts of problems. For a given pair of sememes we have to ask the following two questions: 1) Is there a real equivalence of meaning? and 2) Is there an equivalence of combinatorial properties of the sememes?[4]

3.2.1 Equivalence of Meaning

On the one hand, in many cases the three languages Russian, German and English offer perfect equivalents – see above, the triples in a), b) and c).[5]

On the other hand, in the case of non-equivalence of the meaning of sememes (in two languages A and B), there are two possible constellations: (i) 2 sememes in language A yield 1 sememe in language B, or, (ii) 1 sememe in language A yields 2 sememes in language B.

---

[3] The Russian sememes are taken from (Apresjan, 2006); the English and German equivalents are mine.

[4] In other words, a given Russian sememe and its counterparts in other languages must be examined along the lines that were established for the study of synonyms in NOSS 2004.

[5] For a detailed study of the set of semantic primitives proposed by Anna Wierzbicka in the 1990's and their equivalents in three languages (English – Russian – German) cf. Reuther (2003).

Case (i) is represented by the equivalence of Russ. 'dobro' / 'xorošee' and Germ. 'gut' (Engl. 'good'), cf. the following example:

(5) **blagodarnyj 1.1** (Engl. grateful)
ZNAČENIE. *Blagodarnyj čeloveku A2 za A3* 'Ispytyvajuščij prijatnoe čuvstvo, vyzvannoe tem, čto čelovek A1 priznaet, čto čelovek A2 sdelal emu **dobro** A3, i gotov vyrazit' ėto čuvstvo slovami ili sdelat' dlja A2 čto-to **xorošee** '. (Prospekt, 2010, p. 91)

(5') **dankbarer 1.1.** *Dankbarer dem Menschen A2 für A3* 'Ein solcher[6], der ein angenehmes Gefühl verspürt, welches dadurch hervorgerufen wird, dass der Mensch A1 anerkennt, dass der Mensch A2 ihm **Gutes** getan hat, und bereit ist, dieses Gefühl mit Worten auszudrücken oder für A2 etwas **Gutes** zu tun'.

Another, almost trivial example are the suppletive Russian sememes 'čelovek' (sg.) – 'ljudi' (pl.) and the regular sg – pl Germ. 'Mensch' – 'Menschen'. (Engl 'man' – 'people', 'person – persons'). Cf. the following explanation:

(6) **važnyj 1** (Engl. important)
ZNAČENIE. *Važnyj dlja A2* 'Takoj, kotoryj dolžen učityvat'sja v dejstvijax **čeloveka ili ljudej** A2 ili dejstvijax A2 dlja togo, čtoby oni byli uspešnymi'. (Prospekt, 2010, p. 91)

(6') **wichtiger 1** *Wichtiger für A2* 'Ein solcher, der bei den Handlungen **des oder der Menschen** A2 oder den Handlungen A2 berücksichtigt werden muss, damit sie erfolgreich sind'.

Case (ii) is represented by the relation of the type a) semantic primitive 'znat'' and the two German primitives 'wissen' (to know something) and 'kennen' (to know something/somebody). However, case (ii) is much more typical for sememes of group c), i.e. for intermediate sememes. As an example we can take Russ. 'zvuk' with its equivalents Germ. 'Laut' (sound of a human or animal voice) and 'Geräusch' (sound of other origin), cf. the following explanations:

(7) **bezmolvie 1** (Engl. ?)
ZNAČENIE. 'Polnoe otsutstvie **zvukov** čelovečeskoj reči, imejuščee mesto dolgoe vremja; ... ' (Prospekt, 2010, p. 638) lit 'total absence of sounds of human speech lasting for a long time ... '.

(7') 'Völliges Fehlen von **Lauten** der menschlichen Rede, welches lange Zeit besteht; …'

(8) **zatiš'e 1** (Engl. lull?)
ZNAČENIE. 'Kratkovremennoe otsutstvie gromkix **zvukov**, proizvodimyx prirodnymi stixijami, ... ' (Prospekt, 2010, p. 639) lit 'short absence of loud sounds produced by nature ... '.

(8') 'Kurzes Fehlen von lauten **Geräuschen**, die von Naturkräften erzeugt werden, …'

---

[6] This German wording ('ein solcher, der … ein Gefühl verspürt') is structurally different from the Russian ('ispytyvajuscij cuvstvo'): a head and a subordinated relative clause vs. a noun and an attributive participle.

(9) **tišina** (Engl. silence)
ZNAČENIE. 'Otsutstvie **zvukov**'. (Prospekt, 2010, p. 645) <u>lit</u> 'absence of sounds'.

(9') 'Fehlen von **Lauten und Geräuschen'**

The German explanation (9') must contain both German intermediate sememes 'Laut' (sound of a human or animal voice) and 'Geräusch' (sound of other origin).

3.2.2 Equivalence of Combinatorics

The search for a German-based equivalent of Russian semantic metalanguage of the Active Dictionary necessarily draws attention to the question of how to combine sememes at the level of the lexicon. What we observe is the existence of collocations on a very basic level.

Let us look at the following explanations, and especially, at the verbal collocates of the nouns: 'situacija' – 'imeet mesto'; 'sobytie' – 'proisxodit'; 'osnovanija' – 'est''; 'ob"ekt', 'javlenie' – 'suščestvuet'. The possible German equivalents are 'Situation', 'Grundlage' – 'besteht' (Engl. 'situation', 'basis' – 'exists'); 'Ereignis' – 'tritt ein' (Engl. 'event' – 'occurs'); 'Objekt', 'Erscheinung' – 'existiert' (Engl. 'object', 'phenomenon' – 'exists'. This leads to the following German explanations given next to the Russian originals.

(10) **vse-taki 2** (Engl. nevertheless)
ZNAČENIE. *A1, A2, (i) vse-taki A3* '**Imeet mesto situacija** A1; **imeet mesto situacija** A2; govorjaščij sčitaet, čto esli **imeet mesto situacija** tipa A2, to obyčno ili estestvenno, čtoby **imela mesto situacija** tipa ne-A1; govorjaščij sčitaet, čto v dannom slučae **imeet mesto situacija** A1, potomu čto **imeet mesto situacija** A3'. (Prospekt, 2010, p. 92)

(10') **dennoch 2** *A1, A2, und dennoch A3* 'Es **besteht die Situation** A1; es **besteht die Situation** A2; der Sprecher meint, dass wenn **eine Situation** vom Typ A2 **besteht**, es dann üblich oder natürlich ist, dass **eine Situation** vom Typ Nicht-A1 **besteht**; der Sprecher meint, dass im vorliegenden Fall **die Situation** A1 **besteht**, weil **die Situation** A3 **besteht'**.

(11) **nadejat'sja 1** (Engl. to hope)
ZNAČENIE. *A1 nadeetsja, čto A2* 'Čelovek A1 ožidaet, čto **proizojdet** xorošee ili nužnoe dlja nego **sobytie** A2, i sčitaet, čto dlja takogo ožidanija **est' osnovanija**' [po analogii – o životnyx].

(11') **hoffen 1** *A1 hofft, dass A2* 'Der Mensch A1 erwartet, dass ein gutes oder für ihn unerlässliches **Ereignis** A2 **eintritt**, und meint, dass für eine solche Erwartung eine **Grundlage besteht'**.

(12) **tol'ko 1.1** (Engl. only)
ZNAČENIE. *Tol'ko A3 iz A2* 'V klasse ob"ektov ili javlenij A2, k kotorym prinadležit ob"ekt ili javlenie A3, ne **suščestvuet ob"ekta** ili **javlenija**, otličnogo ot A3, kotoroe možno oxarakterizovat' kak A1' [*Mal'čik s"el* (A1) *tol'ko persik* (A3) [*iz čisla vsex podannyx k stolu fruktov* (A2)]]. (Prospekt, 2010, p. 92)

(12') **nur 1.1** *Nur A3 von A2* 'In der Klasse der Objekte oder Erscheinungen A2, zu denen das Objekt oder die Erscheinung A3 gehört, **existiert** kein **Objekt** und keine **Erscheinung**, das <die> von A3 verschieden ist und das <die> man als A1 charakterisieren kann'.[7]

It is clear that the above collocations of nouns and verbs are of the FUNC0 type, i.e. they are collocations of nouns with so called light existential verbs and represent collocations of the type 'X suščestvuet' ('X exists'), with 'suščestvovat' ('exist') being a semantic primitive.

4   Russian vs German Syntax for Semantic Metalanguages

The last two examples (12) and (12') show syntactic problems of (i) scope and (ii) anaphora.

(i) In Russian, the scope of the negation '**ne** (suščestvuet X **ili** Y)' in (12) is different from the scope of the negative determiner '**kein**' in German in (12'), and therefore we prefer to say 'es existiert **kein** X **und kein** Y' (instead of 'es existiert **kein** X **oder kein** Y'). (Engl. 'does not (exist X or Y)'; 'exists no X and no Y' instead of 'exists no X or no Y'.)

(ii) In Russian, the gender of the Genitive Singular form of the adjective 'otličnogo' ('excellent') in (12) can be either masculine or neuter, and therefore the anaphoric link works with both 'ob"ekt' ('object', masculine), and 'javlenije' ('phenomenon', neuter); on the other hand, the relative pronoun 'kotoroe' ('which', neuter) in (12) works explicitly only with 'javlenije' (neuter), but not with 'ob"ekt' (masculine); however, in the case of different gender, Russian allows for the syntactic reference of the relative pronoun to the positionally closer noun only. For German, the nouns also have different gender, 'Objekt' (neuter) and 'Erscheinung' (feminine), but reference must be in both forms of the relative pronoun, i.e. 'das' (neuter), and 'die' (feminine), cf. (12') above.

Apart from such minor idiosyncrasies, the example of **ždat' 1.1 (**Engl. to wait) in Section 2 above (cf. examples (1) – (3)) pointed to major differences between Russian and German syntax. German word order was one of the differences, and the Russian converb construction was the other.

In the Active dictionary, the converb construction is used to mark a specific part of the explanation – the presupposition, the general rule being the following (Prospekt, 2010, p. 88): "Therefore, e.g., presuppositions, assertions, modal frames, frames of the observer, and other logically different layers of the meaning are built into the explanation in the form of a continuous text. In particular, the presupposition is usually included into the explanation in the form of a preponed converb clause" (Translation T.R.). In Russian, converb clauses are quite frequent, especially in written texts, which is not the case for German – converb clauses do not make good, readable German explanations. Consequently, equivalent means should be used whenever possible. One of them is nominalisation, another is the modificative clause (*wobei*-clause). This leads to the following alternative for the explanation of **warten 1.1**. (cf. explanation (4) above with converb constructions):

---

[7]   The German wording is structurally different from the Russian; the two separate attributive clauses are connected in the form of an adjunctive construction.

(13) **warten 1.1** *A1 wartet auf A2 in A3 im Verlauf von A4* 'Im **Wissen** oder in der **Meinung**, dass sich am Ort A3 ein Ereignis A2, welches für den Menschen A1 unerlässlich ist oder ihn betrifft, ereignen muss oder kann, befindet sich A1 zum Zeitpunkt oder im Zeitabschnitt A4 im Zustand der Bereitschaft für A2, **wobei** sich A1 möglicherweise am Ort A3 befindet und wünscht, dass A2 sich ereignet'.

One last observation is the following: It is a trivial fact that explanations with more than two actants (A1, A2, A3, A4) become more complex and less user-friendly. This might be the reason why popular German explanations, e.g. in the Duden-Online Dictionary (cf. http://www.duden.de/woerterbuch, checked June 17, 2013) come in a simplistic syntactic form without using any symbols, and even without naming the first (subject) agent, cf.

(14) WARTEN (Engl. to wait); cf. (13) above
1.a. dem Eintreffen einer Person, einer Sache, eines Ereignisses entgegensehen, wobei einem oft die Zeit besonders langsam zu vergehen scheint
1.b. sich, auf jemanden, etwas wartend (1a), an einem Ort aufhalten und diesen nicht verlassen

(15) HOFFEN (Engl. to hope); cf. (11') above
a. zuversichtlich erwarten; wünschen und damit rechnen, dass etwas eintreten oder der Wirklichkeit entsprechen wird

5 Conclusion

My conclusion is that the semantic metalanguage of the Active dictionary of Russian makes use of many idiosyncratic characteristics of Modern Standard Russian, and that explanations must be regarded as regular Russian texts constructed in a relatively schematic way. This allows us to define crucial lexical and syntactic problems when it comes to the question whether the Russian prototype can serve as a model to construct an equally "strong" German semantic metalanguage and similar explicit explanations.

In the lexicon of the semantic metalanguage semantic primitives present only a few problems: 'znat'' ('know') and 'wissen'/'kennen', 'dobro'/'xorošo' ('good') and 'gut' were rare examples of one-to-many relations on the level of semantic primitives. More complex sememes will certainly provide similar or more complex problems, cf. 'zvuk' – 'Laut'/'Geräusch' ('sound'/'noise') for a one-to-many-relation on the level of intermediate sememes.

Collocations are part of the lexicon of every natural language, and the semantic metalanguage of the Active dictionary contains them as well. Collocations make user-friendly and readable explanations. Examples from our material were noun – verb collocations with light existential verbs of the type Russ. 'situacija imeet mesto', 'sobytie proisxodit,' 'ob''ekt suščestvuet' and German 'eine Situation besteht', 'ein Ereignis titt ein', 'ein Objekt existiert'.

Syntactic structures are the second crucial area of the semantic metalanguage. The explanations of the Active dictionary are usually complex structures: we mentioned converb constructions, attributive clauses, negation and word order. The latter seems to be one of the main challenges to construct complex but still readable explanations in a German semantic metalanguage.

Acknowledgements

Bibliography

Apresjan, Ju.D. 2006. Osnovanija sistemnoj leksikografii. In *Jazykovaja kartina mira i sistemnaja leksikografija.* Otvetstvennyj redaktor Ju. D. Apresjan, 33–160. Moskva: Jazyki slavjankix kul'tur.

Duden-Online Dictionary (http://www.duden.de/woerterbuch)

NOSS 2004 = *Novyj objasnitel'nyj slovar' sinonimov russkogo jazyka*. 2004. Moskva & Vena.

Prospekt 2010 = *Prospekt aktivnogo slovarja russkogo jazyka*. 2010. Otvetstvennyj redaktor akademik Ju. D. Apresjan. Moskva: Jazyki slavjankix kul'tur.

Reuther, T. 2003. Semantische Primitiva (Englisch, Russisch, Deutsch). In Kuße, H. (ed.) *Slavistische Linguistik 2001*, 287–308. München: Sagner.

In Search of Semantic Links between Nouns and Light Verbs. Some Evidence from Spanish

Sanromán Vilas, Begoña

University of Helsinki
P.B. 24 (Unioninkatu 40 B 514) FIN-00014 Finland
begona.sanroman@helsinki.fi

Track 4: Collocations and Lexical Functions

Abstract

The paper will focus on the assumption that the selection of light verbs *(dar* 'to give', *hacer* 'to make/do') by nouns *(consejo* 'advice' *confesión* 'confession') within light verb constructions *(dar un consejo* 'to give advice', *hacer una confesión* 'to make a confession'), rather than being arbitrary, is based on semantic grounds. I will show that even the most frequent and semantically empty verbs, like the Spanish *dar* 'to give' and *hacer* 'to make/do', are selected with consideration of semantic components shared by the noun and the verb. Finding the semantic components which distinguish *dar* from *hacer,* both including the meaning 'action', will be a step forward towards the elaboration of more precise definitions for these verbs.

Keywords

Light verbs, light verb constructions, semantically motivated collocations, semantically empty/full verbs, semantic compatibility, Explanatory and Combinatorial Lexicology.

1 Introduction

This paper aims to give some evidence to justify that values of the lexical function (= LF) **Oper$_1$** (Mel'čuk, 1996), also called light verbs (= LVs), rather than being arbitrarily selected by their argument, are based on semantic grounds. To support this claim, the *hypothesis of semantic compatibility* (Sanromán Vilas, 2009, 2012) will be tested. According to this hypothesis, collocate verbs, LVs in this case, are connected to the noun, with which they form a light verb construction (= LVC), and to the related full verb existing in the language by means of semantic links. By semantic link I refer to a semantic component which is repeated in the LV and the noun, and in the LV and the full verb.

In this paper I will focus on *dar* 'to give' and *hacer* 'to make/do', two of the most frequent Spanish LVs (Dubský, 1990: 33, Koike, 2001: 84-87) which have been classified as pure Lvs (Alonso Ramos, 2004), that is, verbs which only have a general and taxonomic meaning. In this case, both *dar* and *hacer* mean 'action'. I will try to prove that the claim so often put forward that there is no semantic explanation as to why, for instance, in Spanish *hacemos* and *damos sugerencias* 'we make / give suggestions', but only *hacemos comentarios* 'we make comments' and *damos nuestra opinión* 'we give our opinion', is not fully accurate, as I will show. In fact, the statement can be turned around to underline that, without denying that languages exhibit many disparities in relation to the selection of LVs, they still show a great amount of coincidence, an aspect which has hitherto been largely disregarded. Just as the Spanish noun *consejo* 'advice' selects *dar* 'to give' to form the LVC *dar un consejo* 'to give advice', the same LV is selected in other Romance languages: Fr. *donner un conseil,* It. *dare un consiglio,* Pg. *dar um conselho,* and non-Romance languages: En. *to give advice,* Gm. *einen Ratschlag geben*, Rus. *davat' sovet*, Bq. *aholku eman,* Fn. *antaa neuvoa.* The same holds for the selection of *hacer* 'to make/do' within the Spanish LVC *hacer una confesión* 'to make a confession' and many of its equivalent LVCs in other languages: Fr. *faire une confession,* It. *fare una confessione,* Pg. *fazer uma confissão,* En. *to make a confession,* Bq. *aitortza egin*, Fn. *tehdä tunnustus,* etc.

The main goal of the paper is to show that, apart from the general meaning 'action', the description of Spanish LVs *dar* and *hacer* must contain other semantic components which make *dar* differ from *hacer,* and vice versa. I claim that within an LVC, the noun and the LV share at least one semantic component. Thus, if we come across the same noun co-occurring with both *dar* and *hacer,* we will expect to find that different elements of the meaning of the noun are emphasized when the selected LV is either *dar* or *hacer.*

This research is mainly carried out within the theoretical and methodological framework of Explanatory and Combinatorial Lexicology (= ECL) (Mel'čuk et al, 1995). Data have been collected from two corpora: *Corpus de referencia del español actual (CREA)* and *Corpus del español (CdE)*, and several Spanish monolingual dictionaries.

The remainder of the paper is organized as follows. Section 2 explains the hypothesis of semantic compatibility. Section 3 describes the methodology to select nouns which form pairs of LVCs sharing the same noun and alternating *dar* and *hacer.* Section 4 concentrates on LVCs with verbal communication nouns. We analyze the semantic differences between LVCs with *dar* and LVCs with *hacer* when the nouns remain the same. Through the different nuances expressed by each LVC of the same pair – e.g. *dar una sugerencia* 'to give a suggestion' / *hacer una sugerencia* 'to make a suggestion', – we will find specific semantic components for *dar* and *hacer,* respectively. The validity of these components is ensured by checking separately that they are also present in the meaning of nouns co-occurring only with *dar* or *hacer.* Section 5 offers some generalizations pointing out some limits and implications of this study, and Section 6 draws some conclusions. Appendix contains three tables with the LVCs (and other collocations) discussed in the paper.

2 The Hypothesis of Semantic Compatibility

It is quite frequent for a noun to combine with several collocate verbs as values of **Oper₁**, among other LFs. When this happens, collocate verbs of the same noun are considered synonyms. The Spanish noun *disgusto* 'upset', for instance, can co-occur with verbs such as

sentir 'to feel', *sufrir* 'to suffer', *tener* 'to have', *pasar* 'to pass', *llevarse* 'to carry', etc. However, collocates are not necessarily exact synonyms, but may differ in many nuances (Mel'čuk 1992: 33-34). Only a glance at *sentir* 'to feel' and *sufrir* 'to suffer' will suffice to see clearly that they do not have precisely the same sense. While *sentir* 'to feel' refers to a wider meaning, such as 'to have/experience an emotion/feeling', *sufrir* 'to suffer' is only used when the emotion/feeling is experienced as unpleasant.

Within the ECL (Mel'čuk, 1996) and the lexicon-grammar framework (Vivès, 1993:10, Gross, 1998:27), it is claimed that values of **Oper$_1$** are semantically (quasi-) empty verbs in the context of their collocations. However, I will emphasize here, in line with the studies carried out by Apresjan & Glovinskaja (2007), Apresjan (2009) and Bosque (2004a, 2004b), that there is "enough" meaning left in these verbs to perceive differences among them when combining with the same noun, and to justify that they are selected by the noun on the basis of its meaning. Specifically, I uphold the hypothesis, named *hypothesis of semantic compatibility* (Sanromán Vilas, 2009, 2012), that LVs are connected to the noun with which they form an LVC and to the related full verb by means of semantic links. By semantic link, I mean a semantic component which is repeated in the LV and the noun, and in the LV and the full verb. Referring back to *disgusto* 'upset', it can be shown, through its abbreviated definition, 'X's temporary, unpleasant emotion caused by the undesirable fact Y', that each collocate verb emphasizes a particular component of *disgusto:*

(1)   a. […] sorprendida más del *disgusto* que *sentía* que afectada por las palabras…
      (CREA)
      'more surprised because of the upset she felt than affected by the words'

      b. Jamás *sufrí un disgusto* que no se disipara con una hora de lectura. *(CREA)*
      'I have never suffered an upset which did not dissipate with an hour's reading'.

      c. Recuerdo *el disgusto* que *tuvo* Jesús cuando vino su hermano a visitarle… *(CREA)*
      'I remember the disappointment Jesús had when his brother came to visit him'.

      d. Ya sabes *el disgusto* que *ha pasado* papá pensando que íbamos a casarnos. *(CREA)*
      'You know how upset dad was thinking we were going to get married'.

      e. Lo peor fue *el disgusto* que *se llevó* su padre al enterarse. *(CREA)*
      'The worst was how upset his father was on learning about it'.

When *sentir* (1a) co-occurs with *disgusto,* the emphasis is on the experiencer of the emotion; if *sufrir* (1b) is selected, the accent is on the unpleasant nature of *disgusto. Tener* (1c) points out that *disgusto* denotes a state focusing on the time during which the emotion is experienced. More specifically, *pasar* and *llevarse* underline the temporariness of the emotion: *pasar* (1d) refers to the moment of the experience, but with limited duration, and *llevarse* (1d) focuses on the time the emotion begins to be experienced.

At the same time, the semantic components emphasized by the LVs correlate with that part of the meaning each LV shares with the correspondent full verb (or heavy verb). However, semantic components can be expressed in an abstract way, either metaphorically or metonymically. For instance, if the basic meaning of *pasar* as a full verb is "to go past something" (2), its meaning as an LV could be formulated as "to experience a transitory state"

(3). In other words, as a full verb, *pasar* (2) is a verb of movement and as an LV (3) becomes a verb of "emotional movement" or verb of change of state:

(2) Los espías se *pasaron* de Este a Oeste… (*CREA*)
'The spies passed from east to west…'

(3) El *disgusto* que *pasaría* él yo no me lo quiero ni figurar. (*CREA*)
'I cannot imagine how upset he was.'

In the following sections, I will focus on the Spanish LVs *dar* 'to give' and *hacer* 'to make/do' and attempt to find the semantic components these LVs share with the nouns within the same LVC.

3 Constitution of a Corpus and Methodology of the Study

Having collected a large group of potential pairs of LVCs alternating *dar* and *hacer* with the same noun as direct object, I have looked for examples extracted from *CREA* and *CdE*. Next, the following steps have been taken. Firstly, I have verified that *dar* and *hacer* are pure LVs, specifically values of **Oper$_1$**. Secondly, for nouns, it was necessary to justify that I was dealing with a single lexical unit in the case of polysemous nouns. And thirdly, I have checked that the opposition *dar*/*hacer* is not associated with diatopic variation. In this process, I have paid attention whether one of the LVCs (either with *dar* or with *hacer*) is significantly more productive than the other.

Concerning the verification of *dar* and *hacer* as values of **Oper$_1$**, some pairs of LVCs were eliminated from my data after noticing that the verbs were values of other LFs. The most productive was the opposition *dar*/*hacer* as causative verbs (**CausFunc$_1$**):

(4) a. Me *da* mucha *gracia* pensar que… (*CdE*)
'I find it very funny to think that'
b. No me *hace gracia* la idea de que… (*CdE*)
'I do not find funny the idea that'

(5) a. Pero no me *da miedo*. (*CREA*)
'But I am not afraid.'
b. No le *hacía miedo* el trabajo. (*CREA*)
'S/he was not afraid of working.'

In both sentences, (4) and (5), *dar* and *hacer* means 'to cause': *hacer* (4b) is the most usual in standard Spanish when it is combined with *gracia*, while *dar* (4a) is preferred in some American varieties of Spanish. With emotion nouns, *dar* (5a) is especially productive (see *DiCE*), but *hacer* (5b) is also found in regional varieties nevertheless.

With regard to the sense of the noun under analysis, some LVCs were also left out after corroborating that the lexical unit combined with one of the LVs was different from the lexical unit combined with the other LV. For instance, *hacer* co-occurs with *tregua1* 'cessation in hostilities' in (6a), while *dar* co-occurs with *tregua2* 'break in an ongoing activity' in (6b).

(6) a. […] tenemos que *hacer una tregua* en varios temas, para que la paz reine en el país… (*CdE*)
'We must declare a truce in several areas to establish peace in the country'
b. […] las criadas *daban tregua* a su quehacer y se acercaban silenciosas, se sentaban… (*CdE*)
'Servants took a break from their tasks and they approached silently, sat down'

I have also excluded from my data some LVCs when the alternation *dar/hacer* takes place only when two different varieties of Spanish are encountered. *Dar clases* (7a) in standard Spanish means 'to teach'; however, in Chilean Spanish, *hacer clases* (7b) is used.

(7) a. […] empezó a *dar clases* de Historia de la Literatura en la Universidad de Estocolmo. (*CREA*)
'He began teaching History of Literature at Stockholm University.'
b. […] los artistas no vivían de su pintura, sino de *hacer clases*. (*CREA*)
'Artists did not make living from their paintings, but from giving classes.'

Of the LCVs left I will concentrate on LVCs with nouns belonging to the semantic field of communication: *advertencia* 'warning', *discurso* 'speech', *indicación* 'instruction', etc.

4 *Dar* 'to give'/ *hacer* 'to make/do' and Communication Nouns

According to my hypothesis, LVs are linked syntagmatically to the noun within the same LVC and paradigmatically to the related full verb by means of semantic components. As the goal of this paper is to prove that the selection of *dar* and *hacer* by a communication noun is based on semantic grounds, we should discover which part of the meaning of the noun is emphasized when it combines with *dar*, and which part when *hacer* is chosen. Once we find out the highlighted parts of meaning of the noun, we will be closer to building up the descriptions for the LVs. In addition to nouns, the description of LVs must be related to the meaning of full verbs, because both the full and the light verb are lexical units included in the same vocable. Thus, in order to reach the goal, this Section starts with an overview of some general characteristics of *dar*/*hacer* as full verbs. It continues with an analysis of *dar*/*hacer* with communication nouns.

