Minimodel of semantic synthesis of Russian sentences

Lidija Iordanskaja and Igor Mel'čuk
Observatoire de linguistique Sens-Texte, Université de Montréal

A formal linguistitic model is presented, which produces, for a given conceptual representation of an extralinguistic situation, a corresponding semantic representation [SemR] that, in its turn, underlies the deep-syntactic representations of four near-synonymous Russian sentences expressing the starting information. Two full-fledged lexical entries are given for the lexemes BESPORJADKI 'disturbance' and STOLKNOVENIE 'clash$_{(N)}$', appearing in these sentences. Some principles of lexicalization – that is, matching the formal lexicographic definitions to the starting semantic representation in order to produce the deep-syntactic structures of the corresponding sentences – are formulated and illustrated; the problem of approximate matching is dealt with in sufficient detail.

Keywords: NLP, semantic synthesis, semantic representation, semantic matching, deep-syntactic representation, lexical entries of the Explanatory Combinatorial Dictionary, Russian

1. Stating the problem

The present paper was born out of an innocuous looking attempt to formally describe the operations needed to produce a Russian equivalent for the following English sentence (the headline of a short newspaper note we saw accidentally):

(1) *Violence kills 7 in Iraq.*

This turned out much more difficult than we could foresee – because of a very general meaning of the noun VIOLENCE. It can be used to speak of a wide range of situations: from clashes of two rival bands to pogroms to suicide terrorist attacks. However, based on the contents of the note in question and our knowledge about events in present-day Iraq, we retained one of the possible situations, which can be referred to by the following Russian sentences:

https://doi.org/10.1075/li.00058.lor
Lingvisticæ Investigationes 44:1 (2021), pp. 101–136. ISSN 0378-4169 | E-ISSN 1569-9927

(2) a. *Besporjadki v Irake priveli k gibeli semi čelovek.*
'Disturbances in Iraq led to the death of seven people'.
b. (*Uličnye*) *stolknovenija v Irake priveli k gibeli semi čelovek.*
'(Street) clashes in Iraq led to the death of seven people'.

(3) a. *V rezul'tate besporjadkov v Irake pogiblo sem' čelovek.*
'As a result of disturbances in Iraq, seven people died'.
b. *V rezul'tate* (*uličnyx*) *stolknovenij v Irake pogiblo sem' čelovek.*
'As a result of (street) clashes in Iraq, seven people died'.

NB

1. Sentences (2) and (3) do not, of course, exhaust the set of possible Russian renderings of the information carried by (1); here are, for instance, some other translational equivalents of (1) (of a slightly different structure):
   i. a. *Sem' čelovek pogibli v rezul'tate besporjadkov v Irake*
   'Seven people died as a result of disturbances in Iraq'.
   b. *Sem' čelovek pogibli v rezul'tate* (*uličnyx*) *stolknovenij v Irake*
   'Seven people died as a result of (street) clashes in Iraq'.

   ii. a. *Žertvami besporjadkov v Irake stalo sem' čelovek*
   'Seven people became victims of disturbances in Iraq'.
   b. *Žertvami* (*uličnyx*) *stolknovenij v Irake stalo sem' čelovek*
   'Seven people became victims of (street) clashes in Iraq'.

2. Variants **a**. and **b**. are not fully synonymous: **a**-sentences can be used for a case of "one-way" violence (a group of people attacks another group of people), while **b**-sentences describe only "two-way" violence (two groups of people attack each other).

The problem raised in the present paper can be formulated as follows:

> How can one, proceeding formally, express in a given language the information about an extralinguistic situation extracted from a text of any language and/or from the observation of extralinguistic reality?

In our specific case we will discuss the expression of the information received from English sentence (1) in Russian, by sentences (2) and (3).

We try to solve this problem in the framework of the Meaning-Text linguistic approach (Mel'čuk 2012–2015 and 2016). A sufficient familiarity with this approach is presupposed: several notions and formalisms are used without explanations, and we allow ourselves abbreviations that are not always explicitly indicated, as well as some approximate formulations.

Under semantic synthesis is understood multiple synthesis aimed at producing a maximal quantity of sentences that express the initial information. Sentences being synthesized are not necessarily fully equivalent semantically: they can be only **quasi**-synonymous. The question arises as to what semantic discrepancies are acceptable/non-acceptable; in other words, a method is needed for determining how much approximation is allowed when deciding on semantic quasi-equivalence of two sentences. The principles of lexicalization (Sections 6.2 and 7) are supposed to ensure a reasonable degree of approximation – at least, within the limits of our data.

NB
Semantic quasi-equivalence and its allowed degree are considered in Milićević (2007a, 2007b, 2021), the studies dedicated to semantic and deep-syntactic paraphrasing.

We will discuss the two first steps towards the solution of this problem, namely, establishing two sets of correspondences:

– Those between the starting representation of an extralinguistic situation – that is, its conceptual representation [ConceptR] – and the semantic representations [SemRs] of the corresponding sentences.
– Those between the starting SemR and deep-syntactic representations [DSyntRs] of actual sentences.

These steps were sketched out in the paper Iordanskaja & Polguère (1988), dedicated to a system of text generation. Then Polguère (1990) introduced two items important in this connection:

1. The notion of communicative dependency between semantemes in the SemR; this type of dependency underlies the establishing of correspondences SemRs ⇔ DSyntRs.
2. A procedure for establishing correspondences between a SemR of a set of synonymous sentences and the DSynt-structures of the actual sentences that express this SemR; the procedure is based, in an essential way, on the semantic-communicative structure of the starting SemR.

Finally, Lareau et al. (2018) is the most recent description of a system of sentence synthesis – from a SemR to surface-syntactic sentence structures – that presupposes the use of dictionaries and grammars of various languages.

According to the general scheme of semantic synthesis – ConceptR ⇔ SemR ⇔ DSyntR – the solution to the above-stated problem requires the six following "documents":

1) A notation for the starting information (obtainable, for instance, from sentence (1) and/or other sources), that is, a conceptual representation of a given extralinguistic situation: the starting ConceptR (4), see below.
2) Russian SemR (5), which corresponds to ConceptR (4) and underlies sentences (2) and (3).
3) A fragment of the lexicon "Concepts ⇔ Semantemes" for Russian, necessary for the production of SemR (5) from ConceptR (4).
4) A fragment of a Russian *Explanatory Combinatorial Dictionary* that contains the lexical entries for the lexemes that make up sentences (2) and (3); it is needed in order to construct the DSyntSs corresponding to Russian SemR (5). Only two lexical entries will be given in this paper: for lexemes BESPORJADKI 'disturbance' and STOLKNOVENIE 'clash$_{(N)}$.'
5) Rules for matching lexicographic definitions to a SemR, or, more precisely, a fragment of the set of rules for producing the Russian deep-syntactic representations [DSyntRs] of sentences (2) and (3) from their SemR (5).
6) Rules ensuring the transition DSyntRs ⇔ Surface-PhoneticRs for Russian sentences (2) and (3). These rules, being part of standard modules of the Meaning-Text linguistic model, have been described in numerous publications, so we can ignore them here. See, for instance, Lareau & Wanner (2007).

Let us proceed to the review of documents 1–5.

2. The starting conceptual representation

The proposed ConceptR of the situation in question has been imagined by us on the basis of sentence (1) plus our extralinguistic knowledge. It corresponds to well-known conceptual graphs, which have been used in text generation systems for over several decades (see, for instance, the classical works Sowa 1976 and 1984). In this paper we cannot seriously discuss exactly what kind of ConceptR one needs in the framework of our task. However, we insist on the fact that some form of ConceptR is necessary: it is from this ConceptR that the starting SemR, underlying sentences (2) and (3), is produced.