4.1 *Dar* and *hacer* as Full Verbs

As a basic definition for *dar* 'to give' as a full verb, we propose the following:

'X da Y a Z' 'X gives Y to Z' = 'X causes Y to go from the place where X is located to the place where Z is located'

Dar is a predicate of three Sem-actants (X, Y, Z). In its government pattern (GP) (8a), for each Sem-actant the corresponding Deep-Syntactic actant and the surface means of expressing it in a sentence is specified. In (8b), *alguien* 'somebody' is the grammatical subject (S) of *dar*, *le* 'him', the indirect object and *un regalo* 'a present', the direct object.

(8) a.

X = I	Y = II	Z = III
N	N	*a* N

b. Alguien le *dio* un regalo.
'Somebody gave him a present'

Just as *tener* used to be characterized as a verb of possession in verbal classifications, *dar* may be described as a verb of transfer of possession. In this general use, there is a volitive agent who transfers an object to a recipient. According to Demonte (1990), direct objects can be classified in three major groups concerning their semantic roles: 1) *effective objects*, for direct objects denoting the result or the process of the verb, 2) *affective objects*, for direct objects referring to objects whose state or properties can be impacted by the action of the verb, and 3) *displaced objects*, for those representing the moving object. In this sense, the direct object of *dar* can be characterized as a displaced object, that is, an object which changes its location.

In contrast to *dar*, *hacer* is a predicate of two Sem-actants (X and Y), which can be defined as:

'X hace Y' 'X makes/does Y' = 'X causes Y to come into existence'

In (9a) I present the GP for *hacer* and an example (9b). In (9b) *el carpintero* 'the carpenter' is the S of *hacer*, and *una mesa* 'a table', its direct object.

(9) a.

X = I	Y = II
N	N

b. El carpintero *hace* una mesa.
'The carpenter makes a table'

The description 'to cause to come into existence' can cover meanings such as 'to create/produce something material/immaterial'. In (9b) the direct object of *hacer* is an effected object (executed, produced). Thus, *hacer* can be characterized as a verb of creation or realization.

4.2 *Dar* and *hacer* in LVCs with Verbal Communication Nouns

I will focus on a group of nouns included in the semantic field of verbal communication nouns. Usually these nouns have three Sem-actants (X, Y and Z): X, being the speaker (or communicator); Y, the recipient; and Z, the message. For instance, *la advertencia del hombre* (X) *a la mujer* (Y) *de que no mirara hacia atrás* (Z) 'the man's warning to the woman not to look backwards'. The dominant semantic component of these nouns is 'an action through which X says Z to Y'. In most cases, these nouns are deverbal nouns, morphologically and semantically derived from verbs denoting speech acts: assertives, directives, commissives, expressives, and declarations (Searle, 2005). The dominant types here are the assertive and directive ones. Furthermore, communication nouns are part of vocables which show a clear pattern of regular polysemy (Apresjan, 1974). Besides the lexical unit denoting 'an action of

saying' (10a), some of the nouns have another lexical unit referring to 'a text expressing the content of the action' (10b)[1]:

(10) a. [El político] hace una *advertencia:* Mis clientes no desean ser identificados. *(CREA)*
'[The politician] gives a warning: My clients do not want to be identified.'
b. […] ha publicado una *advertencia… (DSLE)*
'[The management] has issued a warning'

Some of the nouns related to directive speech acts (e.g. *advertencia* 'warning', *recomendación* 'recommendation', *sugerencia* 'suggestion') co-occur with both *dar* and *hacer* LVs:

(11) a. [...] es probable que se decida *dar una última advertencia* a Lagos… (*CREA*)
'It is likely to be decided to give a final warning to Lagos'
b. Te voy a *hacer una advertencia*, muchacho. (*CREA*)
'I will give a warning to you, boy.'

(12) a. [...] *la recomendación que le dio* la curandera [...] o cambias de cama… (*CREA*)
'the recommendation that the healer gave [...] either you change your bed'
b. [...] es conveniente antes de explicar la estructura […] *hacer una recomendación…* (*CREA*)
'It is convenient before explaining the structure […] to make a recommendation…'

(13) a. *Había dado sugerencias*, órdenes… (*CREA*)
'He had given suggestions, orders'
b. [...] me *ha hecho* importantes *sugerencias…* (*CREA*)
'He has made important suggestions'

Examples extracted from *CREA* and *CdE* show a clear preference for *hacer*, underlining its capacity to denote the 'production/creation' of a speech act and, simultaneously, the existence of a volitional agent. However, as the examples refer to communicative acts, it seems that *dar* is also suitable because *dar* has come to be used for the transfer of communication (Newman, 1996:136) from its original meaning of transfer of possession. Indeed, in (11-13) *dar* and *hacer* LVs are mutually interchangeable with respect to the noun under analysis.

To better understand the peculiarities of the alternation *dar*/*hacer* with the same noun, we will compare nouns, somehow similar to the above mentioned, which co-occur exclusively either with *hacer* or with *dar*. In (14), *comentario* 'comment', *aclaración* 'clarification' or *observación* 'remark' co-occur only with *hacer*:

(14) a. […] si Marta *hace algún comentario*, siempre es de orden técnico. (*CREA*)
'If Marta makes a comment, it is always of a technical nature.'
b. [Sus palabras] han sido mal interpretadas, y le comunicó que *haría una aclaración…* (*CREA*)

---

[1]   Instead of the meaning "a text expressing", it is also possible to find the sense "what is said" (e.g. *comentario* 'comment', *aclaración* 'clarification') or 'public act where' (e.g. *mitin).* Milićević & Polguère's (2010) approach to communication nouns in French is done from the perspective of semantic ambivalence rather than from that of polysemy. The authors propose three patrons of semantic tags for these nouns according to the primary and secondary denotation of the nouns: 1) statement and/or linguistic communication, 2) linguistic communication and/or text, and 3) linguistic communication and/or what is said.

'[his words] have been misinterpreted, and he told him he would make a clarification.'
c. El autor *hace una observación* importante al mencionar que el concepto… (*CREA*)
'The author makes an important point when he mentions that the concept…'

In (15), nouns like *opinión* 'opinion' and *información* 'information' only combine with *dar*:

(15) a. Daniel, desde su mesa, le *dio la información* pedida al viejo. (*CREA*)
'Daniel, from his table, gave the requested information to the old man.'
b. Tampoco esta vez me *dio su opinión*. (*CREA*)
'Once again he did not give his opinion.'

When (14) and (15) are compared, the most striking difference among the nouns is related to the nature of the Sem-actant referring to the 'message'. Nouns combining with *hacer* in (14) have a dominant semantic component of action which can be formulated as "creation, production". For instance, in '*el comentario de X a Y sobre Z*' 'X's comment to Y about Z' there is a human agent X who creates a message/text by which s/he develops an aspect included in another text Z (14a). Basically, we can say that nouns, which select exclusively *hacer*, contain a Sem-actant 'message' that did not exist previously. The use of *hacer* highlights that the content of the message is created. Therefore, we can conclude that the 'message' of these nouns is an affected object.

Nouns co-occurring with *dar* in (15) also have a dominant semantic component of action. Nevertheless, here the message is not created, but merely transferred. As it can be noticed in *la información de X a Y de Z* 'X's information to Y about Z', there is a human agent X who transfers a message Z to Y (15a). The content of the message already existed and it was known by X; however, the target is also to make it known by Y. The same applies to *opinión* 'X's opinion about Y' (15b), since the holder of the opinion knows its content and s/he transfers it to Y. In these cases, it can be suggested that the message is a displaced object.

Taking cases in (14) and (15) as granted, it will be expected that when a noun co-occurs with both *dar* and *hacer,* it also contains both components 'transfer' and 'creation' of the message. Returning to the cases in (11-13), it can be noted that in e.g. *la advertencia de X a Y de que Z* 'X's warning to Y that Z' (11), X creates a message Z telling that something bad will happen to Y if Y does something, and at the same time that X is enunciating the content of the warning, X is transferring it to Y. The same holds for *recomendación* 'X's recommendation to Y that Z' (12) and *sugerencia* 'X's suggestion to Y that Z' (13). Thus, in all these cases, the message is both an affected and displaced object.

The contrast between *dar* as a verb of transfer and *hacer* as a verb of creation can be illustrated also with other verbal communication nouns such as *discurso* 'speech', *mitin* 'political speech', *declaración* 'declaration', *descripción* 'description', *referencia* 'mention', *rueda de prensa* 'press conference', etc. Let me illustrate here some examples with *discurso* 'X's speech to Y about Z' (16), in this case, 'X's formal exposition on a topic addressed to a collectivity of individuals' and *mitin* 'X's political speech to Y' (17).

(16) a. […] va a *dar un discurso* magistral ante un foro de estudiantes. *(CREA)*
'He will give a keynote speech to a forum of students'
b. […] el cardenal *hizo* un pequeño *discurso… (CREA)*
'The Cardinal made a little speech.'

(17) a. […] *daba mítines* en el pueblo… *(CREA)*
'He used to give political speeches in the village.'
b. […] se debería *hacer un mitin* muy corto, diez o doce minutos máximo. *(CREA)*
'It should be a very short political speech, ten or twelve minutes maximum.'

While LVCs with *dar* focus on the transfer of the message (either the recipients (16a) or the location (17a) of the speech are expressed), LVCs with *hacer* put the emphasis on the message and its creation. Adjectives such as *pequeño* 'small' accompanying *discurso* (16b) and *corto* 'short' – plus the temporal modifier – with *mitin* in (17b) are used to measure the length of the message, an activity closely related to its creation. In addition to this distinction, previous studies in Spanish about *dar* as an LV have pointed out that this verb always preserves a recipient (Battaner, 2011). However, with communication nouns, not only *dar* but also *hacer* usually expresses the Sem-actant of the noun referring to the receiver as its own SSynt-actant. Other studies have focused in *dar* as a verb denoting a change of localization. Along these lines, de Miguel (2008: 577) considers that the tendency of certain Spanish nouns such as *abrazo* 'hug', *beso* 'kiss', *golpe* 'blow', *grito* 'shout', which denote muscle impulses coming from the body to reach a target, to choose *dar* can be attributed to a feature of path/trajectory included in this verb. And this is also the reason why *explicación* 'explanation', a noun with a "target argument", selects *dar.* On the other hand, de Miguel (2008: 578) defends the fact that a noun like *caricia* 'caress', even if it has a receiver, chooses *hacer.* The reason is because this noun is constructed through the action of a subject without the implication of targets, impulses or trajectories. Thus, she concludes that nouns implying trajectory normally combine with *dar*, while nouns which preferably express an action co-occur with *hacer*. Alonso Ramos (1997) distinguishes two "senses" of the LV *dar* as a value of **Oper₁**.The separation of senses is based on the semantic class of the noun with which *dar* is combined and on the number of actants, but not on the semantic information of the LV itself.

5 Limits and Implications of the Study

My claim that the main difference between *dar*/*hacer* as LVs when they accompany communication nouns is that *dar* focuses on the transfer of the message while *hacer*, on the creation of the message itself, is mainly drawn from the analysis of a set of communication nouns which can combine with both verbs and two sets of communication nouns, one co-occurring only with *dar* and the other one only with *hacer*. Even if these results are based on a homogeneous group of verbal communication nouns, it will be necessary to test other nouns from the same and other fields. Some observations made in other fields show that some nouns selecting both *dar* and *hacer* present the same opposition transfer/creation, e.g. nouns for monetary transactions such as *dar/hacer un préstamo* 'to give/make a loan' or nouns of social relations and activities such as *dar/hacer una visita* 'to pay a visit'. However, other nouns may show a slight different opposition e.g. *dar/hacer una señal*, where the volitive subject is emphasized with *hacer* against a neutral one with *dar*. All the semantic distinctions between LVCs with *dar* and *hacer* as well as a more detail study of the role played by geographical variation, should be integrated in the same model in order to draw a wider picture of these verbs. Some examples in the Introduction suggest that the opposition transfer/creation of a message, expressed by LVCs with *dar*/*hacer*, is not restricted to Spanish. Some communication nouns in other languages can also co-occur with both LVs, presumably based on the same distinction, e.g. Fr. *donner/faire un avertissement* 'to give/make a warning' or En. *to give/make a suggestion* or *to give/make a speech*. It should be important for both theory

and practice to find out to what extent this opposition is universal. As shown in the Introduction, equivalent nouns from different languages select the same verb. However, it is possible to find counterexamples. Certain nouns in certain languages can make a different selection, for instance, Sp. *hacer comentarios* 'to make comments' face Fn. *antaa kommentteja* (lit.) 'to give comments'. In this context, it would be important to verify whether the semantic components of LVs and/or nouns in one language differ from those of their counterparts in the other language.

6   Conclusions

This paper was intended to demonstrate that the selection of collocate verbs by nouns within the context of LVCs is based on semantic grounds. To support this claim we have tested the hypothesis of semantic compatibility. According to this hypothesis, collocate verbs, here *dar* and *hacer*, two of the most frequent Spanish LVs, are connected to the noun within the same LVC and to the related full verb by means of semantic links. In particular, we have used a selected corpus of nouns from the semantic field of verbal communication nouns which have the peculiarity of co-occurring with both *dar* and *hacer*. By contrasting these nouns with nouns which only combine either with *dar* or *hacer*, we have shown that the semantic component of the noun emphasized by *dar* can be paraphrased as "transfer of an existing message", and the semantic component focused by *hacer* can be described as "creation of the message". Thus, we have also verified that both components "message transfer" and "message creation" are present in the definition of nouns which can co-occur with both verbs. In addition to this, it has been confirmed that the "transfer" and "creation" components are also part of the definition of *dar* and *hacer*, respectively, as full verbs. With these results, we are taking a step forward in the elaboration of lexical entries for these LVs. Instead of listing under the entry of *dar/hacer* all the nouns which select *dar/hacer*, we could only specify that nouns including the semantic component "transfer" co-occur with *dar*, and those including "creation" choose *hacer*. However, there is still some way to go before the complete picture of these verbs is achieved.

Bibliography

Alonso Ramos, M. 1997. Coocurrencia léxica y descripción lexicográfica del verbo *dar*: hacia un tratamiento de los verbos soportes. *Zeitschrift für Romanische Philologie* 113:380–417.

Alonso Ramos, M. 2004. *Las construcciones con verbo de apoyo*. Madrid: Visor Libros.

Apresjan, J. 1974. Regular Polysemy. *Linguistics* 142:5–32.

Apresjan, J. 2009. The Theory of Lexical Functions: An Update. In *Proceedings of the Fourth International Conference on MTT*. 1–14. Montreal: OLST. http://olst.ling.umontreal.ca/pdf/ProceedingsMTT09.pdf

Apresjan, J. & M. Glovinskaja. 2007. Two Projects: English ECD and Russian Production Dictionary. In *Proceedings of the 3rd International Conference on MTT*. München: Wiener Slawistischer Almanach http://meaningtext.net/mtt2007/proceedings.

Battaner, P. 2011. Los verbos de frecuencia alta y el diccionario de aprendizaje. In Vázquez Laslop, M.E. et al. (eds). *De la lengua por sólo la extrañeza*, 313–332. México, D.F.: El Colegio de México.

Bosque, I. 2004a. La direccionalidad en los diccionarios combinatorios y el problema de la selección léxica. In Cabré, T. (ed). *Linguistica teórica*: *análisi i perspectives*. 13–58. Barcelona: Universitat Autònoma Barcelona.

Bosque, I. 2004b. *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: SM.

CdE: Davies, M. *Corpus del español* http://www.corpusdelespanol.org/x.as.

CREA: RAE: *Corpus de referencia del español actual* http://www.rae.es.

De Miguel, E. 2008. Construcciones con verbos de apoyo en español. De cómo entran los nombres en la órbita de los verbos. In Moreno Olza, I. et al. (eds). *Actas del XXXIX Simposio de la SEL*, 567–578. Pamplona: Universidad. http://www.unav.es/linguis/simposioel/actas/

Demonte, V. 1990.Transitividad, intransitividad y papeles temáticos. In Demonte, V. & Garza Cuarón, B. (eds). *Estudios de lingüística de España y México*, 115-150. México: El Colegio de México/UNAM.

DiCE: Alonso R, M. 2004. *Diccionario de colocaciones del español* http://www.dicesp.com.

Dubský, J. 1990. Al margen de la complementariedad léxico-sintáctica en español. *Hispanica Posnaniensia* 1:31–37.

Koike, K. 2001. *Colocaciones léxicas en el español actual. Estudio formal y léxico- semántico.* Alcalá de Henares: Universidad de Alcalá/Takushoku University.

Mel'čuk, I. 1992. Paraphrase et lexique. In Mel'cuk, I. et al. (eds). *Dictionnaire explicatif et comibinatoire du français contemporain,* 9-58. Montréal: l'Université de Montréal.

Mel'čuk, I. 1996. Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In Wanner, L. (ed). *Lexical Functions in Lexicography and Natural Language Processing,* 37-102. Amsterdam / Philadelphia, John Benjamins.

Mel'čuk, I., A. Clas & A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire.* Louvain-la-Neuve, Duculot.

Milićević, J. & A. Polguère. 2010. Ambivalence sémantique des noms de communication langagière du français. In Neveu, F. et al. (dirs). *Actes de la section «Lexique et morphologie» du 2e Congrès Mondial de Linguistique Française (CMLF'10),* 1029-1050. La Nouvelle- Orléans: Institut de Linguistique Française (ILF).

Newman, J. 1996. *Give: A Cognitive Linguistic Study.* Berlin: Mouton de Gruyter.

Sanromán Vilas, B. 2009. Towards a Semantically Oriented Selection of the Values of Oper$_1$: The Case of *golpe* 'Blow' in Spanish. In Beck, D. et al. (eds). *Proceedings of the Fourth International Conference on the Meaning-Text Theory,* 327-337. Montréal: OLST.

Sanromán Vilas, B. 2012. La hipótesis de la compatibilidad semántica en la selección de los verbos de apoyo. In Ridruejo, E. et al. (eds). *Tradición y progreso en la lingüística general,* 355-372. Valladolid: Universidad de Valladolid.

Searle, J. R. 2005. Una taxonomía de los actos ilocucionarios. In Valdés Villanueva, L. (ed). *La búsqueda de significado,* 448-475. Madrid: Tecnos.

Vivès, R. 1993. La prédication nominale et l'analyse par verbes supports. *L Information grammatical* 59:8-15.

Appendix

dar/hacer + communication noun	dar + communication noun	hacer + communication noun
dar/hacer una advertencia	dar un consejo	hacer una aclaración
dar/hacer una declaración	dar una explicación	hacer un comentario
dar/hacer una descripción	dar información	hacer una confesión
dar/hacer un discurso	dar su opinión	hacer una observación
dar/hacer un mitin	dar una respuesta	
dar/hacer una recomendación		
dar/hacer una referencia		
dar/hacer una rueda de prensa		
dar/hacer una sugerencia		

Table 1: Spanish LVCs with *dar / hacer* and communication nouns discussed in the paper

dar/hacer + communication noun	dar + communication noun	hacer + communication noun
Fr. *donner/faire una* avertissement	Fr. *donner un conseil*	Fr. *faire une confession*
En. *to give/to make a speech*	It. *dare un consiglio*	It. *fare una confessione*
En. *to give/to make a suggestion*	Pg. *dar um conselho*	Pg. *fazer uma confissão*
	En. *to give advice*	En. *to make a confession*
	Gm. *einen Ratschlag geben*	Bq. *aitortza egin*
	Rs. *davat' sovet*	Fn. *tehdä tunnustus*
	Bq. *aholku eman*	
	Fn. *antaa neuvoa*	
	Fn. *antaa kommentteja*	

Table 2: LVCs with *dar / hacer* counterparts and communication nouns in languages other than Spanish discussed in the paper

LVCs				causative collocations
dar/hacer + noun	*dar* + noun	*hacer* + noun	others	
dar/hacer clases *dar/hacer un préstamo* *dar/hacer una señal* *dar/hacer una visita*	*dar un abrazo* *dar un beso* *dar un golpe* *dar un grito* *dar una tregua2*	*hacer una caricia* *hacer una tregua1*	*sentir un disgusto* *sufrir un disgusto* *tener un disgusto* *llevarse un disgusto* *pasar un disgusto*	*dar/hacer gracia* *dar/hacer miedo*

Table 3: Other collocations discussed in the paper

The "Colorative" Construction in Estonian

Rosina Savisaar

Université de Lorraine
34 cours Léopold, 54052 Nancy, France
University of Oxford
Wellington Square, Oxford, OX1 2JD, United Kingdom
rosina.savisaar@psy.ox.ac.uk

Abstract

The Estonian "colorative" construction is a verbal construction in which the meaning of a verb in the infinitive V_{INF} is modified by a finite verb V_{FIN}, which is its syntactic governor. Often the V_{FIN} is an intensifying collocate of the V_{INF} and should therefore be described using the lexical function (LF) **Magn**. However, the syntactic structure of the construction is incompatible with the use of **Magn**. The complex LF $\mathbf{V_0Magn}$ is proposed as a solution. This LF applies to a verbal lexical unit and returns a verb that takes the keyword as its deep syntactic actant II and functions as an intensifier of the keyword. In addition, examples of the colorative construction where the semantic contribution of the collocate is different from intensification will be briefly introduced, along with proposed descriptions of these collocations at the deep syntactic level.

Keywords

Estonian, colorative construction, dependency reversal, ideophones, lexical functions.

1 Statement of the Problem

The so-called *colorative construction* [=CC] in Estonian consists of a finite verb V_{color} which syntactically governs a verb in the infinitive – "$V_{(color)}$–**synt**→V_{INF}", the semantic function of the finite $V_{(color)}$ being to modify the meaning of its syntactic dependent. For instance, in (1), the $V_{(color)}$ RAIUMA 'to chop' syntactically governs the V_{INF} KÖHIMA 'to cough', while from the semantic point of view the role of the first verb is to intensify the meaning of the second.

(1) | Toomas | + Ø | + Ø | raiu | + Ø | + Ø | + b | köhi | + da |
     | Toomas | SG | NOM | chop | NON-PAST | IND | 3SG | cough | INF |

'Toomas coughs intensely.' (lit. 'Toomas chops to cough.')

The phrase *raiuma köhida* 'cough intensely' (lit. 'chop to cough') is a collocation of the verb KÖHIMA with the verb RAIUMA 'to chop' as the intensifier collocate. RAIUMA 'to chop' cannot be used, for example, to intensify the meaning of the verb TANTSIMA 'to dance' (*raiuma tantsida*), whereas the verb VEHKIMA 'to flail' can intensify the meaning of TANTSIMA but not that of KÖHIMA (*vehkima tantsida* vs. * *vehkima köhida*).[1]

In Meaning-Text Theory (Mel'čuk, 1997), a collocate that intensifies the meaning of its base is usually described using the lexical function [=LF] **Magn**, which – as it fits a modifier – depends on its keyword (Mel'čuk et al., 1995: 136). In case of the CC, however, the syntactic structure of the collocation is incompatible with that of a **Magn**: rather than being its syntactic dependent, the collocate governs the base of the collocation. In other words, the inversion of dependencies that normally occurs between the intensifying and intensified semantemes (or, more accurately, between the corresponding lexical units) during the passage from the semantic structure [=SemS] to the deep- and surface-syntactic structures [=DSyntS, SSyntS] (Mel'čuk, 1997) does not occur in the Estonian CC.

This paper uses the Estonian CC as an example of how the system of LFs can be used to describe collocations where a verb, rather than being intensified by an adverbial (which is more conventional), is intensified by another verb that governs it. In addition, it will be shown that the same reasoning can also be applied to the examples of CC where the semantic contribution of the collocate is not that of intensification. Only examples of the CC that are collocations will be considered.

2   Overview of the Colorative Construction

The CC is widespread in Finnic languages (Saukkonen, 1966). Notably, it also exists in Finnish, where it has been studied much more extensively than in Estonian (see e.g. Rytkönen 1937, Luttinen 2000 or Jarva and Kytölä (2007)). Descriptions of the Estonian CC are mainly found in reference grammars (Erelt et al., 1993; Mihkla et al., 1974) or in the context of research on ideophones (for example, Mikone (2002) whose work is concerned with both Finnish and Estonian). Estonian also has an analogous nominal construction, where the meaning of a noun, rather than being modified through the use of an adjective, is modified by another noun that is also its syntactic governor (Leinonen, 2010) (this is similar to examples such as *a bear of a man* in English or *cet idiot de Jacques* in French).

The head of the CC is usually a verbal ideophone. More generally speaking, an important property of the construction is to be firmly anchored in the sensory world, with the colorative verb expressing not so much what happens but what the striking features of that event are from a sensory, most often auditory, point of view. This property makes colorative verbs

---

[1] Strictly speaking, the glosses provided for these two colorative verbs are inaccurate: even though these verbs belong to vocables that include lexical units signifying 'to flail' and 'to chop', respectively, the colorative verbs themselves are simple intensifiers. These slightly misleading glosses are used to show the metaphor that links the colorative verbs to the other lexical units in their vocable.

notoriously hard to define, although the fact that they very often appear as collocates also contributes to this difficulty.

A somewhat related problem is that of determining the actantial structure of colorative verbs. In Estonian linguistics, the two verbs of the CC are usually described as forming a complex predicate, the actantial structure of which as a whole is inherited from the V_{INF} (Erelt et al., 1993). This description is of course not valid in the current framework as the verbs are two separate lexical units expressing two different predicates and both necessarily have their own actantial structure. In the present paper, V_{INF} is described as filling the second semantic actant slot of $V_{(color)}$ (though see Subsection 4.2. for examples where this is not the case). A description of this kind gives rise to definitions such as the following (the propositional form is in Estonian, while the definition itself has been translated into English for clarity):

'X möirgab Y-da' = 'An animate X makes an intense and low <u>sound</u> while doing Y'

The communicatively dominant node [=CDN] (Mel'čuk, 2001:34) is 'sound' ('hääl' in Estonian), verbalized in the linear definition using its **Oper$_1$** support verb, MAKE (TEGEMA in Estonian).

The main reason for including a slot corresponding to the semanteme 'V_{INF}' in the actantial structure of '$V_{(color)}$' is that the set of verbs that can act as the head of a CC appears to be limited, with many ideophonic verbs that seem to possess the perfect profile for being a colorative verb being restricted from being the head of a CC for no obvious semantic reason. It therefore seems that being able to be the head of a CC is a fundamental feature of these verbs and it is natural to express this property in their actantial structure. There seems nevertheless to be considerable interindividual variation in this respect, with certain speakers being much less restrictive as to the verbs they accept as head of the CC. With respect to the linguistic system of these speakers, the description proposed in the present paper is therefore inaccurate. However, another reason for proposing it nevertheless is that certain colorative verbs obligatorily have a complement, this complement being either V_{INF} within the CC or the $S_0(V_{INF})$ inflected in the partitive case (see, for instance, the verb VEHKIMA 'to flail').

3 The Deep and Surface Syntactic Structure of the Colorative Construction

The CC is analogous to many other constructions in a large number of languages, where the direction of the syntactic dependency between two lexical units is unusual given the semantemes that they express. Most importantly, the CC is the verbal equivalent of a nominal collocation in which an intensifying collocate, rather than being an adjective syntactically dependent on its base, is another noun and syntactically governs the base (as mentioned in Subsection 2, this nominal construction also exists in Estonian). One such collocation appears in one of the examples Milićević (2003) gives for her lexico-syntactic rule of equivalency 21-b:

> (2) *It was a whale* [**S$_0$Magn**(L)] *of a feast* [L]. ≡ *It was a huge* [**Magn**(L)] *feast* [L].

The collocation is described using the complex LF **S$_0$Magn** because the collocate is conceptualized as a nominalization of a **Magn**. An analogous solution is adopted in the

present paper to describe the intensifying CC: the LF V_0**Magn** is a complex LF whose keyword is its deep-syntactic actant II and that, in Estonian, returns a verb whose subject is the lexical unit expressing the semantic actant **1** of the keyword and whose other surface syntactic dependent is the keyword inflected in the infinitive:
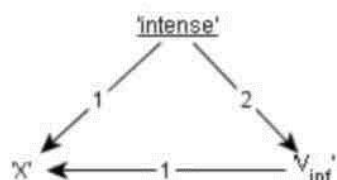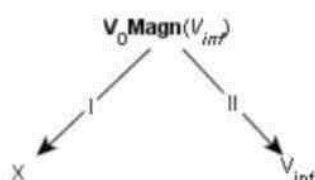

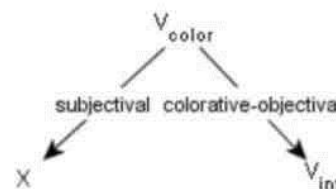Figure 1: SemS of the LF   Figure 2: DSyntS of the   Figure 3: SSyntS of the LF

V_0**Magn**   LF V_0**Magn** V_0**Magn**

The solution proposed above has the advantage of both featuring a strong semantic link with the LF **Magn** and of giving the resulting collocation a plausible syntactic structure. This last property makes it worthwhile to use a new complex LF rather than to simply describe examples of the CC as fused values of **Magn**. Though this latter description would be correct and would not necessitate introducing a new complex LF, it would mean that at the level of the DSyntS, there would be no way of distinguishing between the CC and other fused values a **Magn** can return. However, it is important to be able to set these collocations apart as a class already at the level of the DSyntS, both because of how systematic the CC is as a way of intensifying verbs (and of otherwise modifying their meaning) and because similar cases of dependency reversal exist in many languages (see, for instance, Foolen (2004) for nominal examples from Germanic and Romance languages or some of the glosses in the present paper for possible verbal analogues in English). Moreover, all that is needed is to extend a LF as fundamental as V_0 to apply not only to lexical units but also to LFs. No unnatural extra machinery is thus needed.

Finally, it is most of the time possible to paraphrase a CC by using an **Adv$_0$** of the $V_{(color)}$, see (3-4). This semantic derivative is often also a morphological one (this is not the case for (1) but is so for (3), for instance). The **Adv$_0$** is usually realized in the SSyntS as either the gerundive of the colorative verb or its **S$_0$** inflected in the adessive case (one of the functions of this case in Estonian is to permit the use of nouns in an adverbial role). This paraphrase is of course analogous to the one Milićević (2003) illustrated with the example in (2) and is an argument for describing the collocation in a similar way. The exact type of paraphrase seen in (3)-(4) is described by Milićević as part of her Rule 21-a of lexical-syntactic equivalency.

(3) *Mari* + *Ø* + *Ø* *mürista* + *Ø* + *Ø* +*b* *naer* +*da*
    Mari SG NOM rumble   NON-PAST IND 3SG laugh INF
    'Mari roars with laughter'

(4) Mari + *Ø* + *Ø* *naera* + *Ø* + *Ø* +*b* *mürin* + *Ø* +*al*
    Mari SG NOM laugh NON-PAST IND 3SG rumble   SG ADE
    'Mari laughs roaringly'

The surface-syntactic relation [=SSyntRel] **colorative-objectival** used in Figure 3 is provisional and further research would almost certainly make it possible to use one that does not only pertain to the CC. It is, however, not possible to use the SSyntRel **direct-infinitival-objectival** to describe this construction. Postulating this SSyntRel for Estonian is necessary because the prototypical dependent (Mel'čuk, 2009) of the SSyntRel **direct-objectival** is clearly a noun, however, not all verbs traditionally described as governing direct objects admit nominal dependents (see, for example, the verb JULGEMA 'to dare', the non-pronominal dependents of which can only be verbal).[2] However, one of the properties of the SSyntRel **direct-infinitival-objectival** is that the dependent can be substituted by personal and interrogative pronouns *(julgeb karjuda* 'dares to scream' ~ *julgeb seda teha* 'dares to do it' ~ *julgeb mida teha* 'dares to do what'). With the CC, this is impossible *(raiub köhida* 'coughs intensely' ~ **raiub seda teha* 'does it intensely' ~ **raiub mida teha* 'does what intensely').

It is also not possible to use a representation closely following Erelt *et al.* (1993), who describe the CC as a chain verb in their reference work on Estonian syntax. In Estonian linguistics, the term *chain verb* refers to a set of verbal constructions of the form "V-**synt**→V $_{\text{INF}}$", where the descriptive content is expressed by V$_{\text{INF}}$, with V mainly contributing information relevant to modality, aspect, causality or manner (Erelt et al., 1993). It is not possible to postulate a common SSyntRel that could be used to describe the relationship between the two verbs in all these cases because some of them allow for V$_{\text{INF}}$ to be substituted by a pronoun whereas others do not. Finally, the fact that V$_{\text{INF}}$ is described as deep-syntactic actant II of V$_{(color)}$ means that treating it as an adverbial of V$_{(color)}$ (along the lines of Mihkla *et al.* (1974)) would also be undesirable.

It is hard to say more without undertaking a systematic investigation of the SSyntRels implicated in Estonian verbal constructions, which is why the decision was made to use the very restrictive SSyntRel **colorative-objectival.**

4 More Complex Cases

4.1 Semantic Phenomena other than Intensification

The semantic contribution of a colorative verb to the collocation is not necessarily that of intensification. However, the same reasoning can be used as when producing the LF **V$_0$Magn**. The simplest case is that of **V$_0$AntiMagn**, illustrated here:

(5) *käsi* *+ Ø* *+ Ø* *tuika* *+ Ø* *+ Ø* *+b* *valuta* *+da*
    hand SG   NOM throb PAST 3SG   3SG   hurt   INF
    '(Someone's) hand throbs with pain'

Other examples of the CC hold the same type of relationship that cases of **V$_0$Magn** have to **Magn** to other LFs. For instance, the verb JORISEMA 'sing off key' can be described using the LF **V$_0$AntiBon**:

---

[2]   See Iordanskaja and Mel'čuk (2009) for the same reasoning for French.

(6)   *Mari* *+ Ø* *+ Ø* *jorise*   *+ Ø*   *+ Ø* *+b*   *laul*   *+da*
         Mari SG   NOM sing.off.key NON-PAST IND 3SG   sing   INF
         'Mari sings off key'

Still further CCs can only be described using a non-standard lexical function. This is the case for the collocation *kõõksuma naerda* 'to laugh in a manner that sounds as if one was suffocating', KÕÕKSUMA 'to make a sound as if suffocating' being a collocate of NAERMA 'to laugh':

> as if X was *suffocating(naerma) = kõõksuma*

An interesting property of the CC is that, often, the semantic contribution of the collocate consists simply in emphasizing the sensory aspects of the linguistic situation (Mel'čuk 2004) in question without giving any additional details on it. For example, in the collocation *vuristama kedrata* 'to provoke a whirring sound by spinning', the collocate VURISTAMA 'to provoke a whir' is not very informative as spinning necessarily entails a whirring sound. However, the collocate draws attention to the auditory experience of spinning.

This can also be described using a non-standard LF:

> the speaker emphasizes the senses *(ketrama) = vuristama.*

4.2   Two Types of Colorative Verbs

A second, more problematic issue, is that of the existence of a large number of $V_{(color)}$ that are more precise synonyms of the VIN_F that they govern[3]. This is evidenced by the fact that in a CC containing a $V_{(color)}$ of this type, it is possible to omit VI_{NF} without changing the meaning of the verb phrase[4]:

(7) *Mari + Ø* *+ Ø* *lagista*   *+ Ø*   *+ Ø* *+b*   *naer* *+da*
    Mari SG   NOM bellow.with.laughter NON-PAST IND 3SG   laugh INF
    'Mari bellows with laughter'

(8) *Mari + Ø* *+ Ø* *lagista*   *+ Ø*   *+ Ø* *+b*
    Mari SG   NOM bellow.with.laughter NON-PAST IND 3SG
    'Mari bellows (with laughter)'

Verbs of this type have very different semantic properties than other colorative verbs. As indicated in Subsection 2, the other colorative verbs are at least biactantial (with V_{INF} filling the second semantic actant slot) and their CDN is usually drawn from the lexical field (Mel'čuk et al., 1995) of sensory experience. With verbs like LAGISTAMA 'to bellow with laughter', however, it is not possible to model V_{INF} as an actant of $V_{(color)}$ (because the same semanteme cannot be both the CDN and a constant participant (Mel'čuk, 2004) in a definition) and they can therefore be monoactantial. Moreover, their CDN corresponds to the semanteme expressed by V_{INF}. For example:

---

[3]   The lexical unit JORISEMA, 'sing off key', alluded to above, is among these verbs.

[4]   At least not the semantic structure. There might be differences in the semantico-communicative structure.

'X lagistab' = 'The person X <u>laughs</u> in an intense and resonant manner'

There truly is a difference in the CDN, and it becomes apparent when these verbs are negated: with verbs like LAGISTAMA 'to bellow with laughter', the negation bears on the semanteme expressed by V_{INF} and not on a semanteme of sensory experience – that is to say, the affirmative and the negative convey whether or not the person laughed and not whether or not he or she produced an intense and resonant sound.

However, despite these differences between the two classes of verbs, the DSyntS of all CCs that are collocations can be represented in the same manner. What matters to the LF is the semantic configuration it receives as input: in a language such as Estonian, a verbal semanteme filling the first semantic actant slot of the meaning 'intense' will call upon the LF **V_0Magn** (and other syntactic configurations, notably **Magn**) and will do so independently of the semantic actantial structure of that verbal semanteme. It may of course return no value if the verb in question has no **V_0Magn**. Cases where the V_{color} appears without V_{INF} can simply be described as exhibiting a fused value of **V_0Magn**. For instance:

V_0Magn(*naerma*) = *müristama, ..., (//) lagistama, ...*

One notes that the fused value of **V_0Magn**, for instance, is equivalent to the fused value of **Magn** applied to a verb (**Magn**(*naerma*) = *//lagistama* is also correct). The only reason a description using **V_0Magn** is to be preferred here is that all V_{color} that can be returned as fused values can also be returned as part of a non-fused value when they are the head of a CC (this possibility is indicated by the brackets around the forward slashes). It would not be possible to convey this information simply by using **Magn**. It would also not be possible to present V_{color} of the two kinds as part of the same list of collocates when they can intensify the same V_{INF} (as has been done in the example above).

Conclusion

The case of the CC illustrates the fact that fundamental semantic configurations, such as intensification, that are preferentially associated with a particular deep syntactic structure can also be expressed using other syntactic means. However, by using "verbalized" (or modified in other, similar ways – see the nominalization of a **Magn** in (2)) versions of existing LFs, it is possible to describe collocations involving such phenomena in a way that is both syntactically explicit and conserves the link to well-known LFs.

Acknowledgements

Abbreviations

3SG	3rd person singular	INF	infinitive

ADE	adessive	LF	lexical function
CC	colorative construction	NOM	nominative
CDN	communicatively dominant node	SG	singular
DSyntS	deep syntactic structure	SemS	semantic structure
IND	indicative	SSyntS	surface syntactic structure

Bibliography

Erelt, M., R. Kasik, H. Metslang, H. Rajandi, K. Ross, H. Saari, K. Tael & S. Vare. 1993. *Eesti keele grammatika II. Süntaks. Lisa: kiri,* Tallinn: Keele ja Kirjanduse Instituut.

Foolen, A. 2004. Expressive binomial NPs in Germanic and Romance languages. In *Studies in Linguistic Motivation*, eds. Günter Radden and Klaus-Uwe Panther. Berlin–New York: Mouton de Gruyter. 75-100.

Iordanskaja, L., and I.A. Mel'čuk. 2009. Establishing an Inventory of Surface-Syntactic Relations: Valence-controlled Surface-dependents of the Verb in French. In *Dependency in Linguistic Description*, eds. A. Polguère and I.A. Mel'čuk. Amsterdam/Philadelphia: John Benjamins. 151-234.

Jarva, V. & S. Kytölä. 2007. The Finnish colorative construction and expressivity, *SKY Journal of Linguistics*, 20: 235–272.

Leinonen, M. 2010. Tytöntyllerö ja pojanjolppi suomessa ja virossa, *ESUKA - JEFUL,* 1(2): 193- 217.

Luttinen, S. 2000. *Verbaaliset koloratiivikonstruktiot eräissä pohjoissavolaisissa murteissa.* Master's thesis. Jyväskylän yliopisto.

Mel'čuk, I.A. 1997. *Vers une linguistique Sens-Texte. Leçon inaugurale,* Paris: Collège de France.

Mel'čuk, I.A. 2001. *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences,* Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, I.A. 2004. Actants in Semantics and Syntax I: Actants in Semantics, *Linguistics,* 42(1):1-66.

Mel'čuk, I.A. 2009. Dependency in Natural Language. In *Dependency in Linguistic Description,* eds. A. Polguère and I.A. Mel'čuk. Amsterdam: John Benjamins. 1-101.

Mel'čuk, I.A., A. Clas & A. Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire,* Paris: Duculot.

Mihkla, K., L. Rannut, E. Riikoja & A. Admann. 1974. *Eesti keele lauseõpetuse põhijooned (1. osa),* Tallinn: Valgus.

Mikone, E. 2002. *Deskriptiiviset sanat. Määritelmät, muoto ja merkitys,* Helsinki: Suomalaisen Kirjallisuuden Seura.

Milićević, J. 2003. *Modélisation sémantique, lexicale et syntaxique de la paraphrase langagière.* Doctoral thesis. Université de Montréal.

Rytkönen, A. 1937. Koloratiivinen konstruktio, *Virittäjä,* 41:95-104.

Saukkonen, P. 1966. *Itämerensuomalaisten kielten tulosijain nitiivirakenteiden historiaa II Suomalais-ugrilaisen seuran toimituksia 140,* Helsinki: Suomalais-ugrilaisen seura.

Reconsidering Attenuative (Prefixed Comparative) in Russian: Between Grammar and Lexicon[1]

Dmitri V. Sitchinava

Institute of Russian language
Russian Academy of Sciences
mitrius@gmail.com

Abstract

The paper deals with the Russian prefixed comparative (attenuative) form and focuses on its history, the development of the attenuative semantics and the lexical preferences. It is shown, in particular, that the attenuative semantics was grammaticalized only in the beginning of the 20th century (constructions like гораздо побольше 'far bigger'), and that it is preferably used with the comparatives signifying positive qualities including multiple reactions (like 'accurate', 'smart', 'attentive', 'clever' etc.).

Keywords

Morphology, Russian language, Old Russian, comparative, lexicalization, corpus linguistics

1   Introductory Remarks

The Russian prefixed comparative of the kind *побольше* '[slightly] more, bigger' (cf. *больше* 'more, bigger') has been studied intensively within the MTT framework. It was Igor Mel'čuk who placed this obviously grammatical unit, almost systematically disregarded by Russian grammar tradition, within a broader picture of grammatical categories, as a "quasi-grammeme" labeled "attenuative" (Мельчук 1998). A recent excellent treatment of it was accomplished by Igor Boguslavsky and Leonid Iomdin and presented on an earlier MTT conference (see the English version of their paper: Boguslavsky, Iomdin 2009; the Russian version: Богуславский, Иомдин 2009). This paper presents some results of its corpus-based study with the Russian National Corpus (RNC) as the main source of reference and statistics.

---

2 Earlier History

The comparative with *no-* is not a common Slavic phenomenon and not even a common one for the Eastern Slavic languages. It becomes frequent in Middle Russian after the 15th century, with only few examples, mainly of the adverbial *подале* 'further' in Old Russian (see below). It is a marginal form in Ukrainian and Belarusian, as this is shown by the parallel Ukrainian-Russian and Belarusian-Russian corpora within the RNC. The overwhelming majority of the Russian *побольше-* like forms have simple positive or comparative forms as their Ukrainian and Belarusian counterparts, sometimes with adverbials or suffixes meaning 'as possible' or 'a bit'. The forms with *no-* prefix do exist in these two languages, but they are very rare, mainly the locative and quantitative adverbials (cf. Ukr. *подалі* '[a bit] further', *побільше* '[a bit] more', *поменше* '[a bit] less'). There were adjectives of the kind in Middle Bulgarian like *по-лѣп-ш-аа* 'slightly-beatiful-COMP-neuter.N.PL' (Vaillant 1958). In modern South Slavic (Bulgarian, Macedonian and Serbian) and Baltic languages the *no-* (*pa-*) is prefixed to the positive (not comparative) stem. In Bulgarian and Macedonian it signifies solely comparative as such (Bulg. хубав 'good' — по-хубав 'better' — най-хубав 'the best') while in Serbian and the Baltic languages it only stands for attenuative as such (Serb. *подугачак* 'a bit long' (Крысько 2006: 371), Lith. *padidelis* 'slightly big / too big'), but nowhere do these properties co-occur.

The rise and development of these forms was not studied until the 1990s. The most recent Old Russian grammar has a special section on these forms (Крысько 2006) based on the groundbreaking work by Alexei Gippius (Гиппиус 1994). Gippius (and Andrei Zalizniak, whom he cites) were the first to show that the adverbs like *подале* 'further' are formed from the preposition *no* in its limitative meaning (here, 'until') combined with the still-declined Old Russian comparative form in neuter accusative. Likewise, a class of Old Russian adverbs existed: it was formed with the preposition *въ* 'in, into' and the comparative looked like *въшире*, literally, 'into a broader (place)' > Modern Russian *вширь* 'broadwise'. The forms with *no-* initially meant 'until an X-er (place, degree)': *по дале* 'until a further (place)' (Крысько 2006: 372). It may be not by chance that chronologically the first Old Russian formations of this kind are locative adverbials, like *подале* 'further', *подъле* 'until the larger side' > (Early) Modern Russian *подле* 'near' etc. (ibid: 370, 372, 374). The corpora data, viz. the Middle-Russian subcorpus of the RNC show that the locatives *повыше* '(slightly) higher' and its antonym *пониже* '(slightly) lower' are also attested very early, starting with the 15th century. The quantatives *поменьше* '(slightly) less' and *поболе* '(slightly) more' are among the earlier cases as well. It should be noted that exactly these two groups of prefixal comparatives are the most frequent in Modern Russian (see below Section 3) and the most stable in Ukrainian and Belarusian where the constuction was not fully grammaticalized (see above).

The RNC and the Dictionary of the 11-17th centuries Russian (Словарь XI-XVII) give the earliest dates of the most frequent prefixal comparatives attested in Russian as follows:

поближе '(slightly) nearer', *полегче* 'lighter', *получше* 'better', *помоложе* 'younger', *поранее (пораньше)* 'earlier', *постарее (постарше)* 'older', *почаще* 'more often', *пошире* 'broader' – 16th century; *покороче* 'shorter', *покрепче* 'stronger', *поскорее* 'faster', *похуже* 'worse' — 17th century.

The material of the RNC shows that the form is mainly used in the texts with more popular characteristics (not in the livresque Church Slavonic style), including the Pskov Chronicles of the 15-16th centuries, the routine regulations like the celebrated 16th century Domostroy and

private letters of the 17th century. Some forms attested in the RNC are absent from the Dictionary of the 11-17th centuries Russian, like *посильнѣе* 'stronger' (1578) and *пвѣдмос(т)нѣе* 'in more detail' (1640s).

In Modern Russian, the prefixed comparative, as well as the non-prefixed one, is formed both from adjectives and adverbs and can be used attributively, adverbially and predicatively. It is known that the attributive function is even particularly typical for these forms (see below). In Old Russian these forms initially were only adverbs. The historical and synchronical descriptions do not give a hint when and how the prefixed forms started to be used attributively. The RNC gives the first examples of the attributive attenuatives only in the 16th century: *которые помоложе — тѣ наперед, а которые почестнѣе — тѣ опослѣ* 'those who are younger come first, and those who are nobler come after' [Wedding Regulation]. It is also interesting that Middle-Bulgarian-like declined adjectives with *по-* and comparative morphology existed in Middle and Early Modern Russian as well, exemplified by *двѣ крушки поболшиих* 'two bigger cups' (Словарь XI—XVII, 1697). These forms continue to exist marginally even in later (Early) Modern Russian texts:

(1) *и ты, старый Волх, и ты, постарший Словен, и ты, младший Хорев, и ты, помладший Кощей...* [А. Ф. Вельтман. Кощей бессмертный (1833)]
'and you, the old Volkh, and you, Sloven, who is (yet) elder, and you, Khorev, who is younger, and you, Koschei, who is (yet) younger'

(2) *И никакие Адам и Ева с яблоком и даже со змеем так во мне добра не предрешили, как мальчик — с другим мальчиком, поменьший с побольшим , гадкий — с хорошим, земляничный — с заоблачным.* [М. И. Цветаева. Черт (1935)]
'No Adam and Eve with their apple and even with the serpent predefined the good in me in such a degree as this boy with another boy, the elder one with the younger one, the nasty one with the good one, the strawberry one with the empyreal one'

3   The Attenuative Meaning in Diachrony

The corpora data are helpful in investigating the semantics of the form, viz. the rise and development of the attenuative meaning: 'X is more than Y, and slightly more' (Богуславский & Иомдин 2009). The research has previously indicated that these forms are marked by "subjective" modality rather than by an objective evaluation of difference (Князев 2007: 198).

The meaning of attenuative seems to be linked with the etymology and the delimitative semantics of the preposition *по*. Andrei Zalizniak (p. c.) thinks that the comparatives like *глубже* 'deeper' per se could signify both "a bit deeper" and "much deeper", while *поглубже* meant 'until the limit than could already be called "deeper", i.e. the minimal difference. However, the first Old Russian texts exhibit no traits of this meaning, cf. *10 поприщь подале* 'ten *poprisches* (leagues/miles etc.) further' (Troitsky miscellany, XIV-XV, (Крысько 2006: 371)). There is corpus evidence that the attenuative semantics was not finally grammaticalized until the end of the 19th century.

Russian degree adverbs vary greatly by their compatibility with different degrees of comparison. There are adverbs that can be used only with the adverbs and adjectives of the positive degree of comparison: *очень* 'very', *страшно* 'terribly', *исключительно*

'exceptionally'; the other ones that can be used only with comparatives: *гораздо* 'much', *вдвое* 'two times' and yet the third class that can be used with both, like *немного* 'slightly' (Плотникова 1980: 704). It turns out that the 19th-century Russian widely used the combinations of prefixed comparatives with adverbs of the second type: *гораздо, (на)много,* including the multiplicational adverbs like *вдвое, втрое, вчетверо* 'two, three, four times' and adverbials like *в два, три, четыре раза* (idem). The papers (Богуславский & Иомдин 2009) and (Князев 2010) mark these examples with an asterisk as impossible, and this impossibility is, for these authors, the proof of the fact that attenuative semantics is inherent for the prefixed comparative. Boguslavsky and Iomdin even analyze a counter-example by Mamin-Sibiryak – *в десять раз повернее* 'ten times more certain' – but discard it as an example of language play.

The situation is historically trickier. The combination of the type *гораздо побольше* loses frequency (from 4 to 0.15 ipm) only in the last quarter of the 19th century. Some examples of this kind are attested even in modern texts, although many speakers (as well as linguists) consider them ungrammatical:

(3) *Пелагея, подай стакан пуншу, да гораздо покрепче.* [А. И. Герцен. Кто виноват? (1841-1846)]
'Pelageya, give me a glass of punch, and <u>much stronger</u>'

(4) *Собственный пароходик у него будет, «Батрак», вдвое почище да и побольше вот этой посудины.* [П. Д. Боборыкин. Василий Теркин (1892)]
'His own steamboat *Batrak* is perhaps <u>twice as clean and twice as big</u> as this very vessel'

It is interesting that since the 1980s the prefixed comparative began going together with the adverb *сильно* lit. 'strongly' which can combine, colloquially, with positive adjectives as well (*сильно бедный* 'very poor'). It seems that this possibility is due to the subjectiveness of its semantics, which is linked with qualitative distinction rather than quantification:

(5) *КПРФ — единственная партия, которая имеет среди источников дохода, кроме поддержки крупного бизнеса (она тоже таковую имеет) пожертвования частных лиц, простых людей. Отколупливают от своих нищенских зарплат и шлют папе Зю. Те, которые чуть посостоятельнее, но сильно победнее бизнесменов, снова печальным журавлиным клином двинулись на Запад.* (Валерий Лебедев. Отечество в опасности (2003) // «Лебедь» (Бостон), 2003.11.01)
'The Communist Party of Russia is the only party that has among its sources of income, alongside with the support of business elites (yes, it does also have it), contributions of private individuals, of the common people. They pinch the money off their miserable salaries and send it to Father Zyuganov. Those who are a bit wealthier than them, but still <u>considerably poorer</u> than the businessmen, have again headed to the West like a sad flock of cranes'.

In Modern Russian the prefixed comparative goes even better with the expressive *куда* 'by far', lit. 'where', which is also put under an asterisk in (Богуславский & Иомдин 2009):

(6) *Мы встретились с тобой куда пораньше, а ну-ка вспоминай!* (Эдвард Радзинский. Продолжение Дон Жуана (1990-2000))
'I (we) met you <u>much earlier</u>, hey, try to recall!'

This fact is perhaps related to the well-known non-comparative absolutive uses of the prefixed form (Кустова 2002; Князев 2007: 199; Богуславский & Иомдин 2009). The meaning of the marker "depends on the context and the desire of the speaker, expressing quantity and intensiveness varying from 'a little bit more' to 'in the maximal possible degree' (Guiraud-Weber 1996: 489) (cf. (Camus 1994: 122-129), (Guiraud-Weber 1995: 93, 96-97)).

4   Lexical Preferences of the Prefixed Comparative

The corpus can help us to analyze the general development tendencies of the prefixed comparative and its interactions with the lexicon. The prefixed form is much more strongly lexicalized than the form without a prefix. The first four forms of the type-frequency list – *побольше* 'more, bigger', *поближе* 'nearer, closer', *подальше* 'further', *поскорее* 'faster' – yield 35% of tokens of this form in all the texts. (We may note that these lexemes were among the first prefixed comparatives chronologically). The first twenty forms of the frequency list occupy not less than 67% tokens in the text. This is twice as much as both parameters for the non-prefixed form (18% and 34% respectively).