Such a ConceptR has the following four general properties:

– The ConceptR is a reflection of **any** given extralinguistic situation SIT by an **average** (i.e., naïve) speaker of language **L**, who is not constrained by some

fixed objective parameters of scientific or quasi-scientific nature.[1] In other words, the ConceptR embodies only a subjective perception of the SIT by a profane observer. As a result, a given ConceptR(SIT) may prove insufficient for synthesizing sentences in language **L**: the lexicon and/or grammar of **L** may require additional data on the SIT that are absent from the starting ConceptR(SIT). For instance, a given ConceptR(SIT) may lack information on the manner of X's traveling when stating that X moved closer to Y (Rus. *podošël* 'approached walking' ~ *podplyl* 'approached swimming' ~ *podpolz* 'approached crawling' ~ *podletel* 'approached flying') or on the quantity of objects mentioned ('one' ~ 'more than one'), while this information is indispensable for constructing correct Russian sentences.

- The **vocabulary** of the language of ConceptRs is a drastically reduced vocabulary of a natural language – here, English – which is maximally freed from idiomaticity, that is, from its particular features. The English words used in the ConceptR are called concepts; they are put in angled quotes « ». The result is a kind of Basic English. Thus, the proposed vocabulary of ConceptR is, language-wise, "national." Outside of a narrow special domain, where the concepts are international technical terms, a language-universal ConceptR is impossible.
- The ConceptR **does not include inflectional significations** – grammemes. (In the version of the ConceptR given below, English words are conjugated and declined in order to facilitate the task of the human reader; however, in the formal procedure conjugated and declined forms are not taken into account.) The information to be carried by the grammemes is specified in the ConceptR by the concepts, for instance:
  - the tense of a verb: «before/at/after the moment of speech»;
  - the aspect of a verb: «repeated [= more than once]/completed event»;
  - the number of a noun: «one/more than one».
- The **syntax** of the language of ConceptRs is a hierarchized sequence of statements of the form «Ξ(ξ) : Ψ», where Ξ is a parameter that characterizes a fact or an entity ξ, and Ψ is the value of this parameter for ξ.

Let us insist on the following fact: ConceptR (4) represents only one of several possible interpretations of English sentence (1), based, among other things, on its

1. "Extralinguistic situation" does not mean 'a physically existing situation of real world,' i.e. a "chunk" of the objective reality. The crushing majority of texts do not directly describe physical, observable facts and entities, but deal with their **psychic** reflections, created by human conscience from the underlying information about the "real" reality, this information being extracted, as a general rule, again from texts.

author's knowledge about the state of affairs in today's Iraq. Thus, ConceptR (4) does not reflect the full semantic range of the English lexeme VIOLENCE.

(4) Starting ConceptR

NB

1. Strictly speaking, a Russian ConceptR must be written in Russian; however, for English-speaking readers this ConceptR has to be supplied with approximate English glosses. In order to save space, we allowed ourselves to omit Russian words, whose linguistic properties are irrelevant in a ConceptR, and to present the starting ConceptR just in English.
2. The concepts «cause**1**» and «act**II.2**» correspond to the semantemes 'cause**1**' and 'act**II.2**', see Appendix, pp. 134–135.

«event»	: «subevent **A** caused**1** subevent **B**»
«characterization of the event»	: «completed»
«subevent **A**»	: «either {Xs act**II.2** upon Y^1s, causing**1** damage to Y^1s $\}_{A'}$ or {Xs and Y^2s act**II.2** upon_each_other, causing**1** damage to_each_other$\}_{A''}$»
«quantity of subevents **A**»	: «more than once»
«characterization of actions**II.2** of subevent **A**»	: «simultaneous»; «in one spot»; «unlawful»
«time of subevent **A**»	: «before the moment of speech»
«localization of subevent **A**»	: «Iraq»
«type of X»	: «[a] human»
«quantity of Xs»:	: «numerous»
«type of Y^1»	: «[a] human and/or possessions of this human»
«type of Y^2»	: «[a] human»
«quantity of Y^2s»:	: «numerous»
«subevent **B**»	: «Z died»
«characterization of subevent **B**»	: «completed»
«time of subevent **B**»	: «before the moment of speech»
«type of Z»	: «[a] human»
«quantity of Zs»	: «seven»

Comments

1. The ConceptR (4) is vague in that it does not allow the reader to conclude what type of violence is meant: "one-way" or "two-way" violence. This is expressed by presenting the subevent **A** as a strict disjunction – in order to cover both possible cases.
2. The expression «each_other» (and anyone of its variants: «to_each_other», «upon_each_other», etc.) is an operator, that is, a system of rules that acts upon a concept of the form «**A** and **B** do P», adding to it the meaning of reciprocity:

   «each_other»(«**A** and **B** do P») : '**A** does P to **B, and B** does P to **A, and** **A**'s and **B**'s actions P are simultaneous'.
3. Zs who died (as participants of subevent **B**) can be Xs, Ys or neither (for instance, innocent bystanders).