Which lexemes prefer the prefixed comparative? To answer this question we should find the percentage of the prefixed forms among all the comparative forms. The top ranks list is headed by the form *(по)хлеще* lit. 'more trenching, biting' with 78% of prefixed forms over 22% of the simple form *хлеще*. The fact is that this prefixed form is lexicalized with another, much more general meaning '(expressed) in a more sheer fashion', 'more important and impressing', with the non-prefixal form rarely conveying this sense (usually coupled with the adverb *еще* 'yet'). The same holds for *(по)чище* lit. 'more clear' that is also used in the same contexts as *похлеще*.

Among the comparatives with more than 25% of prefixed forms a semantic group can be discerned: adjectives or adverbs signifying (mainly positive) human qualities, consisting of multiple reactions to multiple stimuli. The forms are the following:

> *(по)аккуратнее* 'more accurate'
> *(по)бойчее* 'more smart, quick'
> *(по)внимательнее* 'more attentive, careful'
> *(по)ловчее* 'more dexterous, cunning'
> *(по)приличнее* 'more decent'
> *(по)пристальнее* 'more fixed, intent'
> *(по)смышлёнее* 'more clever'
> *(по)хитрее* 'more cunning, intricate'

These properties are defined on a continuous non-discreet scale and thus welcome the meaning of attenuative; at the same time, their semantics couples well with modalization – in the so-called selective contexts (*хотелось бы, чтобы ты действовал похитрее* 'I would like it if you were more cunning' / *чтобы мы нашли кого- нибудь побойчее* '...if we found somebody more smart'). Both parameters go well together; e. g. a piece of advice or a reproach may be, due to politeness, expressed with means of attenuative.

It is known that the lower-degree markers are combined well with negative (pejortaive) adjectives in the positive degree, which is explained by politeness requirements. This fact is, perhaps, cross-linguistical. In Russian the diminutive suffix *-оват-* combines with 'bad' and

'silly' (*плоховатый* and *глуповатый*) but is reluctant with 'good' and 'intelligent' (**хорошеватый* and **умноватый*) (Kagan & Alexeyenko 2011: 322). In English we have corpus data saying that the low-degree markers like *slightly* and *a bit* select predominantly negative adjectives in the positive degree (Bylinina & Zadorozhny 2012). It is thus very interesting that this effect is inverse with comparative degree as compared to the positive one, and this issue is open to a further study.

Bibliography

Boguslavsky, I. & L. Iomdin.. 2009. Semantics of Attenuated Comparatives in Russian. In *Proceedings of the 4th International Conference on Meaning — Text Theory. Observatoire de linguistique Sens – Texte (OLST)*, 65—77.

Bylinina, L. & S. Zadorozhny. 2012. Evaluative adjectives, scale structure, and ways of being polite. // M. Aloni, V. Kimmelman et al. (eds.) *Logic, Language and Meaning: 18th Amsterdam Colloquium*, 133—142, Amsterdam: Springer.

Camus R. 1994. *Contribution à l'étude du mot du discours da en russe contemporain*. Thèse pour le doctorat. Paris, Université Paris 7.

Guiraud-Weber M. 1995. Le comparatif russe et ses particularités // *Faits de langues*, № 5.

Guiraud-Weber M. 1996. La syntaxe du comparatif russe : quelques précisions // *Revue des études slaves, Tome 68, fascicule 4*.

Kagan, O. & S. Alexeyenko. 2011. Degree modification in Russian morphology: The case of the suffix *-ovat* // I. Reich et al. (eds.) *Proceedings of Sinn & Bedeutung* 15, 321—335. Saarbrücken: Universaar — Saarland University Press.

Vaillant, A. 1958. *Grammaire Comparée des Langues Slaves*. Tome II: Morphologie. Paris.

Богуславский, И. М. & Л. Л. Иомдин 2009. Семантика смягченной сравнительности: русские компаративы на по– // *Von grammatischen Kategorien und sprachlichen Weltbildern: Die Slavia von der Sprachgeschichte bis zur Politsprache. Festschrift für Daniel Weiss zum 60. Geburststag* (Wiener slawistischer Almanach. Sonderband 73). München, Wien, 2009.

Гиппиус, А. А. 1994. Древнерусские наречия на базе компаратива // *Problemi di morfosintassi delle lingue slave*. 4. Padova, 33—34.

Князев, Ю. П. 2007. *Грамматическая семантика*. М.: ЯСК.

Князев, Ю. П. 2010. Префиксальный компаратив в современном русском языке: синтаксис и семантика // *Международная конференция «Русский язык и литература в международном образовательном пространстве: современное состояние и перспективы: к 55-летию преподавания русского языка в Испании». Доклады и сообщения*. Т. 1. Гранада, 130—135.

Крысько, В. Б. 2006. Наречия, образованные из предложных сочетаний с формой ВП ед. числа сред. рода сравнительной степени // В. Б. Крысько (ред.) *Историческая грамматика древнерусского языка. Т. III. Прилагательные*, 359—374, М.: Азбуковник.

Кустова, Г. И. 2002. О семантике компаратива (абсолютивное употребление) // *Русистика на пороге XXI века: проблемы и перспективы*, 113—120, М.: ИРЯ РАН.

Мельчук, И. А. 1998. *Курс общей морфологии*. Т. II. Вена – Москва, ЯРК.

Плотникова, В. А. 1980. Наречие // Н. Ю. Шведова (ред.). *Русская грамматика.* Т. 1, 703 — 705, М.: Наука.

Словарь XI—XVII. *Словарь русского языка XI—XVII веков*. М., 1975

Towards Machine Translation of Russian Aspect

Barbara Sonnenhauser (1) and Robert Zangenfeind (2)

(1) Institute for Slavic Philology
Ludwig-Maximilians-Universität München Geschwister-Scholl-Platz 1, 80539 München
basonne@lmu.de
(2) Center for Information and Language Processing
Ludwig-Maximilians-Universität München Oettingenstr. 67, 80538 München
r.zangenfeind@lmu.de

Abstract

One of the most intricate problems for machine translation is grammatical categories which are present in the source language but not in the target language. This problem is further complicated if the category in question is highly polysemous. Grammatical aspect in Russian is one example of such a category. On the categorical level, it has only approximate equivalents in non-Slavic languages (such as the progressive form in English). In addition, language-internally, its semantics and interpretation cannot be sufficiently captured with one specific characteristic feature. This paper aims at establishing a basis for the machine translation of the Russian aspect. To do so, the interaction of verb and aspect semantics has to be described in a systematic way. Moreover, the further lexical components contributing to the meaning computation have to be annotated for the aspectual information they provide. This allows for the formulation of rules for machine translation into target languages either lacking grammatical aspect or having a different aspect system.

Keywords

Grammatical aspect, machine translation, ambiguity, semantic features, syntactic features, Russian, German, English, Turkish

1   The Problem of Aspect

Grammatical aspect in Russian is problematic for machine translation in at least two respects: 1) it is a highly polysemous category, 2) it does not have categorical equivalents in some languages, and if it does, the equivalence is hardly one to one.

1.1 Polysemy and Ambiguity

Traditionally, the interpretations that are possible for both aspects are given in the form of lists (e.g. Zaliznjak & Šmelev, 1997: 15-36). This is definitely useful for didactic intentions; other purposes, however, require a different systematization. One case in point is translation, especially machine translation from Russian into other languages.[1]

One possible way of systematizing aspect interpretations is provided by the analysis developed in Sonnenhauser (2004, 2006), based on the combination of a selection-theoretic (Bickel, 1996) and time-relational (Klein, 1995) account. According to this analysis, aspect operators select, and thereby assert, specific part(s) of the event structure encoded by the verb. Assuming a tripartite event-structure (Moens & Steedman, 1988), verbs may encode dynamic phases 'φ_{dyn}' (preparatory processes), boundaries 'τ' (culmination points) and static phases 'φ_{stat}' (consequent states), depending on the eventuality they refer to. By selecting and asserting some part of the coded event structure, aspect establishes a relation between the topic time interval I(TT) as the time the assertion is about and the event time interval I(e) comprising that part of the run time of the denoted event that is selected by the aspect operator. This yields the relations given in (1) and (2), i.e. the inclusion of the boundary in the topic time for the perfective (= PF) aspect and a general overlap relation for the imperfective (= IPF) aspect (a more detailed account is provided in Sonnenhauser, 2006, 2009):

(1) PF $I(TT) \supset I(\tau)$

(2) IPF $I(TT) \bigcirc I(e)$

Both relations are specified in the course of interpretation. For the PF aspect, this mainly concerns the specification of the boundaries of I(TT): the interval may be closed to both sides, i.e. the initial and final points are part of the interval, it may be open to the right or open to the left, i.e. the initial point is part of the interval whereas the final point is excluded and vice versa.[2] This is illustrated with the example in (3a), which can be interpreted in three ways and thus be translated into English as in (3b-d):

(3) a. Ja emu *dala* knigu.
    b. I *gave* him the book [and then...]   I(TT) closed
    c. I *have given* him the book [and now...]   I(TT) open to the right
    d. [After] I *had given* him the book   I(TT) open to the left

---

[1] The other direction of machine translation again has its own specific requirements, cf Mel'čuk & Wanner (2008).

[2] This is based on a set theoretic definition of intervals as proposed, e.g., by Guentchéva & Desclés (1982), cf. also Sonnenhauser (2006, 115-118). According to this conception, intervals consist of sets of points. Unbounded intervals are not segmented, bounded intervals are segmented. With open-bounded intervals I_o the initial (a) and the final point (b) out of the set of points (x) are not part of the interval, i.e. $I_o = \{x, a < x < b\}$, with closed-bounded intervals I_c both are part of the interval, i.e. $I_c = \{x, a < x < b\}$.

The basic relation for the IPF aspect is much more general, since contrary to the PF aspect it is not restricted as regards the selected part of the event – this may be a phase or the event as a whole. Closer inspection reveals three specific relations, which are given in (4), with one example each illustrating the relations:

(4) a. $I(TT) \subset I(\varphi_{dyn})$
Kogda on vošel, ona *čitala* knigu.
'When he came in, she *was reading* a book.'
($I(\varphi_{dyn})$: the time interval of her reading the book, covering only this process excluding beginning or end; $I(TT)$ is included in the reading-process and specified by the moment when he came in)

b. $I(TT) = I(e)$
Ona *rabotala* v universitete.
'She *worked at* the university.' [= She was employed there.]'
($I(e)$: the time interval when she was employed at the university; $I(TT)$ runs exactly parallel to the time interval of her working at the university)

c. $I(TT) \supset I(e)$
Ona uže *rasskazyvala* emu ėtu istoriju.
'She *has* already *told him* this story.'
($I(e)$: the time interval of her telling the story; $I(TT)$ includes the complete story-telling event)

The outlined analysis in terms of specific relations assumes them to be clearly distinct, which in turn suggests ambiguity. Each of these distinct relations may give rise to a specific range of interpretations. For the purposes of machine translation, only the factor of ambiguity is decisive; both the structures underlying the representations and the specific interpretations can be neglected.

Having proposed a solution for the systematization of the manifold interpretations possible for the Russian PF and IPF aspect, the second problem can be addressed: the cross-linguistic similarities and divergences.

1.2 Language Comparison

The justification for postulating the three specifications for the PF aspect is provided not only on language-internal grounds, but also by the fact that these relations can be morphologically coded in other languages, which render it mainly in terms of temporal distinctions. Table 1 illustrates this for Russian, English and German, with the brackets indicating the boundedness-characteristics of the intervals. Note that these correlations hold for the past tense; with morphological present in Russian, the relation "$I(TT)$ closed" yields a future interpretation.

Semantics	Interpretation	Russian	English	German
group I$_{PF}$ TT closed: [---τ---]	eventive	PF	simple past	imperfect / perfect[3]
group II$_{PF}$ TT right open: [---τ---[	perfect (existential, current relevance, extended now, etc.)	PF	perfect	perfect
group III$_{PF}$ TT left open:]--- τ---]	pluperfect	PF	pluperfect	pluperfect

Table 1: Ambiguity of pf aspect

Likewise, the cross-linguistic validity of assuming three basic IPF configurations is suggested by two facts: the three configurations may be coded morphologically in other languages in terms of aspect distinctions, and if coded, they give rise to a similar range of interpretations. This is illustrated in table 2, comparing 'imperfective' grammemes in Russian, English and Turkish (for more details cf. Sonnenhauser 2006).[4] This indicates that even though aspect is grammaticalized in all three languages, they are by no means equivalent as regards the semantic range of the respective grammemes.

Semantics	Interpretation	Russian	English	Turkish
group I$_{IPF}$ TT $\subset \varphi_{dyn}$	processual, conative	IPF	progressive	-iyordu -mekteydi
group II$_{IPF}$ TT = e	habitual, non-actual, potential, permanent, atemporal	IPF	simple form	-irdi
group III$_{IPF}$ TT $\supset$ e	general-factive, durative	IPF	simple form	-di

Table 2: Ambiguity of IPF aspect

---

[3] The distinction between imperfect and perfect is getting blurred in German. It is intact in Northern varieties but has been completely lost in Southern varieties, where the perfect has taken over narrative functions.

[4] The comparison in table 2 is confined to the past, since group III$_{IPF}$ is not possible for the other tenses. Accordingly, the Turkish forms are specified with the past tense morpheme *-di*. Note that being opposed to *-iordu* and *-irdi*, simple *-di* can be interpreted as the morphologically and semantically unmarked element in the past tense aspect system of Turkish (cf. Sonnenhauser 2006).

The ambiguity of Russian aspects and the cross-linguistic validity of the possible disambiguated configurations are crucial for the question of machine translation in that this provides the basis for stating clearly formulated rules.

1.3 Disambiguation

Having pointed out the advantages of assuming a basic ambiguity of the semantics of aspect as regards Russian and the cross-linguistic perspective, the next question we address is that of disambiguation. In natural language communication, interpreting an utterance requires the resolution of the aspect-ambiguity; disambiguation is also the first step towards machine translation.

Disambiguation is achieved by specifying I(TT) in terms of its boundedness-features and – for the IPF aspect – by specifying the relevant part of the Aktionsart that is selected and related to this interval. In Russian, this specification is possible mainly by lexical and syntactic means: as regards the IPF aspect, adverbs like *medlenno* 'slowly' or *postepenno* 'gradually' specify I(TT) as open-bounded, adverbs like *ran'še* 'formerly' as unbounded, particles like *uže* 'already' as closed-bounded, and hence the interpretation as belonging to group I$_{IPF}$, II$_{IPF}$, or III$_{IPF}$ respectively. Concerning the PF aspect, conjunctions like *i* 'and [then]' disambiguate eventive (group I$_{PF}$) from perfect (group II$_{PF}$) interpretations, adverbials specifying a point in time suggest the pluperfect interpretation (group III$_{PF}$), etc. This is due to the fact that here the consequent state following the selected boundary does not hold at the time of utterance (which would yield the perfect interpretation) but at the time specified by the temporal adverbial, which is prior to the time of utterance.

As can be seen from tables 1 and 2, for machine translation from Russian to English, German or Turkish it is enough to solve these basic ambiguities. What is rendered by means of the perfect in English or German has the same interpretational range as the 'perfect' / group II$_{PF}$ specification of the Russian PF aspect, what is rendered by means of the *-irdi* suffix in Turkish may give rise to the same variety of interpretations as group II$_{IPF}$ of the Russian IPF aspect. The same reasoning applies to the other ambiguities.

For an automatic disambiguation, the relevant lexical and syntactic means have to be annotated in the lexical entries of lexemes as regards the aspectual information they contribute to the meaning computation. The computation may then proceed in the form of 'if-then' statements along the lines proposed by Vazov (1999), which is also used by Mel'čuk & Wanner (2008) for aspect-establishing rules in the process of German-Russian translation.

One way to provide the necessary kind of information for the relevant lexemes can be to annotate these lexemes with appropriate semantic features. ETAP-3 (cf Section 2) is a machine translation system that uses dictionaries with semantic and syntactic features.

2 Aspectually Relevant Semantic and Syntactic Features in the Dictionaries of ETAP

The machine translation system ETAP-3 (for an earlier version cf. Apresjan et al. 1992) provides a lot of information for lexemes that can be useful for the interpretation of aspect.

Until now, this information was mainly used for syntactic analysis, and it is given in the dictionaries of ETAP in terms of semantic and syntactic features. Some of the semantic features that are potentially relevant for the interpretation of aspect are 'VREMJA' (to characterize temporal lexemes[5]), 'DEJSTVIE' (for nouns and verbs that denote an action which develops in time and which is initiated by an active subject), 'PROCESS' (for nouns and verbs that denote a process which develops in time and which is initiated by a passive subject) etc. Some important syntactic features for the interpretation of aspect are 'DLIT' which characterizes a period of time or 'NEOPR' for indefinite pronouns, to name just two of them.

Another valuable instrument for our purpose is the classification of predicates by Apresjan (2006). This classification includes 17 classes. Some of them exclude certain disambiguation possibilities and/or make others highly probable. For "dejatel'nosti" ('activities')[6] such as *torgovat'* 'to trade', *upravljat'* 'to rule', for instance, the actual-processual and the general-factual readings are ruled out, whereas a durative interpretation is most likely. For these verbs, therefore, the information 'group II_{ipf}' can already be assigned in its lexical (semantic) information. Other classes, such as "dejstvija", i.e. 'actions' are less explicit and allow for all possible interpretations. For their disambiguation, further information provided by aspectually relevant components of the sentence has to be taken into account.

These relevant components are realized by adverbials, particles and conjunctions.[7] So, besides information about aspect and class of the predicate in a sentence, aspectually relevant information must be provided in the form of semantic (and syntactic) features in the lexical entries of these parts of speech. Another crucial bit of information is provided by tense. Present tense, for instance, excludes ipf interpretations out of group III_{ipf} and all pf interpretations except for the future interpretation (cf section 1.2). The combination of all this kind of information can be the basis for the "calculation" of a temporal and aspectual interpretation of the whole sentence.[8] The next section will show the problems of such a calculation and steps towards a possible solution.

3 Towards a Solution

An example to illustrate which information in a sentence is relevant is given in (5):

---

[5] The descriptions of this and the following features are corresponding to the Russian help manual for ETAP-3 in version 3.1.91 from the year 2008. This is the part of ETAP-3 system which we are very grateful that Leonid L. Iomdin back then placed at our disposal at the Center for Information and Language Processing in Munich.

[6] The English terms for classes of predicates are taken from Apresjan (2005).

[7] These components correspond to the contextual clues (imperfective and perfective triggers) of Mel'čuk & Wanner (2008).

[8] This is by no means to say that there have not been any compositional approaches to aspect before. These approaches (for a basic overview cf. Verkyul 2012) are concerned with the modeling of aspectual composition in order to arrive at a principled syntactic description and the mapping of composition onto syntactic structure, not with possible implementations into NLP. Moreover, they mostly lack a distinction between lexical information and aspect semantics.

(5) Ran'še ja po večeram prodelyval èti gimnastičeskie upražnenja po pjat' raz.[9]
   lit. 'formerly I in evenings do.PAST.ipf these gymnastic exercises each five times'

Lexemes and phrases that are important for our interpretation are the following: *ran'še* 'formerly', *po večeram* 'in the evenings', *prodelyvat'* '[to] do', *upražnenie* 'exercise' and *po pjat' raz* 'five times each'.

The dictionary entries of ETAP provide the following information about these lexemes amongst other characteristics[10]:

- *ran'še* 'formerly' has the syntactic feature 'VREM' which characterizes temporal adverbs;

- *večer* 'evening' has the syntactic features 'DLIT' (to characterize a period of time) and 'VREM' (here to characterize a point of time or a period of time) as well as the semantic feature 'VREMJA' (temporal lexeme);

- *prodelyvat'* '[to] do' has the semantic features 'FAKT' (event) and 'DEJSTVIE' (action, i.e. a situation which develops in time and which is initiated by an active subject);

- *upražnenie* 'exercise' has the semantic features 'FAKT' and 'DEJSTVIE', cf above. For our purposes, this information should be enriched by the following:

- *ran'še* 'formerly' is temporally and referentially (as concerns reference to the event) indefinite and thus excludes group I_{ipf} interpretations; appropriate semantic features in addition to ETAP's features could be 'temporally indefinite' and 'referentially indefinite'[11];

- *po [večeram]* 'in [the evenings]' is a preposition that – when governing a temporal lexeme, i.e *po16*[12] – expresses regularity. An adverbial phrase *like po večeram* 'in the evenings' can be annotated by labeling the preposition *po16* with the feature 'regularity', and thus excludes group I_{ipf} and group III_{ipf} interpretations;

---

[9] Example from Bendixen et al. (2005-2012).

[10] We only cite here the semantic and syntactic features that seem to be relevant for aspect interpretation.

[11] The semantic feature "temporally indefinite" should indicate that there is just a vague temporal specification in terms of localization on the time axis; this feature should also characterize adverbs like *skoro* 'soon', *togda* 'then', *vsegda* 'always' etc. (the semantic feature "temporally definite", on the other hand, should indicate a more precise temporal specification; e.g. for adverbs like *teper'* 'now', *segodnja* 'today' etc.). The lists of adverbs with these and other semantic features, of course, still must be thoroughly examined (the need for a list of such triggers is pointed out also by Mel'čuk & Wanner 2008: 141). "Referentially indefinite" concerns the selection and assertion of a specific part of the event structure carried out by aspect (cf. section 1.1): adverbs like *ran'še* indicate that there is no specific part of the event structure selected by aspect (contrary to 'group I_{ipf}' interpretations, where the dynamic phase is selected by the ipf aspect and asserted to hold within I(TT)). Particles like e.g. *uže* 'already', on the other hand, should be annotated with the features "referentially definite" and at the same time "temporally indefinite".

[12] cf. Slovar' russkogo jazyka 1983.

- *prodelyvat'* '[to] do' is used as a support verb; i.e. it has no semantics, only aspectual information is relevant (here: ipf); semantic information must be provided by the predicative noun in the sentence (here: *upražnenie* 'exercise');

- *upražnenie* 'exercise' is the semantic predicate in the sentence and can be labeled as "zanjatie" ("occupation") according to Apresjan (2006: 83, 86f), i.e. an action whose immediate object is just to accomplish this action[13]; in combination with an ipf support verb such as *prodelyvat'*, it allows for group I_{ipf}, II_{ipf} and III_{ipf} interpretations;

- *po [пять раз]* '[five times] each' is a preposition that – when governing a noun that can have a numeral as syntactic dependent, i.e. *po20*[14] – expresses distributivity of the verbal complement and allows for group I_{ipf}, II_{ipf}, III_{ipf} interpretations[15]; the appropriate feature for the preposition *po20* could be 'distributive'.

Based on the newly added information, the aspectual information given in (5) can be disambiguated as belonging to group II_{ipf}. This is calculated as follows: The predication *prodelyvat' ėti upražnenija po pjat' raz* 'do.ipf these exercises five times each' is ambiguous between all three groups. This range of possibilities becomes restricted by the contribution of past tense and the adverbials *po večeram* 'in the evenings' and *ran'še* 'formerly'. The decisive information is provided by *po večeram,* which excludes two of the three possible specifications and thus overrides the less specific information given by *ran'še,* which excludes only one specification. Based on this specification, the verbal lexeme should be rendered by the simple form in the English translation.[16]

To sum up, this interpretation could be formalized as conditional instructions ("if-then") in the following way:

(6) for language-internal disambiguation:
IF predicate has feature "occupation"
AND IF aspect = ipf
AND IF tense = past
AND IF adverb 'group II_{ipf}'
THEN 'group II_{ipf}' interpretation

---

[13] This means that the information from ETAP (action) for *upražnenie* 'exercise' can be further specified by the feature "occupation".

[14] cf Slovar' russkogo jazyka 1983.

[15] Cf. Mehlig (2008) on the hybrid nature of distributive predications. These predications may be ipf and receive an actual-processual interpretation even though their complement is bounded, as is *upražnenija* by *po pjat' raz* in (5). Note the crucial role of the secondary imperfective *prodelyvat'* here. As Filip (2008: 247) points out, "[predications with secondary imperfectives [...] have sets of partially ordered events in their denotation, due to the contribution of the prefix, but the imperfective suffix on the verb explicitly suspends the requirement that the verb only has maximal events in its denotation [...]".

[16] The most adequate translation would be with the habitual construction "used to". This specification can be solved by means of language-internal paraphrasing rules and is not necessarily an immediate concern of translation.

(7) for translation:
     IF 'group II$_{ipf}$' interpretation
     THEN 'simple form' in English

Formal descriptions like these can be the basis for an implementation in a machine translation system like ETAP.[17]

4   Conclusion

We have argued that, based on the combination of a selection-theoretic and time-relational account, it is possible to systematize the semantics of the Russian verbal aspect and its interpretations. This systematization comprises several groups specifying the relation between topic time interval and event time interval possible for the pf and ipf aspect. These groups may have morphological counterparts in the tense-aspect systems of other languages. In order to choose the right morphological means when translating the Russian verbal aspect it is necessary to disambiguate its semantics. Disambiguation is made possible by annotating all relevant lexemes with specific, aspectually relevant information. This is the starting point for a possible computational implementation of aspect interpretation. The system of semantic and syntactic features as used by the machine translation system ETAP is a workable basis for this implementation. Enriching this system with information taken from Apresjan's classification of predicates and with additional, more detailed semantic features, we illustrated the problems of "calculation" of aspect interpretation and presented steps towards a possible solution.

Our future work will be to develop a refined system of semantic features for verbs (and predicative nouns), adverbials, particles and conjunctions, based on ETAP's features and Apresjan's classification of predicates. With these tools at hand, it is our aim to implement the rules for aspect translation in a machine translation system like ETAP. Besides the practical utility, an implementation in a rule based system has the great virtue to verify the linguistic theory in practice and, with that, to enable to improve the theory.

Bibliography

Apresjan, Ju.D. 2005. Prolegomena to systemic lexicography. In Apresjan, Ju.D. & L.L. Iomdin (eds.). *East West encounter: second international conference on Meaning ↔ Text Theory*. Moscow, 20-29.

Apresjan, Ju.D. 2006. Fundamental'naja klassifikacija predikatov. In Apresjan, Ju.D. (ed.). *Jazykovaja kartina mira i sistemnaja leksikografija*. Moskva, 75-110.

Apresjan, Ju.D. et al. 1992. *Lingvističeskij processor dlja složnych informacionnych sistem*. Moskva.

---

[17] An implementation of rules like these in ETAP's already much elaborated translation from Russian to English could be a first step. But, of course, similar rules can be established for the translation into other languages. For a fragment of the translation from Russian to German as an extension of ETAP, cf Zangenfeind (2011).

Bendixen, B. et al. 2005-2012. *Russisch aktuell.* Wiesbaden.

Bickel, B. 1996. *Aspect, mood and time in Belhare.* Zürich.

Filip, H. 2008. Events and maximalization: the case of telicity and perfectivity. In Rothstein, S. (ed.). *Theoretical and crosslinguistic approaches to the semantics of aspect.* Amsterdam, Philadelphia, 217-256.

Guentchéva, Z. & J.-P. Desclés. 1982. A la recherché d'une valeur fondamentale du parfait bulgare, *Sapostavitelno ezikoznanie* 7/1-2, 44-56.

Klein, W. 1995. A time-relational analysis of Russian aspect, *Language* 71(4): 669-695.

Mehlig, H.-R. 2008. Aspect and bounded quantity complements in Russian. In Rothstein, S. (ed.). *Theoretical and crosslinguistic approaches to the semantics of aspect.* Amsterdam, Philadelphia, 257-290.

Mel'čuk, I.A. & L. Wanner. 2008: Morphological Mismatches in Machine Translation. In *Machine Translation,* 22: 101-152.

Moens, M. & M. Steedman. 1988. Temporal ontology and temporal reference, *Computational Linguistics* 14(2): 15-28.

Slovar' russkogo jazyka 1983. Tom III, Moskva.

Sonnenhauser, B. 2004. Underspecification of 'meaning': the case of Russian imperfective aspect. In *Proceedings of the ACL-04 workshop on "Text, Meaning and Interpretation".* Barcelona, 89-96.

Sonnenhauser, B. 2006. *Yet there's method in it. Semantics, pragmatics, and the interpretation of the Russian imperfective aspect.* München.

Sonnenhauser, B. 2009. Definiteness and specificity of verbal referents. In Birzer, S., Finkelstein, M. & Mendoza, I. (eds.). *Proceedings of the second international Perspectives on Slavistics conference (Regensburg 2006).* München, 115-126.

Vazov, N. 1999. Context-scanning strategy in temporal reasoning, *Modeling and using context. Lecture notes in computer science* 1688: 389-402.

Verkuyl, H. 2012. Compositionality. In Binnick, R. (ed.). 2012. *The Oxford handbook of tense and aspect.* Oxford, 563-585.

Zaliznjak, A.A. & A.D. Šmelev. 1997. *Lekcii po russkoj aspektologii.* München.

Zangenfeind, R. 2011. Transfer of Russian actantial syntactic relations into German, In Boguslavsky, I. & L. Wanner (eds.). *Meaning – Text Theory 2011.* Barcelona, 306-315.

Composites Denoting Nomina Agentis in the Russian Language: Distinguishing Competing Models

Maria Tagabileva

Moscow State University & Higher School of Economics
Moscow
geratagabileva@gmail.com

Abstract

One of the largest blind spots in the studies on Russian word-formation is the problem of interaction of fully or partially synonymic models, their rivalry and distribution both in the vocabulary and in spontaneous speech. We aimed this particular study to describe two competing models used to form composite nomina agents: the zero suffix model and the model with the suffix -ec. In our research we investigated possible constraints imposed on the models in question as well as morpho(no)logical and syntactical features of their derivatives. Trying to determine the factors that influence the choice of one of the competing models we also took into account their productivity and peculiarities of their realization in texts of different genres, styles and time periods. It turned out that, despite the clear tendency of one of the models under investigation to be extruded by the other, there are several factors that do not let this process to be completed, including the following:

1. Morphonological constraints on the zero suffix model, which do not let all the derivatives change their word formation model;

2. Stylistic differentiation of the derivatives, which mostly results from the history of development of the models in question.

Keywords

Morphology, morphonology, derivation, word formation, synonyms, diachrony

1   Introduction

One of the largest blind spots in the studies on Russian word-formation is the problem of interaction of fully or partially synonymic models, their rivalry and distribution both in the vocabulary and in spontaneous speech. We aimed this particular study to describe two competing models used to form composite nomina agentis:

1.   Zero suffix model: "base$_1$ + interfix o/e+base$_2$ + zero suffix" (e.g. *charodey* 'wizard', *zhenol'ub* 'ladies' man', *konokrad* 'horse thief', *zverolov* 'hunter');

2.   Model with suffix *-ec*: "base$_1$ + interfix o/e + base$_2$ + suffix *-ec*" (e.g. *chudotvorec* 'wonder worker', *strastoterpec* 'passion-bearer', *znamenosec* 'flag-bearer'), where base$_2$ is necessarily a verb stem.

In our research we investigated the possible constraints imposed on the models in question as well as morpho(no)logical and syntactical features of their derivatives. Trying to determine the factors that influence the choice of one of the competing models we also took into account their productivity and peculiarities of their realization in texts of different genres, styles and time periods.

2   Syntactic Properties

The first thing we turned our attention to were the categorial and the syntactic properties of motivating bases. As it has already been mentioned in (Sitchinava, 1999), composites with a second verb stem are more easily derived from transitive verbs. This distinction is particularly relevant for composite nomina agentis. It is not only that there are significantly more transitive verb stems in the overall list of bases, but such stems tend to have more derivatives on average, as is clear from the data below:

	overall	intransitive	transitive
-ec	32 (100%)	4 (12,5%)	28 (87,5%)
-0	82 (100%)	10 (12,1%)	72 (87,9%)

Table 1. Motivating stems: trans. vs intrans.

	overall	intransitive	transitive
-ec	165 (100%)	10 (6%)	155 (94%)
-0	414 (100%)	25 (6%)	389 (94%)

Table 2. Number of derivatives: trans. vs intrans.

The first base of the composite may belong to various categories: it may be a noun, an adjective, a pronoun, an adverb or a numeral (both ordinal and cardinal, cf. *pervoprohodec* 'pioneer' and *odnodum* 'a person fixated on a single thought'). As a rule, the first component of the composites with second intransitive verb stem is adjectival or adverbial denoting place or manner (cf. *tonkopr'ad* 'hepialidae moth'). Even if the first stem is substantival, its role in the structure of a composite still cannot be considered as purely substantival (verb argument), but rather as that of an adverbial modifier (cf. *domosed* 'houseling', *skalolaz* 'alpinist', *verhogl'ad* 'skygazer') (verb adjunct). On the contrary, the first components of the composites with second transitive verb stem are mostly substantival (they correspond to one of the initial verb syntactic arguments). The most frequent source for the first stem of such composites is a direct object expressed by a noun in accusative case (cf. *strastoterpec* 'passion-bearer', *vlastoderzhec* 'the person in power', *drovosek* 'woodcutter', *steklodel* 'glass-maker', *mukomol* 'miller'). Arguments and adjuncts with other semantic roles are much more rare, but possible: e.g. instrument (a noun in instrumental case, compare *molotoboec* 'hammerman', *ochevidec* 'witness') or manner (*skorohvat* 'agile and fast person', *pustopl'as* 'superfluous person'), whereas we did not encounter locatives with transitive stems.

For obvious semantic reasons, the first stem denoting a subject is impossible for the composites in question (unlike composites with zero suffix denoting nomina actionis, cf. *snegopad* 'snowfall' and **chelovekopevec* in the meaning of *'a man who sings'*): since the whole derivative denotes an actor, its existence would have been unnecessary and redundant if it had contained the name of the actor in itself. Yet it is possible that such composites are eliminated because of the constraint on external argument incorporation described in (Bogdanov, 2005).

3   Morpho(no)logical Constraints

Zero suffix model also imposes several morphonological constraints on possible motivating stems. Mostly they concern the second (verb) stem. According to our data, the composites with zero suffix may only be formed from one-syllable unprefixed imperfective verb stems. Another constraint that has already been described in (Sitchinava, 1999) does not allow to form such composites from verb stems which end with a consonant cluster (such composites are either impossible or very few). It is worth noticing that many verb stems ending with a consonant cluster per se combine easily with zero suffix, but still do not form composites (cf. such roots as *mysl-, plesk-, hrust-, tresk-* and others). Therefore, this constraint turns out to be not a universal constraint on the combination of zero suffix with verb stems of a certain kind, but a specific trait of the model under investigation. Apart from the second stem, zero suffix model also places constraints on the characteristics of the first stem. Among all the word families in our sample only very few allow multisyllabic first base: only 8 of 89 word families (51 derivative of 426), which constitutes only 9% of the overall number of word families. Thus we can claim that zero suffix model has a propensity to trisyllables. This property does not only constrict the range of possible first stems, but can sometimes lead to abridgement of the first motivating stem. However, increasing productivity of the model with zero suffix makes the constraint on the length of the first stem less rigid: the most numerous word families (which are *vod-, ved- and l'ub-*) allow for derivatives with a multisyllabic first base (cf. *vostokoved* 'orientalist', *maralovod* 'elk breeder'). According to our observations, the model with suffix *-ec* does not involve any morpho(no)logical constraints similar to those put by the zero suffix model and allows to form composites denoting nomina agentis from both multisyllabic prefixed verb stems (cf. *hristoprodavec* 'traitor', *zemleprohodec* 'explorer') and verb stems ending with a consonant cluster (cf. *streloverzhec* 'archer', *samoderzhec* 'autocrat'). In our opinion, this affects the rules of interaction of the two models under investigation, which will be discussed in the sections below. Historically the absence of morphonological constraints on the model with suffix *-ec* may be explained by its initial high productivity (on the history of the models see the next section), as in our opinion high frequency of a word-formation model gradually leads to elimination (full or partial) of morphonological constraints.

4   Doublets

Though most of the verb stems "prefer" only one of the two models to form composite nomina agentis, some of them, nevertheless, allow variation. These stems may be divided into two types: the ones that form composites with different first stems through different models (cf. *maralovod* 'elk breeder' – *flotovodec* 'naval commander', *medonos* 'honey plant' – *ordenonosec* 'order bearer') and the ones that allow what at first sight seems to be full

doublets – cognates with identical meaning, formed through different models (cf. *serdceved – serdcevedec* 'interpreter of human nature', *pravdol'ub – pravdol'ubec* 'truth-lover'). Usually full derivational doublets are rare, the same way that full semantic doublets – full synonyms – are rare or, according to some theories, even impossible. Cf. a well-known pair of Russian lexemes *olivki – masliny* 'olives' for which the native speakers of Russian "invented" a distinction to avoid full synonyms, though originally the words denoted the same object, with the only difference that one of them was a loan word and the other was not.

The same holds true in our case. On closer examination it becomes obvious that such pairs cannot be considered full doublets. There is either a difference in the shades of meaning, or one of the lexemes has additional meanings (i.e. the sets of meanings overlap but do not coincide, cf. *bogomol* 'mantis' – *bogomolec* 'pilgrim'), or there is a distribution of the members of the pair across different genres and styles, or we can observe clear trends in the time of occurrence, or one of the lexemes is considerably more frequent than the other and thus we can consider the second one to be occasional. Nevertheless, usually it is the examination of such pairs that gives the researchers the necessary information on the properties and the development of competing models. Thus we made a list of existing doublet pairs and compared the frequencies (according to the data of the Russian National Corpus (RNC)) and the years of first occurrence in written texts inside each pair and for each of the word-formation models as a whole. Resulting data is represented in the table below. We have highlighted the cases where one of the lexemes in the pair is considerably (more than five times) more frequent than the other.

We also compared lexeme frequencies across separate time periods (we divided the period from the beginning of the XVIII century to present day into several periods of 50 years each) in those pairs where the overall frequencies of the members are comparable. The results are represented in the histograms below.

-0	-ec	*Total number*	*Total number*	The first occurrence	The first occurrence
dushegub	dushegubec	263	108	1833	1766
zhenol'ub	zhenol'ubec	33	16	1892	1843
zhiznel'ub	zhiznel'ubec	43	16	1906	1877
pravdol'ub	pravdol'ubec	29	86	1848	1724
samol'ub	samol'ubec	4	20	1765	1765
svobodol'ub	**svobodol'ubec**	2	21	1923	1718
seb'al'ub	**seb'al'ubec**	3	61	1985	1829
slastol'ub	**slastol'ubec**	1	117	1786	1764
trudol'ub	**trudol'ubec**	4	25	1907	1760
chelovekol'ub	**chelovekol'ubec**	6	52	1846	1717
chestol'ub	**chestol'ubec**	2	254	1894	1766
serdceved	serdcevedec	36	61	1867	1757
verhovod	verhovodec	30	1	1862	1843-1847
morehod	morehodec	315	125	1825	1766
kitoboy	kitoboec	268	9	1855	1939
bogomol	**bogomolec**	495	1104	1765	1682-1709
vol'nodum	**vol'nodumec**	12	296	1784	1779
odnodum	odnodumec	7	3	1887-1894	2002

| inover	**inoverec**	2	334	1873	1709
starover	staroverec	483	8	1733	1833
duhobor	duhoborec	251	86	1890	1845
kr'uchkotvor	kr'uchkotvorec	30	24	1799	1772
vodonos	vodonosec	61	4	1792	1826

Table 3. Doublets

The table demonstrates that in most of the doublet pairs (in all of them, except for *kitoboy – kitoboec, odnodum – odnodumec, starover – staroverec, vodonos – vodonosec*) the derivative with suffix *-ec* appears in written texts earlier (and for large majority considerably earlier (for half a century on average)) than its correlate with zero suffix. However, it is also obvious that the frequency of the zero suffix correlate is much higher in most of the doublet pairs. Judging by the collected data, we can come to the conclusion that the active expansion of the zero suffix model of composites denoting nomina agentis started considerably later than the expansion of the model with the suffix *-ec* with the same meaning. To confirm this theory we turned to earlier sources: to the dictionaries of Old East Slavic containing data on written sources starting from the XI century.

According to the Dictionary of the Russian language of the XI – XVII centuries (Bogatova, 1975-2012), there were 35 word families of zero suffix composites denoting nomina agentis in the Old East Slavic language, and for most of the motivating verb stems there were parallel derivatives with the suffix *-ec*. Despite the comparatively large number of word families, the number of derivatives was small (all in all about 60 derivatives), which means that every verb base, which could produce such composites, had only one or two (rarely more) composite derivatives. This fact indicates the occasional nature of such composites. Nevertheless, it is clear that this word formation model existed in Old Russian and had already acquired (on a previous level of language development) the potential to denote an agent. On the contrary, the model with *-ec* was, according to the Dictionary of the Russian language of XI – XVII centuries, very productive and accounted for 84 word families and more than 400 derivatives (not counting graphical variants). Yet, as we have already mentioned before, many of those derivatives had already had doublets with zero suffix, which testifies to the early beginning of the competition of the two word formation models. Thus the data of Old East Slavic written texts confirms the hypothesis of earlier expansion of the model with *-ec* in the meaning of nomina agentis. Bearing this in mind, how can we explain the fact that in modern Russian the number of composite derivatives with zero suffix is much higher than the number of the derivatives with *-ec* and the same holds true for their frequency? Let us turn to the history of some of the doublet pairs functioning in modern Russian.

Let us consider those pairs in which the frequencies of the doublets are comparable. Such pairs as *dushegub – dushegubec* 'murderer', *zhenol'ub – zhenol'ubec* 'ladies' man' , *zhiznel'ub – zhiznel'ubec* 'swinger', *morehod – morehodec* 'sailor', *duhobor – duhoborec* lit. 'spirit wrestler', *kruchkotvor – kruchkotvorec* 'chicaner' and *vodonos – vodonosec* 'water-carrier' show the same common trend: the derivative with zero suffix appears much later, but over time it becomes much more frequent than its doublet with *-ec* (cf. the histograms below). In most cases expansion of the zero suffix derivative is accompanied by a simultaneous drop in the frequency of the derivative with *-ec,* which shows the process of extrusion of one word formation model by the other. According to the RNC data, this process begins in the XIX century and reaches its climax in the first half to the middle of the XX century. Sometimes it leads to a complete extinction of the derivative with *-ec*.
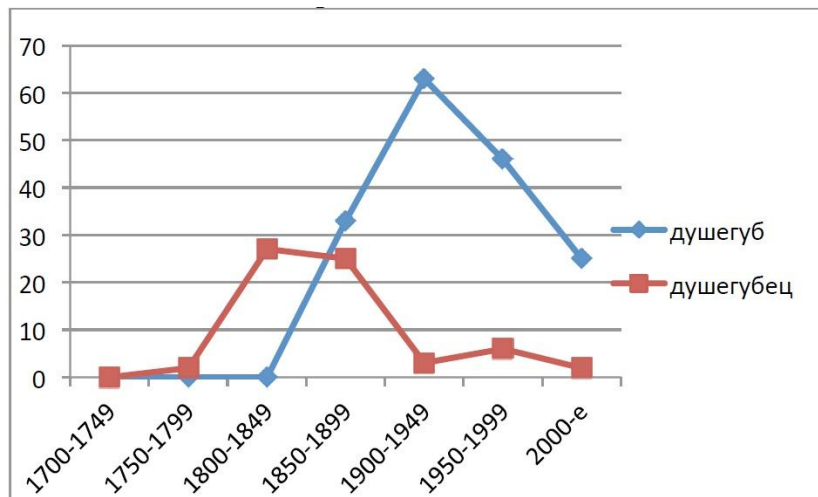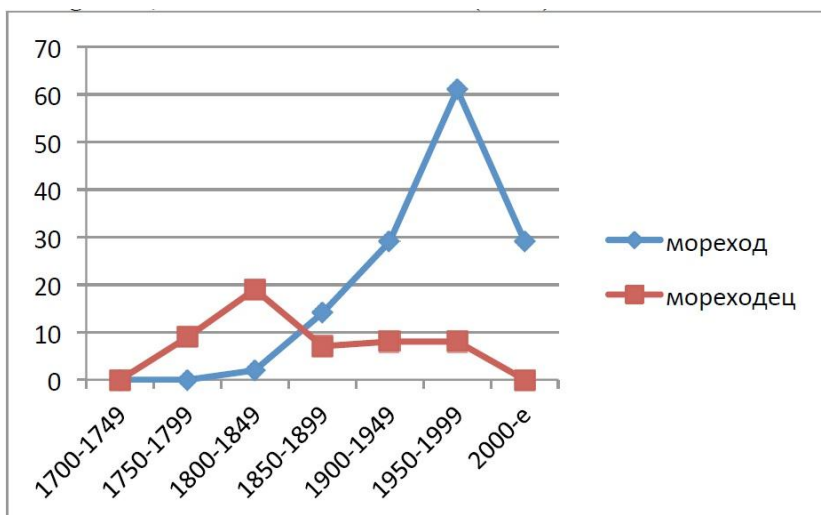
Histogram 1, dushegub – dushegubec (263 : 108)
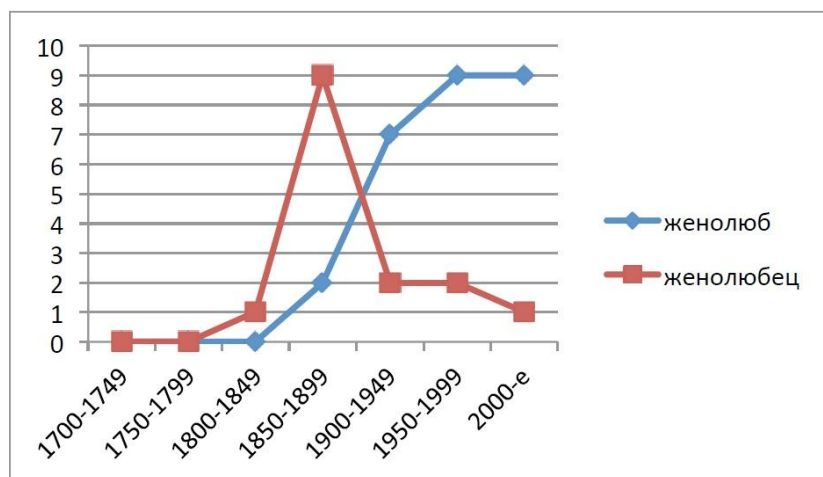
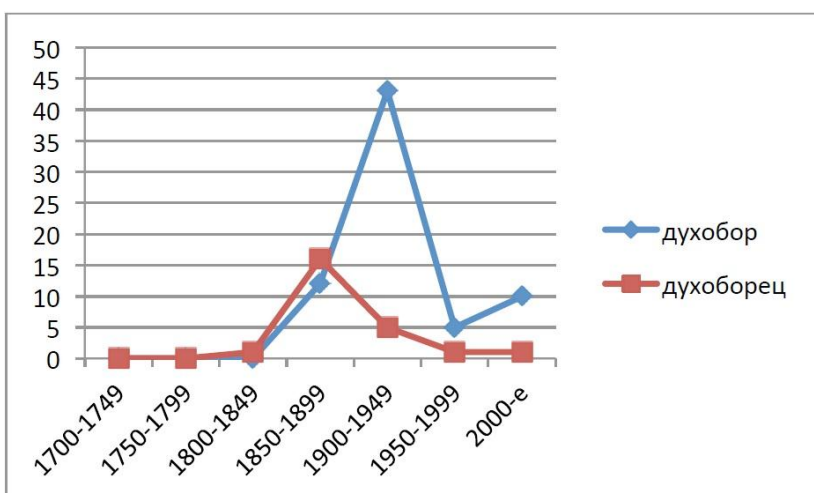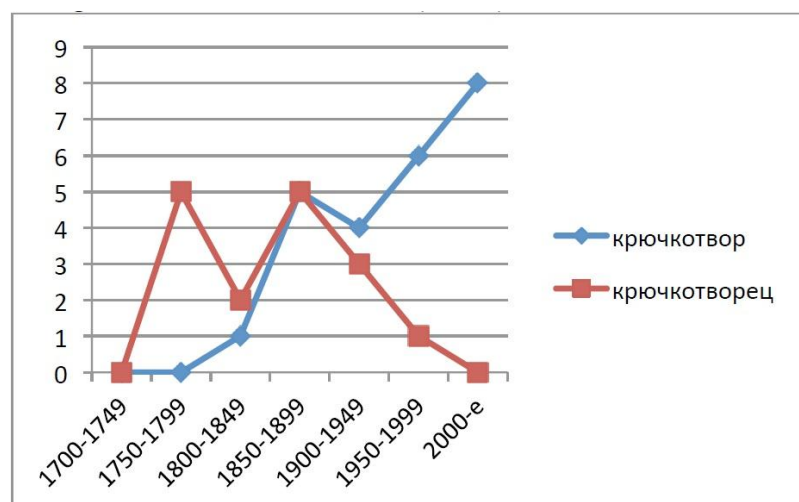**Histogram 2, zhenol'ub – zhenol'ubec (33:16)**
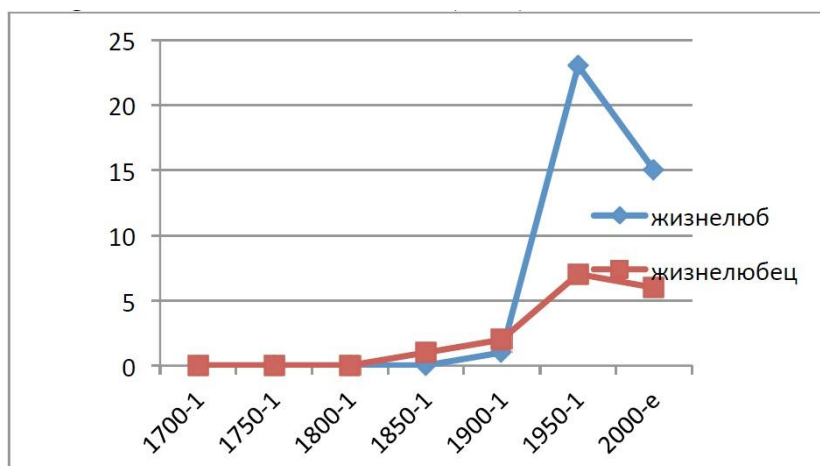
Histogram 3, zhiznel'ub – zhiznel'ubec (43:16)

Histogram 4, morehod – morehodec (315:125)

Histogram 5, duhobor – duhoborec (251:86)

Histogram 6, kr'uchkotvor – kr'uchkotvorec (30:24)

If we consider those doublet pairs, where the derivative with zero suffix seems (according to the corpus data) older than its fellow derivative with *-ec,* we will see that those cases do not in fact contradict our conclusions. In our opinion the pair *starover – staroverec* 'Old Believer' is not representative because of the great difference in frequencies (483 occurrences of the derivative with zero suffix against 8 occurrences of the derivative with *-ec*): we can consider the equivalent with *-ec* to be an occasional result of the competition and blending of the models under investigation in the language.

One of the two pairs of doublets, where the zero suffix derivatives actually turn out to be older than their correspondents with *-ec,* is the pair *kitoboy* 'whaler' – *kitoboec* 'whale boat', though its existence does not disprove our theory. Basically, these composites appear rather late (compared to other lexemes under investigation), which is clearly the result of previous absence of the realia. So the realia itself appears in the time period when the model with zero suffix had already become more common and, as a result, more productive than the competing model with *-ec.* No wonder the speakers had chosen this particular model when forming a new word. So why would the doublet with *-ec* even appear at all? Here we have to mention that the central meaning of the word *kitoboec* is not a name of a person (an agent), but the name of the boat used for this particular kind of fishery (whale hunting). Indeed, we can see that the meaning of the model with suffix *-ec* changes over time: being at some point "oppressed" by the more productive zero suffix model, it takes the adjoining semantic zone and acquires the meaning of machines and tools for particular activities: cf. *minonosec* 'torpedo boat', *avianosec* 'aircraft carrier'. At the same time, since the model had not yet lost the meaning of animated agent and since there still are many frequent derivatives with this meaning in the language, blending becomes almost inevitable and the word *kitoboec* 'whale-boat' sometimes (though rather rarely) is used in the meaning of *kitoboy* 'whaler':

(1) *Ah, chto tvorilos, kogda kitoboyci prihodili iz vosmimes'achnogo rejsa!* [Roman Karcev. *Maloy, Sukhoy i Pisatel'* (2000-2001)]
'What was going on when the whalers returned from the 8-month journey'

A similar process happens to another pair of doublets, where the cognate with zero suffix is older than the derivative with suffix *-ec: odnodum – odnodumec.* Here the lexeme with *-ec* is obviously a rather recent formation, its first occurrence in the RNC dates to the year 2002:

(2) *On rad privetstvovat' v etom zale, kotoriy skoro zapolnyat zriteli, svoih soratnikov i odnodumcev.* [Leonid Zorin. Jupiter (2001) // *Znamja*, 2002]
'He is happy to greet in this hall, which will soon be filled with the audience, his comrades and those who think alike'.

Initially the meanings of the two words differed from each other: *odnodum* stood for 'a person who thinks about one and the same thing all the time' and *odnodumec* meant 'a person who has the same position on an issue'. In order to form a word with a new meaning based on the roots that have been merged once before, the language chooses a different word formation model (though a less productive one). But later, **after** the word *odnodumec* had already emerged and probably under its influence, the word *odnodum* acquired a second meaning, which, according to the corpus data, had not been attested before:

(3) *Vot i zdes' sverstniki i odnodumi avtora iz prezhnej sovetskoj zhizni nesostarivshimisya i polnimi sil desantiruytsa v dostatochno poganuy nineshnyy deistvitel'nost' s ejo kriminalom i demokraticheskimi viborami.* [Vladimir Baranov.