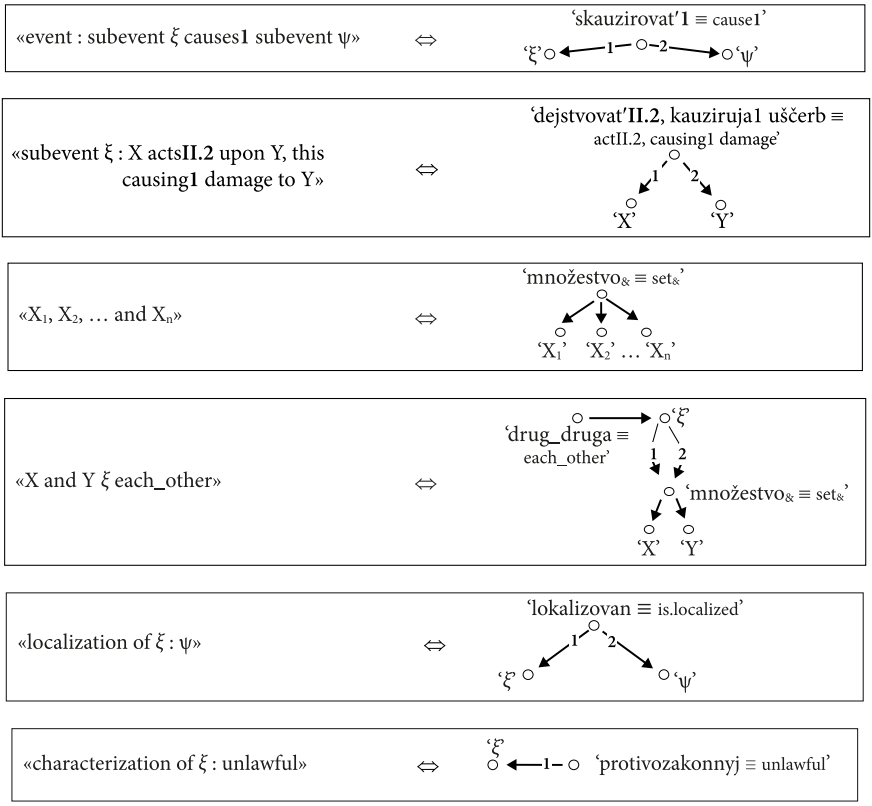
3. A fragment of the "Concepts ⇔ Semantemes" dictionary for Russian

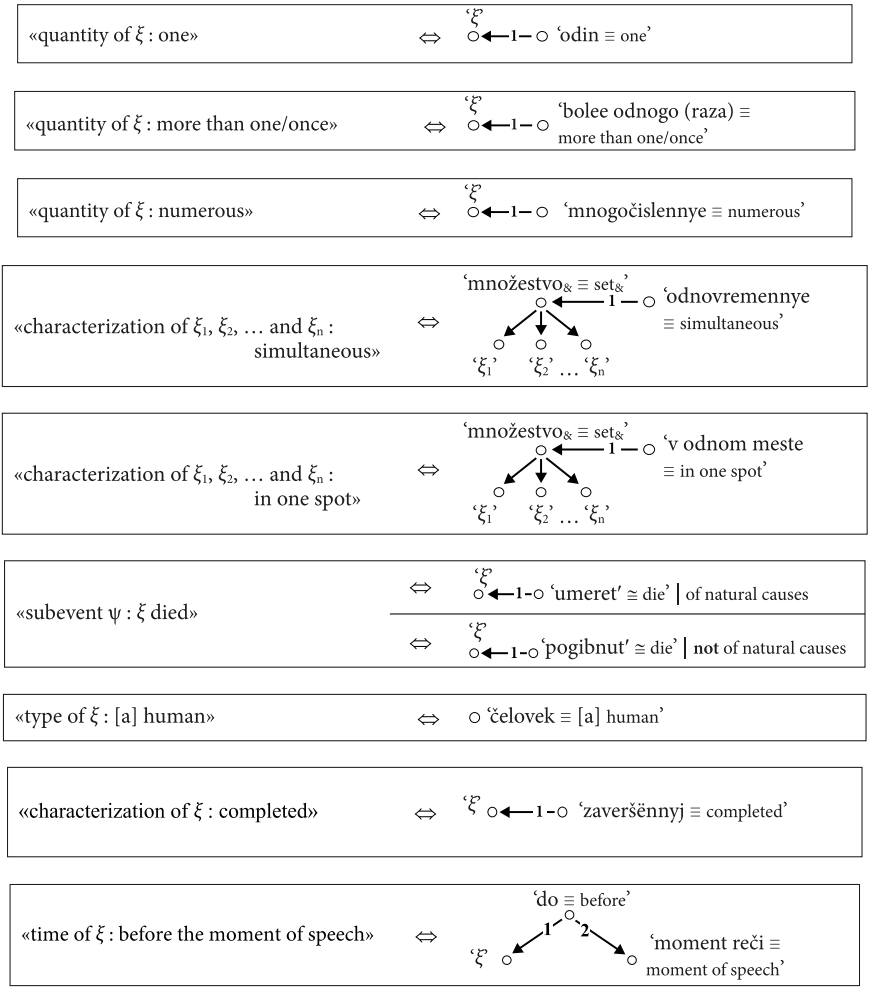
The very first step of the transition from ConceptR (4) to sentences (2) and (3) is the construction of a Russian SemR corresponding to ConceptR (4) (this SemR is then "translated" into alternative DSyntRs of Russian sentences by the semantic module of a Meaning-Text linguistic model for Russian). This step is performed by means of a special "Concepts ⇔ Semantemes" dictionary for Russian.

NB

1. Our illustration is substantially simplified; as a result, «concepts» and 'semantemes' stand, most of the times, in one-to-one relation. However, in linguistic reality, this is by no means the case.
2. Semantemes whose semantic content is not obvious are explained in the Appendix.

Here are a few entries of such a dictionary necessary for the transition from ConceptR (4) to the SemR (5), which underlies sentences (2) and (3):





It goes without saying that a "Concepts ⇔ Semantemes" dictionary is itself not sufficient for the synthesis of a SemR corresponding to the starting ConceptR. This transition requires as well a system of rules that would ensure the reunion of semanteme configurations that have been selected into a well-formed network. However, these rules are not considered in the present paper.

4. A starting Russian semantic representation

While constructing a SemR, based on the starting ConceptR, the Speaker[2] is free to omit some fragments of the ConceptR – following his own intentions, the circumstances of his communication act, the needs and abilities of his audience, etc. Thus, SemR (5) below corresponds only to the disjunct **A″** in the conceptual description of subevent **A**, that is, (5) specifies reciprocal acting of two groups of people upon each other (the Speaker chose to ignore the disjunct **A′**).