Buduschego ne budet (2003) // *Lebed'* (Boston), 2003.07.28]
'And here the peers and comrades of the author from the past soviet realm, unaged and full of vitality, land into the quite disgusting present reality with its crimes and democratic elections'

So it turns out that interaction of derivational doublets formed by different models is not limited to purely temporal relations. It is important that when such doublets coexist in the language within the same period of time, the language tends to draw a semantic or (at least) a stylistic dictinction between them. We can clearly observe the same trend on the material of the following doublet pairs. In modern Russian the words *bogomol* and *bogomolec* have different sets of meanings, cf:

БОГОМОЛ,·муж. (устар.). Богомольный человек.
Bogomol (m, archaic) 'a prayerful person'

II. БОГОМОЛ, муж. (зоол.). Большое прямокрылое насекомое теплых стран, иначе — богомолка.
II bogomol (m, zool.) 'a large orthopteran that inhabits warm regions'

БОГОМОЛЕЦ, богомольца, м. 1. Ходящий на богомолье. Толпы богомольцев шли к монастырю. 2. Любящий молиться. Старик у нас богомолец. Не богомольцы вы (крестьяне) весной. Д. Бедный. 3. Молящийся за кого-н. богу (устар.) (Д.Н. Ушаков «Большой толковый словарь современного русского языка»)
Bogomolec (m) 1. A person who comes to bogomol'je. 2. A person who likes to pray. 3. A person who prays for someone (archaic)
[D. Ushakov, *Bol'shoj tolkovyj slovar' sovremennogo russkogo jazyka*]

Yet according to the corpus data the central meanings of the words used to coincide:

(4) *Vse my zhe – bogomoly-brat'ja, l'udi bozhii, svoj narod [V. Krestovskij. Peterburgskije Truschoby. Kniga o sytyx i golodnyx. Roman v shesti chastjax. P. 4 (1864)]*
lit. 'All we are bogomols-brothers, people of God, our nation'

Bogomolec is clearly of earlier origin: unlike *bogomol* it occurs already in Old Russian texts:

(5) *Ot davnyx dobryx vremen i dosele chto zovutsja bgomol'cy i ves' popovskyj chin tem ne nadobe ni kotoryje poshliny*
'Those who from old good times and to now are called bgomol'cy, and all the priests, are not obligated to pay any levy'

(6) *Ne nadobe ni ego ludem ni vsem crkonym bgmolcom popom i chernecom...ni kakova dan'. [Sreznevsky, (1890-1912)]*
'Neither his people nor any of the church people are obligated to pay any levy'

Also the lexeme may be found in the texts of late XVII century, the first occurence in the Russian National Corpus dates back to 1682-1709:

(7) *O chem iz'javja i vashego svjatejshestva, v Duxe Svjatom otca nashego i bogomol'ca, prosim, daby za takoje neizrechennoe Bozhie miloserdie soborno i kelejno molebnoe blagodarenie vozdavali i o nashem zdravii i vsego voinstva molili*

[I. Zheljabuzhskij. Dnevnye zapiski (1682 – 1709)]
'Having told this to your holiness, our father and bogomolec in the Holy Spirit, we ask that they pray in cathedrals and in cells for the health of ours and of all the army'

The word *bogomol* appears only in the second half of the XVIII century and has the initial meaning of 'pilgrim', which is nowadays, according to the data of dictionaries and corpora, outdated: for the last time it appears in this meaning in the middle of the XX century. It is possible that this pair has undergone the same process which sometimes happens to full synonyms (we would call this process "dissimilation": one of the doublets has changed its central meaning over time because two words with identical meanings (in our case even two cognates formed by the same means) create a muddle and are obviously redundant in the vocabulary of the language. The competition of the two lexemes over time may be observed on the histogram below (the graph for the word *bogomol* was made taking into account only its occurences in the meaning of 'pilgrim'; in the meaning of insect *(mantis* religiosa) the lexeme is still frequent):
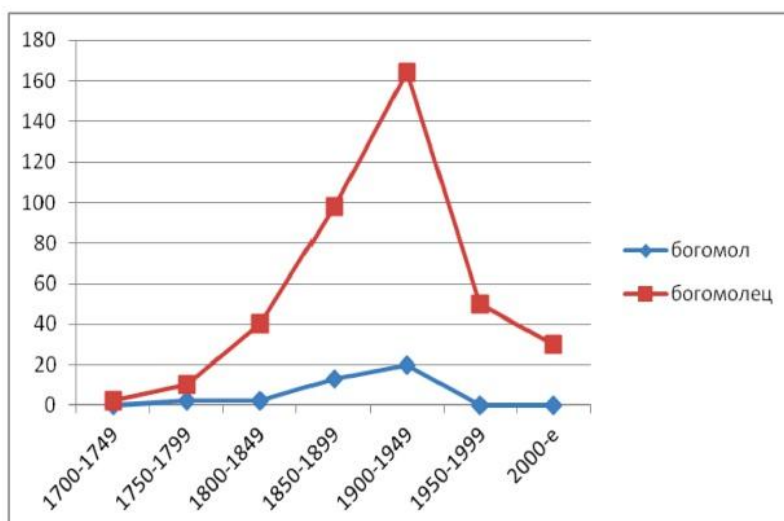
Histogram 7, bogomol – bogomolec (495 : 1104)

The pair *serdceved – serdcevedec* 'interpreter of human nature' is also an interesting example of semantic and stylistic differentiation of doublets. The lexeme *serdcevedec* emerging much earlier than its correspondent is initially used only to refer to God:

(8) *Kak mozhesh ty Serdcevedcy skazat': uslyshi Gospodi pravdu moju: kogda vopl' ot tvoego pritesnenija strazhduschix vxodit vo ushi Gospoda Savaofa?* [arxiepiskop Platon (Levshin) Slovo na osvjaschenije xrama (1777)]
'How can you say to Serdcevedec [God]: "Hear my truth, Lord", when the cry from your oppression of the sufferers reaches the ears of the Lord?'

(9) *K tebe vozzovu s Daviom: Uslishi mja vo dni pechali, ibo ty esi bog kajuschixsja; gospodi, sogreshix i nesm' dostoin pomilovanija. No, sercevedec vsezryaschij! dusha moja pred toboju est'.* [A. Radischev. S. I. Sheshkovskomu (1790)]
'I will call to you with David ['s name]: "Hear me in the days of sorrow, because you are the god of repentants; Lord, I sinned and I am not worth the mercy. But, sercevedec all-seeing! my soul is before you"

By the beginning of the XIX century the meaning of the word expands and it is used in the meaning of 'a person who understands the feelings and the thoughts of other people':

(10) *Tut malo byt' zakonovedom i bespristrastnym: tut nado byt' serdcevedcem etogo naroda.* [A Bestuzhev-Marlinskij. *Mulla-Nur* (1836)]
'It is not enough to know the laws and be impartial, one must be a sercevedec [=understand the mentality] of this people'

Yet the initial meaning of the lexeme is still relevant nowadays; it is interesting that, beginning with the end of the XX century, it is encountered only in texts somehow related to religion:

(11) *Dalee serdcevedec Gospod' napominaet nam, chto stoit za nashimi zemnymi zabotami, za nashej mrachnoj ozabochennostju, meshajuschej nam vo vsjom videt' Bozh'ju slavu* [S. Averincev. Iz propovedej (1993) // *Kontinent*, 2004]
'Then God-serdcevedec reminds us what stands behind our earthly worries, our grim anxiety that prevents us from seeing God's glory in everything'

(12) *No pochemu zhe eto imja prisvoil Serdcevedec Xristos tishajshemu i nezhnejshemu iz uchenikov Svoix – Apostolu Ioannu?* [mitropolit Vladimir (Ikim) Slovo v den' pamjati apostola i evangelista Ioanna Bogoslova (2004) // *Zhurnal Moskovskoj patriarxii*, 2004.05.24]
'But why has serdcevedec Christ name so his most quiet and most gentle apprentice – John the Apostle?'

The word *serdcevedec,* having appeared much later (its first occurrence in RNC dates to the year 1867 and it is not attested in the dictionaries of Old Russian), has never been used to refer to God during its existence. Apparently, these doublets have developed a clear semantic distinction (and as a consequence a distribution over texts of different topics), but such conclusion would not have been fully justified as both lexemes are too rare in the modern language (6 occurrences of each in the period from the second half of the XX to the beginning of the XXI centuries).

5   Discussion

Despite the described trend of extrusion of one of the models by the other, many composites with the suffix *-ec* are still in use and sometimes much more frequent than the corresponding derivatives with zero suffix. What influences the competition and does not let the model with the suffix *-ec* become extinct in the meaning of animate agent in the Russian language?

Apparently, there are several factors. In the first place, the morphonological constraints on the zero suffix model do not let all the derivatives to change their word formation model by replacing the suffix. That is why, in our opinion, such words as *chelovekol'ubec* and *svobodol'ubec* are considerably more frequent than the corresponding *chelovekol'ub* and *svobodol'ub,* which have nevertheless appeared under the influence of the common tendency.

Stylistic differentiation which has already been mentioned above also has a rather big influence on the competition of the models in question: in the modern language composites with the suffix *-ec* (in the meaning of nomina agentis) mostly belong to written texts and are

used as a means of archaicism. So they acquire a function which a neutral zero suffix model does not have.

Bibliography

Bogatova, G. 1975-2012. *Dictionary of the Russian language of the 11th – 17th centuries.* Moscow.

Bogdanov, A. 2005. Slozhnye slova v russkom jazyke: leksicheskie ili sintaksicheskie ogranichenija? // *Proceedings of the second conference on typology and grammar for young scholars*. Saint Petersburg.

Sitchinava, D. 1999. *Ogranichenija na vtoruju glagolnuju osnovu v russkih slozhnyh slovah s nulevym suffiksom*. Kursovaja rabota (manuscript).

Sreznevsky, I. 1890-1912. *Dictionary of the Old Russian Language on the Basis of Written Records.*

The Regularities of Diminutive Formation from Russian Quasisubstantive Adverbs

Margarita Tyrenkova

National Research University Higher School of Economics
20 Myasnitskaya Ulitsa, Moscow, 101000, Russia
kusiar8@gmail.com

Abstract

The paper considers the regularities which govern the adverbial diminutive formation in Russian. The focus is placed on the adverbs with suffixes *-om[1]/-em*, *-oj/-ej*, *-ju*, *-ami/-jami* (e.g. *vecherom* 'in the evening', *noch'ju* 'at night', *peshkom* 'on foot', *polzkom* 'crawling') and the adverbs with *v-...-k-u* structure (e.g. *vrazvalku* 'in a waddling manner', *vpripryzhku* 'in a jumping manner', *vtikhomolku* 'stealthily'). The paper shows that the diminutive formation is influenced by phonetic, accentological and frequency factors. Moreover, the semantic conditions which determine this process are discussed.

Keywords

Word formation, diminutives, adverbs, semantics, accentology, frequency, dissimilation

1   Introduction

One of the most well-known features of Russian language is its productive diminutive formation. This word-formation mechanism could be applied not only to nouns and adjectives, but also to adverbs, which are usually thought to have smaller word-formative potential. Although diminutive adverbial forms are comparatively frequent in Russian, there is hardly any systematical and comprehensive description of this phenomenon. The existence of such forms is stated with several examples in (Zemskaja 1973: 298), their morphological structure is discussed in (Lopatin 1972), and the semantic features of some of these forms are described in (Zaliznjak & Shmelev 2005: 42). However, these works do not consider the frequency of these forms or the regulations which govern the diminutive formation.

In this paper we focus on the adverbs which are morphologically productive as far as diminutive formation is concerned: **(1)** the adverbs with suffixes *-om/-em*, *-oj/-ej*, *-ju*, *-ami/-*

---

[1]   Here and further we use transliteration GOST 16876-71

jami, which are formally identical to flexions of instrumental case (e.g. *vecherom* 'in the evening', *noch'ju* 'at night', *peshkom* 'on foot', *polzkom* 'crawling') and **(2)** the adverbs with *v-...-k-u* structure, where *v-* is formally identical to the preposition *v* 'in' and *-u* is one of the flexions of accusative case (e.g. *vrazvalku* 'in a waddling manner', *vpripryzhku* 'in a jumping manner', *vtikhomolku* 'stealthily') We call these adverbs "quasisubstantive", as they are formed with the suffixes identical to the declensional endings in nouns. Such structure makes these adverbs behave like nouns in view of word-formation and they form diminutives using the substantive suffixes, mainly the suffix *-к-*. Although these adverbs are in the same morphological conditions for diminutive derivation, the frequency of their diminutive forms is different. The aim of this paper is to determine the factors which influence this process.

The research is based on the data of the Russian National Corpora (RNC)[2]. For each adverb with structure **(1)** and **(2)** listed in (Zaliznjak 2003) the diminutive form was searched. Depending on whether the diminutive was frequent (the percent of diminutive forms is greater than or equal to 1), rare (the percent is less than 1) or does not occur in RNC at all, the adverbs were divided into three groups. Moreover, all these adverbs were divided into five groups in accordance with their meaning.

2 Restrictions

First of all, it seems useful to mention the restrictions that complicate the formation of diminutives from some adverbs. In this section we list these factors and discuss the mechanisms of their influence. The concrete examples will be given in each section.

2.1 Accentological Factor

Among several Russian diminutive suffixes, it is the suffix *-ok* that is most frequently used in adverbial diminutive formation. But it has a combinatory power restriction caused by its accentological properties: this suffix (in diminutive meaning) could not be attached to a strong base component (i.e., to a component which does not tend to have a movable accent), except the stems with *к*, *г*, *х* at the end (Zaliznjak 1985: 81). First of all, the stems which consist of prefix and root or of two roots belong to this group.

2.2 Phonetic Factor

Some differences in frequency of adverbial diminutive forms could be explained by the influence of phonetic factor, in particular, by the tendency to avoid repetition of the same sound in adjacent syllables (for more details, see (Browne 1999; Itkin 2005)). Concerning the adverbial diminutives, the repetition of sound *ch* at the end of the stem and in the suffix is avoided.

---

[2]   www.ruscorpora.ru, as of 13.04.13

2.3   Frequency Factor

The last factor which influences the frequency of diminutive form of the adverb is the frequency of its primary form. If the adverb is rare, it will not tend to be productive from the point of view of diminutive formation.

3   Adverbs of Manner

This group includes the adverbs denoting the way of movement, the way of speaking and some other adverbs of manner.

3.1   Adverbs Denoting the Way of Movement

diminutive form is frequent		diminutive form is rare		diminutive form does not occur in the RNC
	% of diminutives		% of diminutives	
shagom[3]: 6891 shazhkom: 107	1.5[4]	polzkom: 464 polzochkom: 1	0.2	kubarem, begom, cugom, zadom, peredom, naezdom, zaezdom, mimoezdom, proezdom, samokhodom, prokhodom, juzom, guzhom, porozhnjakom, naskokom, volokom, broskom, petushkom, valom, peshedralom, khodunom, kuvyrkom, mimoletom, naletom, opromet'ju, vdogonku, vperebezhku, vperegonku, vpripryzhku, vpriskochku, vrastrusku
peshkom: 6775 peshochkom: 63 peshechkom: 29	1.3	galopom: 551 galopchikom: 1	0.2	
rys'ju: 1306 ryscoj: 346	26.5			
vperevalku: 75 vperevalochku: 28	37.3			
vrazvalku: 177 vrazvalochku: 105	59.3			
vraskachku: 83 vraskachechku: 1	1.2			

---

[3]   It is often difficult to distinguish the adverbs of this morphological type from the corresponding nouns in the instrumental case (AG-80: 398)

[4]   Although they do not occur in RNC, some of them could be found on the Internet. However, their frequency is not high.

The accentological factor makes it impossible to derive the diminutive forms from such adverbs as *naezdom*, *samokhodom* or *mimoletom* due to their stem structure ("prefix + root" or "root + root"). The phonetic factor complicates diminutive formation from the adverbs *vraskachku* and *vpriskochku* because of the sound *-ch-* at the end of the stem. The frequency factor can explain low frequency of the adverb *polzkom*, which is rare itself and which form a rare diminutive *polzochkom.*

However, the formal restrictions cannot explain the low frequency of the diminutive forms from the adverbs *begom* 'at a run', *kubarem* 'head over heels', *kuvyrkom* 'topsy-turvy', *broskom* 'with one throw' and *opromet'ju* 'at top speed'. They differ from the adverbs with frequent diminutive forms (*shagom* 'at a walk' or 'step by step', *peshkom* 'on foot', *vrazvalku* 'in waddling manner', *vperevalku* 'in waddling manner') by their meaning: all of them denote fast impetuous movement, while the diminutive productive adverbs denote "careless" movement with a low speed.

The diminutives from the adverbs *rys'ju* 'at a trot' and *galopom* 'at a gallop' join this group. The diminutive form *galopchikom* (0,2%) is much less frequent than the diminutive form *ryscoj* (26,5%) probably because gallop is faster than trot.

3.2   Adverbs Denoting the Way of Speaking

diminutive form is frequent		diminutive form does not occur in the RNC
	% of diminutive forms	
vrastjazhku: 110 vrastjazhechku: 6	5.5	vperebivku
shepotom: 7957 shepotkom: 180	2.3	molchkom
vtikhomolku: 639 vtikhomolochku: 20	3.1	tikhomolkoj
tikhomolkom: 104 tikhomolochkom: 1	1	

In this group the number of the adverbs which do not form diminutives is insignificant. However, the adverbs with frequent diminutive forms have a common semantic feature: all of them denote quiet slow speech.

As for the adverbs with rare diminutive forms, this could be explained by means of influence of restrictions. Thus, the adverb *vperebivku* '(to speak) interrupting each other' is rare in its original form and the adverb *molchkom* 'silently' has *-ch* at the end of the stem, which makes diminutive formation difficult because of the phonetic factor.

The adverbs *tikhomolkom, tikhomolkoj* and *vtikhomolku* 'on the quiet', which are connected with the adverbs of speech in a metaphorical way, also demonstrate the same regularities of diminutive formation. The adverb *tikhomolkoj* (2 tokens in RNC) is rarer than *tikhomolkom* (104 tokens) and *vtikhomolku* (639 tokens), and, correspondingly, there are no diminutive forms from this adverb in RNC.

3.3 Other Adverbs of Manner

diminutive form is frequent		diminutive form is rare		diminutive form does not occur in the RNC
	% of diminutives		% of diminutives	
golyshom: 115 golyshkom: 2	1.7	bosikom: 1501 bosichkom: 8	0.5	khodenem, zapoem, putem, migom, propadom, gradom, poedom, chudom, razom, nizom, porjadkom, samotekom, celikom, durikom, tajkom, silkom, tolkom, nenarokom, schipkom, mel'kom, sploshnjakom, navalom, ogulom, nakhrapom, zalpom, skopom, darom, dobrom, khodorom, fuksom, rikoshetom, ekspromtom, optom, gurtom, chokhom, dukhom, slykhom, teleshom, nagishom
ladom: 58 ladkom: 50	86	sledom: 4823 sledkom: 1	0.02	
		prjamikom: 687 prjamichkom: 1	0.1	
vprikusku: 134 vprikusochku: 6	4.5			vprigljadku, vzatjazhku, vpervinku, vdogonku, vprishcurku, vprikhvatku, vsukhomjatku, vsmjatku
vnakladku: 48 vnakladochku: 4	8.3			
				volej, nevolej, storicej, dorogoj,ukradkoj, obydenkoj, samouchkoj, siloj, zimoj, vesnoj, tolpoj, poroj, osen'ju, nasyp'ju, rossyp'ju, propis'ju, chast'ju, polnost'ju

Among the other diminutive-productive adverbs of manner there are the adverbs *bosikom* 'barefoot' (0.5% of diminutive forms) and *golyshom* 'stark naked' (1.7% of diminutive

forms), which are often used when speaking about a child (cf. the adverb *nagishom*, which is a synonym of the adverb *golyshom*, but is not used in the "child" context and does not form diminutives). Moreover, there is a significant number of diminutive forms from the adverbs *vnakladku* '(to drink tea) with sugar in' (8.3% of diminutives) and *vprikusku* '(to drink tea) holding a lump of sugar in one's mouth' (4.5% of diminutives). The diminutive forms tend to occur together with the noun *chaj* 'tea' in the diminutive form (*chajok*):

(1) Nadejat'sja teper' ne na kogo, a mezhdu tem privykshaja k kusku guba net-net da i sprosit libo supca s khoroshej, nastojashheju, kak v starinu byvalo, govjadinkoj, libo **chajku vnakladochku**, libo vinca krasnen'kogo, kotorogo do toshnoty zakhotelos' babenke… [A. I. Levitov. Moskovskie komnaty snebil'ju (1863)]

(2) ― Sladkoe brashno, mamasha, ― v pisanii govoritsja, ― tol'ko gortan' veselit, kormit zhe chervja neusypajushhego… khe-khe… A vy mne vot **vnakladochku… chajku-to…** ― I-i… [F. D. Krjukov. Set' mirskaja // Russkoe Bogatstvo, 1912]

Similarly, diminutive forms from the adverbs *prjamikom* and *sledkom* occur when the same or the adjacent sentences contain diminutive forms from nouns:

(3) ― **Shokoladochku**, **shokoladochku** vdogonku, ― zakhlopotal Florovskijj. ― Nu, i my **sledkom**. [Semen Daniljuk. Rublevaja zona (2004)]

(4) A luchshe idi vse **prjamichkom**, po avtobanu, na volju bozhiju. Ehkh ty, vojaka nasha siraja! ― I ottogo, chto bol'she nechem bylo snabdit' v dorogu, pokrestila ee **razochka** dva. ― Nu, stupajj svoejj dorogojj… da **shejku-to** beregi, **bylinochka** moja! [L. M. Leonov. Russkijj les (1950-1953)]

This could be explained by the fact that the diminutive meaning is spread over the whole phrase and the diminutive marker appears on several components of this phrase (not one) (Rusakova 2009: 33).

4 Adverbs Denoting Place and Position in Space

In this group there are two sub-groups: the adverbs denoting the position of the objects towards each other (in the upper part of the table) and the adverbs denoting the form of the object (in the lower part of the table).

diminutive form is frequent		diminutive form is rare		diminutive form does not occur in the RNC
	% of diminutives		% of diminutives	
krugom: 17 578 kruzhkom: 612 kruzhochkom: 11	3.5			vnakidku, vraskidku, vrazrjadku, vnatjazhku, vperemezhku, vrassypku, vnakhlestku,
rjadom: 55 058 rjadkom: 327	1.7			

rjadyshkom: 636				vperevertku, vperemeshku, vpritychku, vrazbivku vrazbrosku, vrastopyrku
osobnjakom: 705 osobnjachkom: 18	3.7		0.3	
vpritirku: 75 vpritirochku: 8	10.7	vpovalku: 354 vpovalochku: 1	0.3	
raskorjakoj: 60 raskorjachkoj: 2	3.3			nichkom torchkom bitkom
		dybom: 953 dybkom: 3		

In this group the influence of phonetic and frequency factors can be observed. Thus, the synonymous adverbs *vpritirku* and *vpritychku* 'very tightly' have different number of diminutive forms: the diminutive form *vpritirochku* is frequent, whereas the form *vpritychechku* does not occur in RNC, which could be explained by the influence of phonetic factor (the stem *vpritir-* has no *-ch* at the end, unlike the stem *vprity**ch**-*). Concerning the frequency factor, in RNC there are no diminutives formed from the rare adverbs *vrassypku* and *vperevertku* (which are registered in RNC one time each).

The semantic factor is also relevant for this group. The adverbs *vpritirku* 'very tightly' and *rjadom* 'closely' are diminutive productive because they denote the small distance between the objects, which is emphasized by means of using the diminutive suffix.

5   Adverbs Denoting a Point in Time

This group consists of the adverbs denoting time of the day: *vecherom* 'in the evening'*, utrom* 'in the morning'*, dnjom* 'in the afternoon' and *noch'ju* 'at night'. The adverbs *vecherom* and *utrom* form a significant amount of diminutives (1,8% and 0,6% of diminutive forms), while the adverbs *dnjom* and *noch'ju* do not (0,01% and 0,006% of diminutive forms). The difference between these adverbs is in their semantics: in the perception *vecherom* and *utrom* denote the periods of time which are shorter and which are intermediate between day and night. Moreover, according to (Zaliznjak, Shmelev 2005: 42), the appearance of endearment suffix in the diminutive forms from the adverb *utrom* could be connected with the human activity which takes place in the morning.

6   Conclusion

It has been shown that adverbial diminutive formation is influenced by a complex of factors. Besides the restrictions related to accentology, phonetics and frequency, it is governed by a semantic factor. To sum up, it seems useful to list the semantic components which cause the diminutive endearment suffix to be attached to an adverb with the meaning of:

- low speed of movement;

- low volume and speed of speech;

- small distance between the objects;

- short period of time.

In addition, uses of diminutive form of the adverb could be governed by the context of utterance: for instance, the presence of the diminutive form from another part of speech or addressing a child.

Bibliography

Browne 1999 – Browne W. Dissimilarity of consonants as a factor in suffixation // *Guard the Word Well Bound*: Proceedings of the Third North American-Macedonian Conference on Macedonian Studies. In Honor of Professor Horace Gray Lunt on the Occasion of His 80th Birthday. Indiana Slavic Studies. V. 10. 1999.

AG-80 – *Russkaja grammatika* (pod redakciej N.Y.Shvedovoj). M., 1980

Itkin 2005 – Itkin I.B. Ob odnom ogranichenii na sochetaemost' suffiksov s osnovoj v sovremennom russkom jazyke // *Slavjanovedenie*. 2005. № 4.

Lopatin 1972 – Lopatin V.V. O strukture suffiksal'nykh ocenochnykh narechij // *Razvitie sovremennogo russkogo jazyka 1972: Slovoobrazovanie. Chlenimost' slova*. M., 1975.

Rusakova 2009 – Rusakova M.V. *Rechevaja realizacija grammaticheskikh elementov russkogo jazyka.* (abstract of the Doctor of Letters dissertation) SPb., 2009

Zaliznjak 1985 – Zaliznjak A.A. *Ot praslavjanskoj akcentuacii k russkoj* . M., 1985.

Zaliznjak 2003 – Zaliznjak A.A. *Grammaticheskij slovar' russkogo jazyka*. Izd. 4-e, ispr. i dop. M., 2003.

Zaliznjak, Shmelev 2005 – Zaliznjak Anna A., Shmelev A.D. Vremja sutok i vidy dejatel'nosti // Zaliznjak A.A., Levontina I.B., Shmelev A.D. *Kluchevye idei russkoj jazykovojkartiny mira.* M., 2005.

Zemskaja 1973 – Zemskaja E.A. *Sovremennyj russkij jazyk. Slovoobrazovanie.* M., 1973.

Russian Double Verbs in the 1st PL Imperative

Daniel Weiss

University of Zurich Slavisches Seminar
dawe@slav.uzh.ch

Abstract

In this paper, serialized chains of verb forms of the type *pojdem posmotrim* are first examined with respect to their grammatical homonymy (1PL.IMP vs. FUT1PL) and lexical, syntactic and contextual cues that allow to disambiguate it. In section 2, a thorough analysis of data from the Russian National Corpus serves to capture the aspectual and lexical constraints of such chains. The last section contains a systematic comparison of the mentioned forms with serialized imperatives in the 2nd PL.