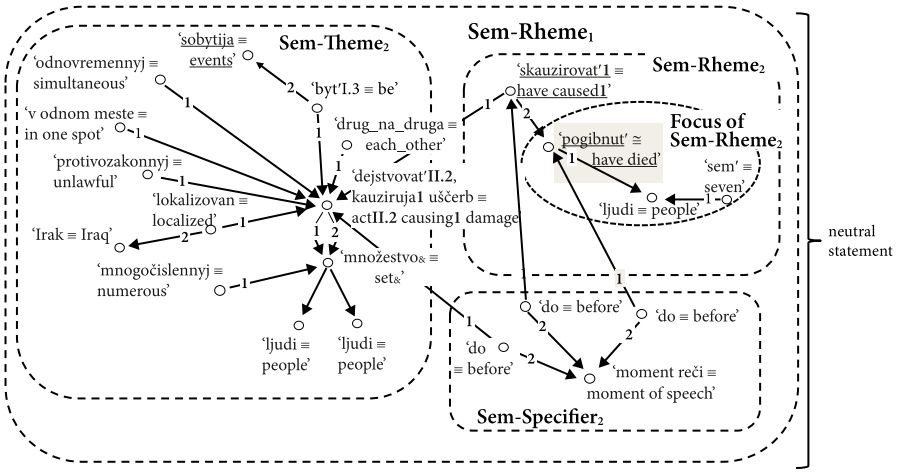
Since ConceptR (4) corresponds to the title of a piece of news, the output Russian sentences are all-rhematic: their SemR does not have a primary semantic Theme, so that each of them answers the underlying question "What happened?" and thus expresses only a semantic Rheme (for semantic-communicative structure of sentences, see Mel'čuk 2001). However, the all-embracing SemR of these sentences contains a secondary Rheme-Theme division, introduced by the Speaker as a function of his communicative goals. It is also the Speaker, of course, who specifies Comm-dominant nodes in Sem-communicative subareas. Therefore, SemR (5) presents only one of possible secondary Rheme-Theme divisions.

The Sem-communicative structure imposed on the SemS in SemR (5) must, of course, be preserved in any sentence to be synthesized from it, and such is the case of sentences (2) and (3);[3] sentences (i) and (ii) in **NB 1** on p. 102 express radically different Sem-CommSs and are not considered in this paper.

(5) One of the possible semantic representations corresponding to ConceptR (4)
To ensure a better readability, this SemR uses the following two abbreviations:
- In some cases grammemes are represented not by configurations of semantemes (as they should be), but directly by morphological form of the corresponding Russian word. Thus, instead of 'čelovek←bolee.odnogo' we write simply 'ljudi', instead of 'dejstvovat′**II.2**←bolee.odnogo.raza' we use 'dejstvovat′**II.2**' in the imperfective aspect, and instead of 'pogibnut′←zaveršënnyj', 'pogibnut'' in the perfective aspect.
- The node labeled 'dejstvovat′**II.2,** kauziruja**1** uščerb ≡ act**II.2** causing**1** damage' in Sem-Theme$_2$ should be, strictly speaking, be labeled as 'set$_{\&}$ of actions**II.2**, causing**1** damage'. We allow ourselves not to do this – in order to avoid cumbersome formulations in the subsequent discussion.

2. "The Speaker," with a capital *S*, denotes the author of the given utterance, that is, the initiator of the given speech act; "the speaker," with a small *s*, stands for a speaker of a language.

3. Sentences (3) do not have the same Sem-Comm-structure as sentences (2), but this is explained by the different choices of the entry node in ther starting SemR, see Subsection 6.1.



For details on semantic representation, see Mel'čuk (2012–2015: vol. 1, Part II, p. 183*ff*).

5. Lexical entries of two Russian lexemes: BESPORJADKI 'disturbance' and STOLKNOVENIE 'clash$_{(N)}$'

As indicated above, p. 104, we limit ourselves to lexicographic descriptions of two Russian lexemes appearing in sentences (2) and (3): BESPORJADKI 'disturbance' (*Krovavye besporjadki v Turcii oxvatili vsju stranu* 'Bloody disturbances in Turkey engulfed the whole country'.) and STOLKNOVENIE 'clash$_{(N)}$' (*Ežednevnye stolknovenija meždu silovikami i boevikami Rabočej Partii Kurdistana prodolžajutsja* 'Everyday clashes between the forces of order and the fighters of the Kurdistan Workers' Party continue'.). We give their lexicographic definitions in three formats:[4]

- Ordinary textual lexicographic definition in standard format adopted for *Explanatory Combinatorial Dictionaries*.
- Schema of the textual lexicographic definition in the form of a sequence of separate statements.
- Formal lexicographic definition in the form of a semantic structure, that is, of a semantic network. In this network the same abbreviations concerning

4. The proposed lexicographic definitions correspond to the methodological principles most recently expounded in Mel'čuk & Polguère (2018).

grammemes and the node 'dejstvovat'II.2, kauziruja1 uščerb ≡ actII.2 causing1 damage' are made, as in the starting SemR, see above.

The generic, i.e. the Comm-dominant, component of a definition is shown by underscoring.

All three formats of lexicographic definitions are equivalent.

The textual version of a lexicographic definition and its schema must be maximally readable. For this, "empty" (= grammatical) words necessary for grammatical correctness of Russian sentences are used. This is, of course, not the case of the formal version of a lexicographic definition, so that lexemic differences are possible between the textual and the formal versions of a definition.

The reader is supposed to be sufficiently familiar with the basic notions and formalisms used; when needed, Meľčuk (2012–2015: vol. 2, Chapter 11) or Meľčuk (2016) can be consulted.

5.1 BESPORJADKI 'disturbance'

5.1.1 *Lexical entry of BESPORJADKI 'disturbance'*

BESPORJADKI 'disturbance', noun, plural only, uncountable, individualizable

Lexicographic definition

– Textual lexicographic definition

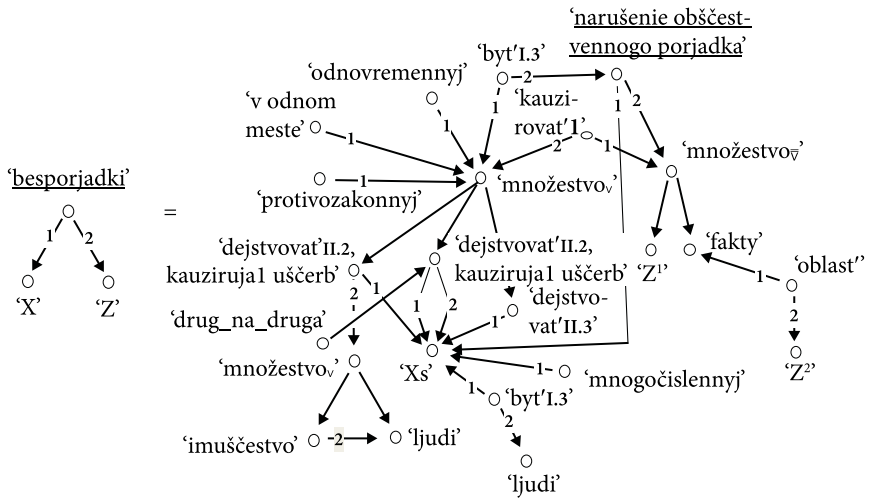
'besporjadki, ustraivaemye X-ami iz-za Z-a' = 'disturbance by Xs because of Z' =
'narušenie obščestvennogo porjadka X-ami iz-za Z-a, sostojaščee [= byt'I.3] v tom, čto
'breach of public order by X-s for reason Z consisting [= be] in that
mnogočislennye ljudi X {dejstvujutII.2, kauziruja1 uščerb, libo na ljudej ne-X i/ili ix imuščestvo, libo drug_na_druga} i/ili {X-y dejstvujutII.3}, pričëm èti dejstvijaII.2–3 odnovremenny, proisxodjat v odnom meste, protivozakonny i kauzirovany1 libo Z^1-om, libo faktami iz oblasti Z^2'
numerous people Xs {actII.2, causing1 damage, either upon people non-X and/or their possessions, or upon each_other} and/or {Xs actII.3}, these actionsII.2–3 being simultaneous, taking place in one spot, being unlawful and caused1 either by Z^1 or by facts from a domain Z^2'

– Schema of the textual lexicographic definition

'besporjadki, ustraivaemye X-ami iz-za Z-a' =
'narušenie obščestvennogo porjadka X-ami iz-za Z-a,

sostojaščee v tom, čto:
{{libo X-y dejstvujut**II.2**, kauziruja**1** uščerb, na ne-X-ov i/ili ix imuščestvo,}
{libo X-y dejstvujut**II.2**, kauziruja**1** uščerb, drug_na_druga}
{i/ili {X-y dejstvujut**II.3**}};
X-y – ljudi i mnogočislenny;
ne-X-y – ljudi;
dejstvija**II.2–3** X-ov odnovremenny;
dejstvija**II.2–3** X-ov proisxodjat v odnom meste;
dejstvija**II.2–3** X-ov protivozakonny;
dejstvija**II.2–3** X-ov kauzirovany**1** libo Z^1-om, libo faktami iz oblasti Z^2'