Keywords

Imperative, verb serialization, double verbs, colloquial Russian, Russian National Corpus, grammatical homonymy, lexical constraints, verbal aspect

1   Introduction: A Tricky Case of Grammatical Homonymy

This study continues a series of articles on the so-called double verbs (for definition see Weiss 2012, 613) in colloquial Russian and written genres influenced by the latter.[1] These papers were devoted to the description of the different subtypes based on the continuum between the prototype (the semantic and prosodic merger of both verbal components denoting one single event) and the prosodic/semantic twins and their conditions of use in modern colloquial speech and the language of traditional folklore. Moreover, Weiss (2000) tackled the question of how this construction should be modelled in an MTT framework of dependency syntax, Weiss (2003) examined new corpus data from the language of the contemporary press and fiction and established a parallel to similar constructions in Finno-Ugric languages spoken in European Russia, and Weiss (2008) focused on the interpretation of double verbs with perfective aspect, whereas Weiss (2012) treats the whole construction as an instance of SVC (serial verb construction) and elaborates on the Finno-Ugric parallels, including Mordvin and some relics in Balto-Finnic. Whereas the latest papers were based on a corpus of

---

[1]   The relevant list includes letters, diaries, literary fiction, press genres such as reportage, feuilleton etc., Internet communication (blogs, forums, chats), traditional and modern (urban and religious) folklore and dialects. It goes without saying that not all these genres could be equally extracted.

slightly more than 800 tokens in all possible grammatical forms extracted mostly by hand, in a recent study (Weiss 2013) we exclusively investigated one word form, viz. the imperative of the 2PL in the basic corpus ("osnovnoj korpus") of the Russian National Corpus (NKRJa). The results of that search will be contrasted with the findings of the present study in section 3.

As suggested through the use of 1PL.IM, the forms in question serve to realize directive speech acts including both one or more addressee and the speaker. In principle, such forms may be **morphologically, syntactically** and/or **lexically** marked, or else remain unmarked. A morphological expression exists in Polish, which offers a special verb form, cf. *pójdźmy* 'let us go' vs. *pójdziemy* 'we will go'. The syntactic alternative is available in languages that use a special word order pattern, cf. German *gehen wir!* (imper.) vs. *wir gehen* (indic.), or require the omission of the subject pronoun, cf. French *allons!* (imper.) vs. *nous allons* (indic.). Lexical means may consist of modal verbs such as German *Lass(t) uns gehen* or English *let us go,* but also of particles*,* cf. German *gehen wir mal!*. Russian, however, has no obligatory marking: in both the perfective (cf. *pojdem!* 'let's go!') and imperfective aspect *(budem pljasat'!* 'let's dance') the imperative is homonymous with the corresponding pf. or ipf. future tense FUT1PL. In spoken language, **prosody** gives the decisive hint as to which reading is the preferred one. When dealing with written data, however, we have to rely on additional cues. The imperative reading may be optionally marked by (i) the postfixed **particle -ka,** (ii) the **semi-particles**[2] *davaj, davajte* and/or (iii) the **postfix -te,** which expresses plurality of addressees and/or the politeness form, cf. *pojdemte!* 'let us go' (you all + me / you_pol [all] + me). All three procedures may co-occur with each other, cf. *Davajte-ka sjademte!* 'Let us sit down'. As for (iii), *pojdemte* is synonymous with *davajte pojdem!* It should, however, be pointed out that despite Isačenko's view to the contrary, the meanings expressed by the type *pojdemte* may also be realised by the unmarked 1PL.IMP *pojdem,* which thus covers all four values of the two features ">1adressee" and "±polite". And finally, **punctuation** partially serves as another disambiguating device: the exclamation mark supports the imperative reading, whereas the question mark blocks it. However, even the exclamation mark does not provide an infallible proof since it may occur with the FUT1PL as well. To make things worse, even *-te* turns out to be less impervious than described in text books since it sometimes appears in sentences marked as questions, cf.

(1) *Хотите, — пойдемте погуляем? .. Чудная погода!* [В. В. Вересаев. К жизни (1908)]
Xotite, — pojdemte pogulajem? .. Čudnaja pogoda!
lit.'Do you want – let's go for a walk? .. it's beautiful weather!'

Here, the inclusive imperative postfix *-te* conflicts with the question marked by both punctuation and the preceding *xotite;* most probably, the question mark is a misprint. But what are we to do with an utterance marked simultaneously as a question and an exclamation? Cf.

(2) *— Это ты на кого намекаешь, сволочь! Пойдем выйдем?!* [Ильдар Абузяров. Ненормативная лексика (2002)]
Eto ty na kogo namekaješ, svoloč! Pojdem vyjdem?!
'Whom are you referring to, you scoundrel! Let's go outside!'

---

[2] Since these forms still distinguish number they do not meet the decisive criterion 'lack of inflexion'. Note by the way that *davati* in Common Slavic could have the same modal meaning as English *let* or German *lassen*; in Russian, this use is more constrained than in West Slavic by now, cf. von Waldenfels 2012.

Here, either *pojdem vyjdem* functions as a pragmateme with unequivocal directive meaning (1PL.IMP; see next section), or the speaker presents two alternative solutions of the conflict (FUT1PL): 'Let's go outside, or take your words back!'

On the other hand, according to (Isačenko, 1975: 308) one has to take into account a negative criterion: if the sentence contains the overt subject *my* 'we', the indicative reading is mandatory. This is, however, not a thoroughly reliable criterion since spoken Russian even in the indicative freely allows for the omission of the subject pronoun. In this regard, Russian is located in an intermediate position between French and a real PRO-drop language like Italian, where the subject pronoun is only added when required by the communicative structure (cf. *andiamo* 'let us go!' or 'we are going'). But Isačenko's restriction should be verified anyway. After all, with imperatives of the 2SG and 2PL the subject pronoun does occur either when being rhematised as in *pojdi **ty**!* or in contrasting themes as in *Ty₁ schodi za produktami, ty₂ uberi komnatu, a ty₃ vozmis' za domašnie zadanija!* 'You₁ go do the shopping, you₂ clean the room, and you₃ get down to your homework'.[3] Are these contexts really impossible with the 1PL.IMP? Our corpus does not contain any examples that would allow for either a contrasting or a focalised interpretation of the type *Davajte **my** pojdem!* or *Davajte pojdem **my**!,* but native speakers accept such examples.

In this way, we face a rather messy situation: most likely, any corpus analysis will provide a large amount of ambiguous contexts. What could then be the additional contextual cues that would impose or at least facilitate an unequivocal interpretation? First of all, the **referential potential** of the first person plural calls for closer scrutiny (Norman 2002, Szymański 1990). Its systematic account should not only involve politeness and the number of addressees, but also allow for both the exclusive reading 'speaker + > 1 person not addressed (e.g. being absent)' and the cumulative reading 'speaker + > 1 person not addressed + > 1 addressee'. This amounts to a total of 14 possible combinations.[4] Among them, the exclusive case is not relevant for our purpose since 1PL.IMP presupposes the existence of at least one addressee. Therefore, if a given context triggers the exclusive reading we may be sure that we are dealing with FUT1PL. Moreover, the number of non-addressed persons within the cumulative reading does not matter. The 8 remaining cases, however, are still ambiguous between FUT1PL and 1PL.IMP. Unfortunately, the limited length of the excerpts from the Russian National Corpus seldom allows for an exact assessment of the referential situation. This holds in particular for the exclusive reading, which imposes the indicative interpretation of the utterance. Only by means of additional adverbs such as *vmeste* 'together' can the exclusive reading be ruled out, cf. *Pojdem použinaem vmeste* 'Let's go have dinner together'.

And finally, due to pragmatic reasons such as empathy or baby talk the reference of the first person plural may even be shifted on the addressee by excluding the speaker (Weiss 2007, 372-375). This is illustrated by the following example:

(3) *Davaj **pomoemsja** / golovku pomoem // Igruški moet / net čtoby samoj myt'sja! Xolodnaja / net / Svet?* (Zemskaja/Kapanadze 1978. 248)

---

[3] The subject pronoun may also be found in other contexts, such as magic spells, cf. *Pojdi ty, xvor', vo čisto pole…* 'Go, illness, to the wide field'.

[4] As is well known, the inclusive interpretation may be made explicit by means of *my s toboj / my s vami*. This marker is, however, associated with the indicative reading.

'Let's wash / let's wash your head // [she] is washing her toys instead of washing herself! Cold [water] / is it / Sveta?'

The initial reflexive *pomoemsja* and the subsequent *myt'sja* indicate that the real subject is the child herself, i.e. the addressee.

Another cue is offered by **verbs denoting speech acts**. A fragment of direct speech introduced by *vzmolilas'* 'she begged', *ja tebja umoljaju* 'I implore you' or followed by *prošu tebja* 'please' or *sdelaj odolženie* 'do me a favour' imposes the imperative reading. In a similar vein, a string like *pojdem potancuem, – priglasil* on 'Let's go [and] dance, – he invited [her]' prevents the indicative reading. Even the verb *skomandovat'* 'order' was found: *Pojdem otdyšimsja, – skomandovala Svetka* 'Let's go [and] take a breath! Svetka ordered'. The perlocutionary verb *ugovorit'* 'persuade, argue into' likewise describes a request, not a statement, therefore it refers to utterances in the 1PL.IMP, cf.:

(4) *Мне удалось **уговорить** ее зайти со мной в этот парк погулять. — Бабонька, пойдем погуляем чуть-чуть в парке! Я там никогда не был!* [Павел Санаев. Похороните меня за плинтусом (1995) // "Октябрь", 1996]
*Mne udalos **ugovorit'** eje zajti so mnoj v etot park pogulat' – Babon'ka, pojdem pogulajem čut'-čut' v parke! Ya tam nikogda ne byl!*
'I managed to persuade her to go for a walk to this park. – Granny, let's go for walk a little bit in the park! I've never been there!'

This list can be continued: a context such as *Brat zovet: "Pojdem pokurim s mužikami na dvore"* invites the imperative interpretation 'My brother calls / invites me: "Let's go have a smoke with the guys in the court'. The request may be indicated by a gesture, cf. *On menja manit pal'cem: Pojdem sxodim v bar* 'He attracts me with his finger: Let's go to the bar'. Other verbs admit both readings: for instance, *predložit'* 'propose' combines not only with exclamation marks, but occurs in questions as well, which only admit the FUT2PL reading, cf. *…predložil: – Vyp'ete, otcy?* 'and proposed: What about having a drink?'. Moreover, *predložit'* may refer to quotations with **modal adverbs** that indicate the speaker's hesitation, cf. *Možet, pojdem razberemsja, – predložila Ramil'* 'Perhaps, we should go and figure it out?' This is another unequivocal marker of the indicative reading, since *možet* never combines with imperatives, cf. **Možet pojdi! *Idi, možet!* '*Go perhaps!' The disambiguating effect of modal adverbs can be further illustrated by *naverno* 'probably', cf. also: *naverno pojdem použinaem vmeste* 'probably, we'll go have dinner together'.

Sometimes the **reaction of the addressee** gives a clear indication that he interpreted a given expression as a request: he then may *soglasit'sja* 'agree' or else refuse by *da net, neoxota* 'no, I am not in the mood / I don't feel like this'. In other cases, a 2SG/PL.IMP may prepare the ground for a subsequent 1PL.IMP, cf. *Provetri kak sleduet, nakuril. Pojdem prinesem postel'* 'Air the room properly, you have smoked too much. Let's go fetch the bedclothes'; again, this criterion is rather shaky since the indicative interpretation is not completely excluded.

On the other hand, the illocutionary force of the FUT1PL also merits a closer look: as is well known, in Russian the FUT.PRF2SG/PL may express an indirect directive, thus functioning as an equivalent of the imperative, if the addressee is somehow socially subordinated to the speaker. Such relations hold between a child and their parents, a soldier and an officer, a clerk and a boss, but also a husband and a wife (though not vice versa!); in all these settings, an utterance

such as *Musor vyneseš' potom* lit. 'you'll take the garbage out' would be appropriate. In view of this, one is tempted to interpret certain examples as an instance of FUT.PRF1PL.

To sum up: we have found quite a few contextual cues supporting either the imperative or the indicative interpretation, but many of them have turned out not to be completely reliable in that they may conflict with counterevidence or be overruled by other contextual factors. Thus, our former expectation that a search of double imperatives of the 1st PL in the Russian National Corpus will provide a large amount of ambiguous results still remains valid.

2 Analysis of the Data

2.1 Imperfective vs Perfective Aspect: a Mismatch

Let us now turn to the search for double verbs in the 1PL.IMP in the Russian National Corpus. For reasons of space, the query was limited to verb pairs in immediate contact, i.e. without another word form separating them from each other. As for punctuation, only verb pairs separated by a space or hyphen were considered; the motivation for this may be found in (Weiss 2013). Contrary to the latter study, dashed forms such as *polučim – otdadim* 'if we get X, we'll give X back' were no longer taken into account since most often they mark asyndetic links between two clauses. Moreover, as in (Weiss 2013), pairs of lexically identical verb forms were excluded.

The first thing to note is a striking aspectual asymmetry: the overwhelming majority of all serialised instances of 1PL.IMP were perfective. To begin with, the imperative model *budem igrat'* 'let's play', described as infrequent in grammars, did not provide one single extended example of the type *budem pet' igrat'* 'let's sing [and] play' in the basic corpus of the RNC. In this connection, it should be emphasised that such constructions with double infinitives are structurally ambiguous in that the second infinitive may be governed by the first. This turned out to be the case in nearly all of the examples. The query with "V,imper,pl,1p,ipf na rasstojanii 1 от V,inf,ipf -bmark na rasstojanii 1 от V,inf,ipf –lexred" led to 115 results, but only one of them represented a double verb (*Zavtra my budem dumat' obsuždat'* 'tomorrow we'll think [and] consider', and this is an obvious instance of FUT1PL. The search for imperatives with *davaj(te)* + double infinitive, which is considered more frequent than the variant with *budem*, was even less successful: none of the 24 instances represented the serial construction. The same holds true for the negated variant *davajte ne budem*.

The search for perfective double imperatives, however, provided 521 hits that met the query "V,imper,pl,1p,(pf | ipf) -bmark na rasstojanii 1 от V,imper,pl,1p,pf". Among them was only a small amount of noise (20 excerpts), e.g. *umrem uvidim* 'when we'll die, we'll see'. Before going into the details, it should be mentioned that 27 hits consisted of the IPF form *idem(te)* 'we go / let's go' and a PF second verb. Most of these examples can be considered instances of 1PL.IMP. This is perfectly in line with the findings on double verbs in the 2PL.IMP (Weiss 2013), where *idite* 'go.PL!' occurred in 33 pairs with mixed aspect. Moreover, our collection contains a double imperative with the IPF *bežim* 'let's run' as first component.

2.2 The Homonymy ꜰᴜᴛ1ᴘʟ vs. 1ᴘʟ.ɪᴍᴘ

Let us first filter out clear instances of ꜰᴜᴛ1ᴘʟ. To begin with, the collection comprises 8 cases of the idiom *poživem uvidim* lit. 'we'll live we'll see', an epistemic marker meaning approximately "may be, probably". This idiom is usually written with a dash (174 hits). The next criterion is the subject pronoun *my* which was present in 22 examples. The question mark followed the double verb forms in 33 examples, two of them with the IPF *idem* as the first verb. The modal adverb *možet* 'perhaps' appeared in 6 examples not yet included in the group with a question mark, *navernoe* 'probably' in 1. All three groups together amount to a total of 70 clear instances of the indicative meaning.

The remaining 431 excerpts were then checked for evidence in favour of the imperative interpretation. The verb forms were preceded by 11 instances of *davaj* and 1 instance of *davajte*. 20 of the first imperatives had the postfix *-ka*, another one had *davajte-ka*. The particle *nu-ka* only occurred together with the explicit markers *-te* and *predložila.* These figures should now be summed with the results of a separate search for the postfix *–te,*[5] which was found with 53 first imperatives. Additionally, in 4 examples both imperatives were marked this way; three of them were by I. S. Turgenev (mid-nineteenth century). The most recent one is the following:

(5) ***Идемте купимте*** *торт и отпразднуем это событие.* [К. К. Вагинов. Труды и дни Свистонова (1928-1929)]
Idemte kupimte tort i otprazdnuem eto sobytie.
'Let's go let's buy a cake and celebrate this event'.

The third verb form *otprazdnuem* is again ambiguous: it can represent either ꜰᴜᴛ1ᴘʟ or 1ᴘʟ.ɪᴍᴘ. The indicative reading seems to be more convincing, although our data contains a whole range of obvious triplets with the third verb coordinated with the second of the type *davaj sjadem, obsudim i rešim* lit. 'let's sit down, discuss and decide' or *pojdem prisjadem na skamejku, pogovorim* 'let's go have a seat on the bench [and] talk'. The only real triplet is the following:

(6) *Пойдем за сарай ляжем полежим, душа опять болит.* [А. П. Платонов. Счастливая Москва (1936)]
Pojdem za saraj l'ažem, poležim, duša op'at' bolit.
'Let's go behind the barn, lie down [and] lie there a bit, the soul is aching again'.

And finally, 29 double verbs were followed by an exclamation mark without additional markers of the 1ᴘʟ.ɪᴍᴘ reading and without evidence to the contrary.

A lexical indicator is the pragmateme (Iordanskaja & Mel'čuk 2007, 305 ff., 311) *pojdem vyjdem* 'let's go outside!', which (much as its English counterpart) functions as a request to settle a conflict through a physical fight and occurs only in the imperative. It was attested 12 times.

Further lexical cues are offered by the speech act verbs mentioned in the previous section. Not less than 29 contexts with *predložit'* / *predlagat'* 'propose' either in the previous or the subsequent sentence and without the additional marker *-ka* invite the 1ᴘʟ.ɪᴍᴘ reading. The

---

[5]   In the RNC these forms are denoted by a separate category called "imperativ 2"

remaining verbs provide another 6 contexts with 1PL.IMP, 3 contexts contained the verb *(po)zvat'* 'call, invite' pointing in the same direction. So far, our total of double imperatives with explicit markers sums up to 169 instances. Moreover, 17 preceding contexts exhibited a verb in the 2nd SG or PL imperative; however, as mentioned above, this is a rather shaky indicator of the 1PL.IMP reading.

The following table summarises these results (recall that double coding such as *-te + -ka* was counted only once):

N = 558

marker	1PL.IMP	FUT1PL
davaj(te)	12	
-ka	21	
-te	57	
pojdem vyjdem	12	
speech act verbs	38	
exclamation mark	29	
total	169 = 30,3%	
after 2SG/PL.IMP	(17 = 3%)	
subject pronoun		22
poživem uvidim		8
modal adverbs		7
question mark		33
total		70 = 12,6 %
unambiguous contexts	238 = 42,9%	

The ratio of 1PL.IMP increases only insignificantly if we take into account additional context features, for instance subsequent occurrences of verbs denoting agreeing or disagreeing. The resulting total of much more than 50% ambiguous cases is, of course, less than satisfactory. On the other hand, one may argue that the users are probably not very concerned about this type of ambiguity. In particular, the hearer of a given utterance could not care less whether it is formulated in the imperative or as a question as long as both variants convey the meaning of the same speech act, e.g. a proposal to do something together. Moreover, it seems a fair

assumption that the default interpretation of the construction under scrutiny is the 1PL.IMP. The following observations will corroborate this view.

2.3   Lexical Restrictions

If we now turn to the **lexical composition** of our data, we face a strikingly monotonous picture. It should be recalled that the first component (V_1) of a prototypical double verb construction is lexically more restricted than V_2 (Weiss 2012, 615ff). This holds in particular for the double 1PL.IMP: in 458 cases **V_1** is *pojdem(te)*, its ipf counterpart *idem(te)* is attested in 27 cases, in 30 instances we find another motion verb out of the following list: *zajti* 'come in', *vyjti* 'come out', *prijti* 'come [by foot]', *sxodit'* 'go [by foot]', *poxodit'* 'walk around', *poexat'* 'go [on a vehicle], *zaexat'* 'come [for someone/something, on a vehicle]', *bežat'* 'run', *sbegat'* 'fetch'. Two inversions of V_1 and V_2 are attested, viz. *posmotrim pojdem* 'let's watch-go' and the following one:

> (7) *... предложил Николаю: – Может, пропустим пойдем по стопарику? У меня взято. Николай отказался.* [Роман Сенчин. Елтышевы (2008) // "Дружба Народов", 2009]
> *...predložil Nikolaju: – Možet, propustim pojdem po stopariku? U menya vzyato. Nikolaj otkazalsya.*
> '...he proposed to Nikolaj : – Perhaps, [lit.] down go a shot [of vodka]? I have some with me. Nikolaj refused.'

In other words: 516 examples or **92,4%** of the total represent verbs of **physical motion**, out of which 486 or 87% belong to the basic verb *pojti / idti* 'go'. Among them there were no cases of semantic bleaching comparable to imperatives of the 2ⁿᵈ PL such as *podite pojmite ego* lit. 'go understand him' = 'It's impossible to understand him'.

The remaining part comprises 22 instances of 5 posture verbs (*sest'* 'sit down' (7), *leč'* 'lie down', *vstat'* 'get up', *posidet'* 'to sit for a while' (11) and *postojat'* 'to stand for a while' (2)). To this should be added the 7 instances of *poživem,* see above. Only 4 other verbs are involved : *poprobovat'* 'try', *podoždat'* 'wait', *sobrat'* 'collect' and *pocelovat'* 'kiss'.

Said lexical constraint helps to explain why we did not find a single instance of the **negation** before or in-between the two verbal components:[6] it would be fairy senseless to announce or ask sb. to go somewhere just for doing nothing. In this regard, the 1ˢᵗ PL differs radically from the 2ⁿᵈ PL (see next section).

As for the second component **V_2**, its inventory exhibits a greater diversity. However, the main groups are not very variegated: we found 146 uses of a perception verb,[7] 121 uses of another motion verb including *vyjdem* as in the above-mentioned pragmateme *pojdem vyjdem* 'let's go outside', 48 uses related to speech activities (*speak, discuss, ask, invite, pray, say goodbye* etc.), 22 uses of posture and related verbs (e.g. *sprjačemsja* 'let's hide'), 17 uses denote

---

6   To be precise, there was one dialectal use of FUT1PL illustrating a folkloristic subtype of double verbs, cf. *Lačim ne ulačim* approximately meaning 'we are hugging - won't [can't] hug enough' (V.Apresjan, p.c.).

7   Here are the detailed figures: *posmotrim* 'let's / we'll see' – 112, *poslušaem* 'let's listen' – 8, *uvidim* 'we'll see' – 8, *pogljadim* – 5, *gljanem* – 4, *vzgljanem* – 3, *zagljanem* – 2, other – 4. As this ranking demonstrates, visual perception prevails, amounting to no less than 92%.

various physical activities, e.g. picking fruit, cutting grass, weighing maize or freeing prisoners, and 10 refer to mental operations.

As can be seen, most possible combinations of V_1 and V_2 easily lend themselves to the imperative interpretation. On the other hand, it cannot be denied that in many pairs of motion verbs the first verb *pojdem* is almost semantically empty; with the imperative meaning, it thus is turning into another marker of the directive illocution, much as *davaj(te)*. Such a development is of course not exclusively characteristic of Russian; motion verbs such as 'go' or 'come' tend to desemanticise in many languages all over the world, see (Majsak 2005).

Moreover, we are now in the condition to state that all examples found represent the **prototypical** double verb construction, where one verb semantically modifies the other. In terms of the typology of verb serialization sketched out by (Aikhenvald 2006), this corresponds to the asymmetric type of VS. Non-prototypical DVs, or else: semantic twins (Weiss 2012, 625-632) of the type *est'-pit'* 'eat-drink', *spat'-počivat'* 'sleep-rest', *stirat'-gladit'* 'wash-iron', *razdevat'-odevat'* 'undress-dress' etc., where the two components are on a par, simply do not occur. This again contrasts with the picture to be found with 2nd PL imperatives, which will now be briefly characterized. Before tackling this subject, however, a brief remark on the impact of our observations on the syntactic representation of double verbs in a MTT framework may not be out of order. In (Weiss 2000) we argued that in most cases the natural solution would be to posit a dependency relation $V_1 \rightarrow V_2$, and not vice-versa. The findings of the present paper point in the same direction: if more than 90% of the total contain a motion verb that may take an infinitive as second actant, one might consider whether in the serial construction V_2 simply occupies this valency, cf. *Pojdem → poguljat'* vs. *pojdem → poguljaem.*[8]

3   The Contrast: a Brief Glance at Double Imperatives in the 2nd PL

The results to be presented here are discussed in more detail in (Weiss 2013). That paper covered a comparable number of data (533 instances), but unlike the present study, it also included uses with one intermediary component, e.g. *smotrite ne prostudites'* 'watch out – don't catch cold', *ver'te ne ver'te* 'believe it or not' or *rasskazyvajte svjazno davajte* 'tell the story in a coherent way'. The main contrasts concern the following criteria: (i) lack of homonymy, (ii) negation, (iii) inversion of V_1 and V_2, (iv) lexical composition, (v) desemanticisation of V_1 and (vi) aspectual properties.

Strings of the type *pojdite sprosite* 'go.PL ask.PL' are not subject to systematic grammatical homonymy, as this was the case in the data examined in the present study. Negation, which has been shown to be completely absent from the 1PL of double verbs, affected 12% of the 2PL.IMP; its scope exclusively encompassed V_2. Inversion of V_1 and V_2 occurred in 5% of the cases; the corresponding figure in the present study is 0,4%. Lexical composition reveals several contrasts. First of all, not all instances represented the prototype: symmetric pairs such as *pejte-eš'te* 'eat drink', *izvinite-prostite* 'excuse-apologise', *zdravstvujte-proscajte* 'welcome goodbye' constituted 5% of the total. As for the meaning of V_1, verbs of motion provided

---

[8]   In redundant uses, such as *pojdem vyjdem*, the infinitive alternative would, however, not work, cf. *pojdem*vyjti*.

55% of all uses with immediate contact of V_1 and V_2, but much less so in distant position. Other typical representatives of V_1 are *davajte, voz'mite* 'take', 'do suddenly', *poprobujte* 'try' and several politeness markers. Contrary to the picture obtained with 1.PL, where we only have two idioms *(poživem uvidim, pojdem vyjdem),* with 2PL.IMP we find quite a few set phrases, including a syntactic phraseme (Mel'čuk 1995), viz. *smejtes' ne smejtes'* 'laugh.PL don't laugh.PL' and phraseologised V_1s such as *bud'te dobry* 'be so kind' or *sdelajte odolženie* 'do [me] a favor'. Moreover, unlike with 1.PL, many V_1s tend to desemanticise or rather acquire a new pragmatic meaning: such is the case of *podite* as a marker of rhetorical directives, *smotrite* as a marker of warnings, but above all *davajte* which functions as a mere signal of the directive illocution comparable to *hajde(te)* in South Slavic languages. And finally, double 2PL.IMP forms are not bound to PF aspect, but allow for many IPF verbs; this leads to a large portion of aspectually mixed pairs (26%). Most of them (89%) show the pattern $V_{1ipf} + V_{2pf}$, which is due to the high rate of aspectually unpaired and simultaneously desemanticised V, such as *davajte, smotrite, bud'te (dobry).* On the whole, we may thus conclude that 2pl.imp exhibits much more structural diversity than 1.PL regardless of the two readings (FUT vs. IMP) of the latter. They share, however, one common verb form *(davajte)* which functions as V_1 with 2PL.IMP and an additional marker of the imperative reading with 1PL and one common tendency: V_1 tends to undergo semantic bleaching and to eventually acquire new pragmatic meanings.

Bibliography

Aikhenvald, A.Y. 2006. Serial Verb Constructions in Typological Perspective. In Aikhenvald A.Y.& R.M.W. Dixon (eds.): *Serial Verb Constructions.* A Cross-linguistic Typology. 1-68. Oxford: University Press (Explorations in Linguistic Typology).

Iordanskaja, L.N. Gardiner & I.A. Mel'čuk. 2007. Moskva: Jazyki slavjanskix kul'tur.

Isačenko, A. 1975. *Die russische Sprache der Gegenwart. Formenlehre.* München: Max Hueber

Majsak, T. 2005. *Tipologija grammatikalizacii konstrukcij s glagolami dviženija i glagolami pozicii*. Moskva: Jazyki slavjanskix kul'tur.

Mel'čuk, I.A. 1995. *The Russian language in the meaning—text perspective.* Moskau-Wien: Wiener Slaiwstischer Almanach, Sonderband 39.

Norman, B Ju, 2002. Russkoe mestoimenie *my:* vnutrennjaja dramaturgija. *Russian linguistics* 26(2). 217-234.

Szymański. M, 1990. Znaczeniowa kategoria osoby. In: *Studia z Filologii Polskiej i Stowiańskiej* 26. 75-82.

Weiss, D. 2000. Russkie dvojnye glagoly: kto xozjain, a kto sluga? In L.L.Iomdin, L.P. Krysin (ed.). *Slovo v tekste i v slovare. Sbornik statej k semidesjatiletiju akademika Ju. D. Apresjana.* 356-378. Moskva: Jazyki slavjanskix kul'tur.

Weiss, D. 2003. Russkie dvojnye glagoly i ix sootvetstvija v finnougorskix jazykax. *Russkij jazyk v naučnom osvescenii,* 6(2): 37-59.

Weiss, D. 2008. Vremennaja sootnesennost' dvojnych glagolov soveršennogo vida. In *Dinamičeskie modeli: Slovo. Predloženie. Tekst. Sbornik statej v čest' E.V. Padučevoj,* 155-177. Moskva: Jazyki slavjanskix kultur.

Weiss, D. 2012. Verb serialization in North East Europe: the case of Russian and its Finno-Ugric neighbours. In B. Hansen & B. Wiemer & B. Wälchli (eds.), *Grammatical replication and grammatical borrowing in language contact.* 619-655. Berlin-Boston: Walter de Gruyter Mouton.

Weiss, D. 2013. Dvojnye glagoly v russkoj razgovornoj reči v zerkale Nacional'nogo korpusa. Formy imperativa mn.č. In E.Vel'mezova (ed.), *Schweizerische Beiträge zum XV. Internat. Slavistenkongress in Minsk,* August 2013, Bern-Frankfurt: Peter Lang.

von Waldenfels, R. 2012. *The grammaticalization of 'Give' + Infinitive. A comparative study of Russian, Polish and Czech.* Berlin New York: De Gruyter Mouton.

Zero Subjects in Active and Passive Sentences

Anton Zimmerling

Institute of Modern Linguistic Research,
Moscow

Abstract

The paper discusses syntactic constructions with zero subject, focusing on the data from three languages: Russian, Ukrainian and Modern Icelandic. Zero subjects are defined in the paper, according to the MTT tradition, as symbolic objects with an empty signifier and non-empty signified. Zero subject lexemes have role-and-reference semantics of their own and their meaning is not equivalent to the meaning of any non-zero signs in the same language. Zero subjects share some syntactic properties with overt sentence elements taking the subject position in the same language; notably, they act as agreement controllers on the verb and can be case-marked in languages where overt grammatical subjects are case-marked and control verbal agreement. The working hypothesis raised in the present paper is that zero subjects in Russian, Ukrainian and Icelandic are always specified as non-referential Agents {+ Agent, – referential}, while the specification of zero subjects as '+Human' or '– Human', as well as their agreement morphology (singular, plural form), can vary across languages.

Keywords

Syntax, agreement, control, case, zero lexemes, subjects, semantic roles, argument structure

1 Zero Subjects in Standard Active and Passive Structures

1.1 Active Sentences

1.1.1 *Zero Generic Subjects of Verbal Sentences*

The paper discusses syntactic constructions with zero subject, with focus on data from 3 languages – Russian, Ukrainian and Modern Icelandic. Zero subjects are defined in the paper according to the MTT tradition (Mel'čuk 1995) as symbolic objects with an empty signifier (i.e. an empty string of phonemes – **Ø**) and non-empty signified. Zero subject lexemes have role-and-reference semantics of their own and their meaning is not equivalent to the meaning of any non-zero signs in the same language. At the same time, zero subjects share some

syntactic properties with overt (i.e. non-zero) sentence elements taking the subject position in the same language, notably, they act as agreement controllers on the verb and can be case-marked in languages where overt grammatical subjects are case-marked and control agreement features of the verb (Zimmerling 2009). The working hypothesis raised in the present paper is that zero subjects, at least in Russian, Ukrainian and Icelandic, are always specified as non-referent Agents {+ Agent, – referential} while the specifications of zero subject as '+Human' or '–Human' as well as their agreement morphology (singular, plural form) vary across languages.

In Russian, Ukrainian and Modern Icelandic, zero subjects have generic semantics and are non-referential Agents i.e. typically have a feature matrix {+ Agent; – Referential}. I am adopting the version of the event structure analysis where the notion of Agent of a transitive process is generalized both for events with an uncontrolled process / activitity caused by a non-Human factor, cf. Russ. *Ulic-u* Acc.Sg.F *zasypal-o* Prt.3Sg.N *peskom* Instr.Sg. 'The street got covered with sand <due to the impact of a natural force>' and for events denoting activities which may be caused and controlled by human beings, cf. Russ. *Ulic-u* Acc.Sg.F ***nazlo žitelyam₁ naročno₂ zasypal-i*** Prt.3Sg.N *peskom* Instr.Sg 'The street has been **intentionally₂** covered with sand <by some people acting with a malicious intent> **to spite the street's residents₁**'. The data of Russian, Ukrainian and Icelandic zero subject constructions seem to support the broad definition of Agent as a semantic role, since zero Agents specified as {+ Human; – referential} and {– Human; – referential} have similar syntactic properties. In Russian and in Ukrainian active sentences with non-referential Agents specified as {+Human} are marked by the use of the 3rd person plural, cf. *Ulič-u* Acc.Sg.F *zasypal-i* Prt.3Sg.N *peskom* Instr.Sg. In Russian linguistics this sentence pattern is called 'indefinite-personal constructions' (Russ. *'neopredelenno-ličnye predloženiya'*). In a framework where zero subjects are not recognized, this is either a descriptive tag or backing of a hypothesis that Russian sentences in the 3rd person plural signaling a generic meaning do not project a position of an overt subject marked by the nominative case. In the MTT tradition, Russian 'neopredelenno-ličnye predloženiya' are analyzed as zero subject constructions i.e. as structures projecting a subject position filled by a zero sentence element. Mel'čuk (1995) in his seminal work on zero lexemes identifies the generic subject of Russian sentences with the 3rd person verb form as a zero lexeme Ø^People with the role-and-reference features {+ Human, – referential}. Zimmerling (2007) interprets the element Ø^People in Russian as a 3rd person zero pronoun in the nominative case, plural with role and-reference characteristics of a generic Agent: the complete array of features for this element is {+ Agent, + Human, – referential, 3rd person, + plural, + nominative case}.

1.1.2 *Transitive Impersonals*

Russian sentences with verbs in the 3rd person singular denoting uncontrolled processes and lacking overt nominative subjects are traditionally called impersonals and are analyzed as structures that do not project any subject position. This kind of analysis is standard for the Russian linguistic tradition and adopted in some versions of generative syntax, cf. Babby (2002) who straightforwardly treats all Russian and Ukrainian impersonals as subjectless. A special focus of interest has been made on transitive impersonals like Russ. *Ulic-u* Acc.Sg.F *zasypal-o* Prt.3Sg.N *peskom* Instr.Sg since such constructions falsify a postulate of Chomsky's grammar known as Burzio's generalization and wrongly predicting that only verbs with nominative subjects assign accusative case to their objects, cf. (Reuland 2000). Although the validity of Burzio's generalization has been challenged and there are obvious empiric counterexamples

to it, cf. the discussion in (Woolford 2003), a number of attempts have been made in the recent years to justify the existence of transitive impersonals and save the initial assumptions of Chomsky's syntax. Sigurðsson (2011) proposed an "accusative-of-fate theory": if an Icelandic sentence is about uncontrolled events, the predicate becomes defective (in phrase-structural terms, gets a defective VP) and does not project a subject element (called "internal argument" in Chomsky's Minimalist Program) but can still assign structural accusative case to its object. Likewise, Svenonius (2002) and Richardson (2007) stipulate that Icelandic and Russian transitive verbs only license transitive impersonal constructions in certain contexts, presumably linked with telic readings and projecting a special layer of VP called Aspect Phrase. The main drawback of such theories is that they treat transitive impersonals as a deviation from the principles of Universal Grammar and claim that transitive impersonals are only possible in certain deviating contexts. In addition, the authors of such theories are prone to stipulating such ad hoc constructions as "accusative-of-fate", "accusative of nausea", "accusative of diarrhea", etc., which is not an economic solution[1]. MTT offers a simple and elegant solution. Mel'čuk (1995) identifies the generic subject of Russian transitive impersonals with the verb in 3rd singular **as a zero lexeme** $\emptyset^{Elements}$ with the role-and-reference features {– Human, – referential}. Zimmerling (2007) interprets the element $\emptyset^{Elements}$ in Russian as a 3rd person zero pronoun in the nominative case, singular with a role-and-reference characteristics of a generic Agent: the complete array of features for this element is {+ Agent, – Human, – referential, 3rd person, – plural, + nominative case}. A similar analysis in phrase-structural terms has been proposed by Lavine & Freidin (2002) who, however, do not acknowledge syntactic zeroes as grammatical subjects.

1.1.3   An MTT Approach to Zero Subjects in a Typological Perspective

Mel'čuk's (1995) analysis of zero subjects in Russian can (with slight modifications) be used for Ukrainian, Icelandic and other European languages with similar morphosyntactic conditions and nominative case marking on the grammatical subject. In these languages, most verbs licensing transitive impersonals also license structures with an overt nominative subject, cf. Russ. *zasypat'* 'to cover', 'to charge', *pronesti* 'to carry' or 'to have diarrhea', Icel. *reka* 'to drive'. Zimmerling (2007) argues that Old Slavic and Germanic languages had generic zero subjects with other roles than Agent, but Modern Icelandic and Ukrainian zero generic subjects are always Agents. All three languages have transitive impersonals with verbs assigning accusative case. Icelandic also licenses zero subjects with verbs assigning structural dative case to their object, cf. (1). Impersonal verbal structures like (1) are active and denote uncontrolled processes while impersonal structures with non-agreeing participle II and a copula in 3rd person singular are passives and denote results/activities controlled by a human Agent, cf. the contrast of (1) and (2).

(1)   Icel. *Bátu-num* $_{Dat.Pl.Det}$ $\emptyset$ $_{3Sg}$ *hvolf-di* $_{Prt.3Sg}$ **viljandi* $_{PartI}$
      'The boats capsized *on purpose'.

(2)   Icel. *Bátu-num* $_{Dat.Pl.Det}$ $\emptyset$ $_{3Sg}$ *var* $_{Prt.3Sg}$ *hvolf-t* $_{PartII.Sg.N}$ *viljandi* $_{PartI}$
      'The boats have been turned down on purpose'.

---

[1]   For the criticism of Svenonius and Richardson's "Aspectual theory", from the positions of Minimalist syntax, see (Lavine 2012)

It is apriori unclear whether (1) and (2) have one and the same kind of zero subject or two different ones. In this paper we opt for the first analysis and postulate only one type of generic subjects invariably associated with the 3rd person singular and having the features {+ Agent, – referential}. The specification '+ Human' depends on the voice – in active sentences only {–Human} subjects are possible, in passive sentences – only {+ Human}. Note that Icelandic lacks 'neopredelenno-ličnye predloženiya' of the Russian type. Modern Icelandic does not have zero subjects associated with the 3rd person plural, while generic human subject is normally expressed by an overt indefinite pronoun *maðr* 'one'[2] in the Nominative case singular. I conclude that the 3rd person form in Icelandic is linked both with generic non-human Agents, cf. (1) above, and with generic human Agents, cf. (2).

Standard Icelandic does not favour impersonal passives from transitive and ditransitive verbs (i.e. verbs like *give* take both accusative and non-accusative objects), though colloquial Icelandic reportedly has these constructions, cf. (Sigurðsson 2011). On the contrary, transitive impersonal passives from accusative verbs are grammatical in standard Ukrainian, cf. (3) which is a structure with a zero copula 'be' in the present indicative[3].

(3) Ukr. *Oficeriv*$_{Acc.Pl}$ Ø$_{3Sg}$ *zalyaka-n-o*$_{Part.3Sg.N}$ Ø$_{3Sg}$ *zaturka-n-o*$_{Part.3Sg.N}$, Ø$_{3Sg}$
zaklbova-n-o$_{Part.3Sg.N}$, *usi* $_{Nom.Pl}$ *robl'at'*$_{Prs.3Pl}$ *use i vodnočas ne robl'at'* $_{Prs.3Pl}$ *ničogo*.
'The officers are <u>bullied</u>, <u>scared</u> and <u>cowed</u>, all of them <u>do</u> everything and at the same time <u>do</u> nothing'

The main puzzle with Ukrainian data is that Ukrainian also has active 'neopredelenno-ličnye' constructions of the Russian type, with the verb in 3rd person plural, cf. (4).

(4) Ukr. *Oficeriv*$_{Acc.Pl}$ Ø$_{3Pl}$ <u>*zalyaka-l-i*</u>$_{Prt.3Pl}$ Ø$_{3Pl}$ <u>*zaturka-l-i*</u>$_{Prt.3Pl}$, Ø$_{3Pl}$ <u>*zakhova-l-i*</u>$_{Part.3Sg.N}$.
'The officers are <u>bullied</u>, <u>scared</u> and <u>cowed</u>'

The active (3) and passive (4) construction with a generic zero Agent seem to be synonymic in Ukrainian and to signal the same value {+ Agent, + Human, – referential}. Therefore, in spite of the fact that Ukrainian, unlike Icelandic, has generic zero Agents associated both with 3rd person singular and 3rd person plural, the singular/plural distinction is of less importance for this language than for Russian. Russian which lacks impersonal transitive passives from accusative verbs has an unambiguous mapping of role semantics {+ Human Agent} and agreement features of zero subject (3rd person singular/plural).

Human generic Agent	Non-human generic Agent
3rd person plural (active sentences only)	3rd person singular (active sentences only)

Table 1: Mapping of role semantics and agreement features of zero subjects in Russian

---

[2] This pronoun is a grammaticalized form of the noun 'man' in the nominative case, singular, indefinite form. Remarkably, Old Icelandic had a half-grammaticalized form of the same noun – *menn* in the nominative case, plural, indefinite form. The O. Icel. *menn* is an overt counterpart of the Modern Russian Ø^{3Pl}.

[3] For the reasons of space I do not gloss the zero copula "be" in Russian and Ukrainian examples.

Ukrainian has an across-the-voice synonymy of 3rd person singular and plural. In addition, the 3rd person singular form in zero subject constructions remains two-way ambiguous between {+ Human Agent} and {– Human Agent} constructions, just as in Icelandic.

Human generic Agent	Non-human generic Agent
3rd person plural (active sentences)	3rd person singular (active sentences)
3rd person singular (passive sentences)	

Table 2: Mapping of role semantics and agreements features of zero subjects in Ukrainian

2 Zero Subjects in Non-Standard Structures

According to the hypothesis of Zimmerling (2012), Russian has oblique subjects marked with the dative case at least in two types of constructions – with an infinitive (*emu eto **ne osilit'*** 'he **cannot manage** that') and a nominal non-agreeing predicative (*emu bylo **xolodno*** 'he was **cold**'). The verification or falsification of this theory is beyond the scope of this paper. In any case, it is obvious that even if the oblique subject hypothesis is wrong, these two dative constructions cannot be explained in terms of Mel'čuk's zero subjects $\emptyset^{People}$ and $\emptyset^{Elements}$, since these predicates lack an Agent argument, and the core argument of dative-predicative structures like *emu bylo **xolodno*** is an Experiencer, not a Patient.[4] Therefore, it is important to check whether $\emptyset^{People}$ and $\emptyset^{Elements}$ (in our notation – $\emptyset^{3Pl}$ and $\emptyset^{3Sg}$) combine with the tentative dative subjects. We did not find any instances of a combination '$\emptyset^{3Pl}$ + subject-like dative element' but the combination '$\emptyset^{3Sg}$ + subject-like dative element' was found in two relatively rare and obscure Russian constructions.

2.1 Ditransitive Active Impersonals in Russian

Ditransitive impersonals of the first type are the active constructions with $\emptyset^{3Sg}$, an overt Patient argument in the Accusative case, specified as {– Human} and an overt Experiencer/External Possessor argument in the Dative case, specified as {+ Human}, cf (5) and (6). Traditional grammar explains such sentences as subjectless since the verbal form does not have any overt controller and stands in 3rd person plural, neutrum[5]: the alternative is to analyze the dative element specified as {+ Human} as a subject. However, the main predicate is a verb, not a nominal predicative, and the accusative argument can be easily interpreted as a Patient. Therefore, we have the reasons to postulate a zero Agent subject $\emptyset^{3Sg}$ ($\emptyset^{Elements}$ in Mel'čuk's original notation).

---

[4]   One can certainly postulate for Russ. *emu bylo **xolodno*** and similar sentences in Ukrainian and Icelandic a zero subject with the role of Stimulus ($\emptyset^{Stimulus}$), but this solution does not have any advantages over for the Oblique subject hypothesis or Babby's (2002) analysis of such sentences as subjectless.

[5]   In the past tense.

(5) Russ. *Emu* $_{3Sg.DatM}$ {+ Human} *nogu* $_{Acc.Sg.F}$ {− Human} Ø$_{3Sg}$ *sve-l-o* $_{Prt.3Sg.N}$
'He got a cramp in his leg'.

(6) Russ. *Emu* $_{3Sg.Dat.M}$ {+ Human} *pam'at'* $_{Acc.Sg.F}$ {− Human} Ø$_{3Sg}$ *otšib-l-o* $_{Prt.3Sg.N}$
'He had a lapse of memory'.

The semantic role of the dative argument in (5) and (6) is ambiguous between the values of 'Recipient' and 'Possessor'. Anyway, from a formal viewpoint, this element is an indirect object of an impersonal verb, so the oblique subject hypothesis does not make sense here.

2.2 Ditransitive Middle Impersonals in Russian

Ditransitive impersonals of the second type are middle constructions where the verb has a reflexive marker *-s'ya/ -s'*. Russian verbs on *-s'ya/ -s'* derived from ditransitives (cf. *pokazat'* 'to show something to someone') and taking dative elements specified as {+Human} are of course common, but only a few verbs on *-s'ya/ -s'* reconstruct the initial event structure described by the non-reflexive ditransitive. For instance, Russ. *nasypat'* 'to fill something with something', 'to pile up' underlies the reflexive verb *nasypat'sa* 'to be scattered over' that occasionally retains the dative valency, cf. (7), while Russ. *pokazat'* 'to show something to someone' does not bring about the same effect: in the *pair pokazat'* 'show' → *pokazat'sa₁* 'seem'[6], 'think' the reflexive verb does not describe any situation 'X thought that Z' derived from the situation 'X showed Z to Y', cf. (8a-b). In other words, the role of the Experiencer is not inherited by any Russian reflexive verb taking a dative argument even if the correlative non-reflexive verb takes a dative argument too. On the contrary, the role of the Recipient marginally can be inherited from a correlative non-reflexive verb, see (7).

(7) Russ. *Mne* $_{1Sg.Dat}$ {+ Human} *daže v karman* $_{Prep.Acc}$ Ø$_{3Sg}$ *nasypa-lo-s'* $_{Prt.3Sg.N.Refl}$
'I got it <the snow> even in my pocket. <due to some uncontrolled process>'[7]

(8) Russ. a. *Katya* $_{Nom.Sg.F}$ *pokaza-l-a* $_{Prt.3Sg.F}$ *Vane* $_{Dat.Sg}$ *knigu* $_{Acc.Sg}$
'Kate showed a book to John.'
b. *Vane* $_{Dat.Sg.M}$ *pokaza-lo-s'* $_{Prt.3.Sg.F}$ *čto Katya* $_{Nom.Sg.F}$ *nedovol'na* $_{Adj.Pred.Sg.F}$
'John thought that Kate is displeased.'

Verbs like *navalit'sa* in the meaning 'to be loaded', 'to be piled up', also *nabit'sa* 'to be filled up', *nalit'sa* in the meaning 'to be filled with liquid' behave in the same way as *nasypat'sa* 'to be filled', 'to be scattered over'. All these verbs in the specified meanings describe uncontrolled processes, cf. (9).

(9) Russ. a. *Mne* $_{1Sg.Dat}$ {+ Human} Ø$_{3Sg}$ *nali-lo-s'* $_{Prt.3Sg.N.Refl}$ *v rukav* $_{Prep.Acc.}$
'I got some liquid poured in my sleeve <not because of any intentional activity of any X>'.
b. *Katya* $_{NomSg.F}$ *nali-l-a* $_{Prt.3Sg.F}$ *mne* $_{1Sg.F}$ *židkost'* $_{Acc.Sg.F}$ *v rukav* $_{Prep.Acc.}$
'Kate poured some liquid in my sleeve.'

---

[6] Russian also has a homonymic verb *pokazat'sa₂* 'appear' which is derived from *kazat'sa* 'seem', 'think' by prefixation.

[7] In the context where the author heard a native speaker utter (7) she referred to the effect of a sudden blizzard.

I conclude that Russian sentences like (7) and (9a) show a zero generic subject in 3rd person singular, neutrum with the value {– Human Agent} while (8b) does not have a zero subject.[8] The role of Recipient marked by the dative case is inherited from the underlying non-reflexive verbs, almost in the same way as in the Icelandic pair of examples (1) vs (2) the dative marking on the object is retained in the passive. An interesting characteristic of Russian examples (7) and (9) is that their form (reflexive/middle marking on the verb) does not correspond to their semantics, which is close to passive. We have an event structure with three arguments – Agent (X), Recipient (Y) and Patient (Z). On the one hand, in the reconstructed sentence with the corresponding non-reflexive active verb *nalit'* 'to pour' *X nalil Y-u Z v rukav* 'X poured some liquor Z in Y's sleeve' the overt Agent (X) can only be intentional, so (9) does not look as a canonic passive. On the other hand, in the derived structure (9a) the zero Agent ($Ø^{3Sg}$) stands not only for the Agent, but also for the Patient in (9b). Therefore, despite (9b) is not a direct source of (9a), (9a) is a kind of a semi-passive or a mediopassive construction where the zero argument $Ø^{3Sg}$ stands for two arguments in the active structure and the verb is morphologically marked as middle.

Active structure with a non-reflexive verb	Derived structure with a reflexive verb
X (Agent) – Nominative case	X ( ) – $Ø^{3Sg}$ Nominative case
Y (Recipient) – Dative case	Y (Recipient) – Dative case
Z (Patient) – Accusative case	**X ( ) – $Ø^{3Sg}$** Nominative case

Table 3: Ditransitive middle impersonals in Russian

3   Summary

All three languages have zero subject constructions with zero lexemes specified for role-and-reference features and acting as agreement controllers in structures traditionally analyzed as subjectless, notably in verbal transitive impersonals licensed by verbs of different classes: accusative, dative and ditransitive verbs, and in passive and middle constructions. An MTT-based approach to the problem of zero subjects helped to reveal non-trivial cross-linguistic similarities between different types of non-standard passive and middle constructions with participles and reflexive verbs. An oblique subject analysis does not make much sense in the constructions projecting an Agent argument.

---

[8]   It is plausible that (8b) has a sentential subject, with the complement clause filling in the subject position.

Acknowledgements

Bibliography

Babby, L. 2002. Subjectlessness, External Subcategorization, and the Projection Principle. *Journal of Slavic Linguistics*. Vol. 10.

Lavine, J. & R. Freidin. 2002. The subject of defective Tense in Slavic. *Journal of Slavic linguistics* 18: 1, 101-130.

Lavine, J. 2012. Varieties of v and the Structure of 'Anti-Burzio' Predicates. *Typology of morphosyntactic parameters,* Moscow 14-16 November, 2012.

Mel'čuk, I. 1995. Syntactic, or Lexical Zero in Natural Language. // The Russian Language in the Meaning-Text Perspective. *Wiener Slawistischer Almanach. Sonderband 39.* Moskau; Wien, 1995.

Reuland, E., ed. 2000. *Argument and case: Explaining Burzio's Generalization*. Amsterdam: John Benjamins.

Richardson, K. 2007. *Case and Aspect in Slavic*. Oxford: OUP.

Svenonius, P. 2002. Case in uninterpretable aspect. *Proceedings of Perspectives in Aspect*, University of Utrecht.

Sigurðsson, H. 2011. On the new passive. *Syntax* 14: 148-178.

Woolford, E. 2003. Burzio's Generalization, Markedness and Locality Constraints on Nominative Objects. Ellen Brandtner and Heike Zinsmeister (eds.). *New Perspectives in Case Theory*. 299-327.

Zimmerling, A. 2007. Zero Lexemes and Derived Sentence Patterns. *Wiener Slawistischer Almanach*. Sonderband 69. Wien.

Zimmerling, A. 2009. Dative Subjects and Semi-Expletive pronouns. G. Zybatow, U. Junghanns, D. Lenertová, P. Biskup (eds.). *Studies in Formal Slavic Phonology, Syntax, Semantics and Information Structure*. Frankfurt am Main; Berlin; Bern; Bruxelles; New York; Oxford; Wien, 2009.

Zimmerling, A. 2012. Nekanoničeskie podležaščie v russkom yazyke. *Ot značenija k forme, ot formy k značeniju: Sbornik statej v čest' 80-letija Aleksandra Vladimiroviča Bondarko.* Moscow: Yazyki slavyankix kul'tur. P. 568-590.