– Formal lexicographic definition



Government Pattern

'X' ⇔ I	'Z' ⇔ II
1. *sredi* 'among' $S_{PL,\ GEN}/S_{(collective)SG,\ GEN}$ 2. $A_0(S)$ 3. $Loc_{in} + S_{loc}(S)$ 4. $A_0(S_{loc}(S))$	1. *iz-za* 'because.of' S_{GEN} 2. *na počve* 'on the.soil' S_{GEN} 3. *na* $A_{\emptyset}(S)$ *počve* 4. *po povodu* ⟨*po slučaju*⟩ 'on the occasion/case' S_{GEN} 5. *po pričine* 'because.of' S_{GEN} 6. $A_{\emptyset}(S)$

1. **Non-desirable**: $^{?}A_0 + A_0$ [$^{?}$*studenčeskie religioznye besporjadki* ‘student religious disturbance’].
2. **If** ‘Z’ = ‘Z²’, **then** $C_{II} \neq C_{II.1}$ [**besporjadki iz-za religii* ‘disturbance because of religion’].
3. **If** ‘Z’ = ‘static facts’, **then** $C_{II} = C_{II.1/5}$ [*besporjadki iz-za ⟨po pričine⟩ grubosti direktora* ‘disturbance because of the director’s rudeness’].
4. **If** ‘Z’ = ‘dynamic facts’, **then** $C_{II} = C_{II.1/4/5}$ [*besporjadki iz-za ⟨po povodu⟩ uvol'nenij* ‘disturbance because of firings’].

$C_{I.1}$: *besporjadki sredi škol'nikov ⟨sredi molodëži, sredi naselenija⟩*
‘disturbance among students ⟨ among youth, among population ⟩ ’

$C_{I.2}$: *molodëžnye ⟨rabočie, soldatskie, studenčeskie⟩ besporjadki*
‘youth ⟨ worker, soldier, student ⟩ disturbance’

$C_{I.3}$: *besporjadki v školax ⟨v tjur'max, v kazarmax, v universitete, na zavode⟩*
‘disturbances in schools ⟨ in prisons, in military barracks, at the university, at the factory ⟩ ’

$C_{I.4}$: *škol'nye ⟨tjuremnye⟩ besporjadki* ‘school ⟨ prison ⟩ disturbance’

$C_{II.1}$: *besporjadki iz-za uvolennyx rabočix ⟨iz-za uvol'nenija rabočix⟩*
‘disturbance because of the fired workers ⟨ because of the firing of workers ⟩ ’

$C_{II.2}$: *besporjadki na počve religioznoj rozni* ‘disturbance on the soil of religious strife’

$C_{II.3}$: *besporjadki na religioznoj počve* ‘disturbance on the soil of religious strife’

$C_{II.4}$: *besporjadki po povodu ⟨po slučaju⟩ avarii* ‘disturbance because of the accident’

$C_{II.6}$: *religioznye ⟨èkonomičeskie, ètničeskie⟩ besporjadki* ‘religious ⟨ economical, ethnic ⟩ disturbance’

$C_{I.1} + C_{II.4}$: *ètničeskie besporjadki sredi mladšix oficerov* ‘ethnic disturbance among junior officers’

Lexical functions

Syn_n : stolknovenie ‘clash$_{(N)}$’; volnenija ‘unrest’; bunt ‘mutiny’, mjatež ‘rebellion’, vosstanie ‘uprising’; pogrom
Gener : narušenie ‘obščestvennogo porjadka’ ‘breach of public order’
Mult : volna ‘wave’ [~ov]
S_1 : učastnik/-ca ‘participant’ [~ov]
Magn : bol'šie ‘big’, krupnye ‘large’, masštabnye ‘large-scale’
$Magn_1^{quant}$: massovye ‘mass [~]’
IncepPredPlus : usilit'sja ‘intensify’
CausPredPlus : usilit' ‘enhance’ [~i]
CausPredMinus : umen'šit' ‘diminish’ [~и]
Loc_{in} : 'v xode' ‘in the course of’ [~ov]
$Oper_1$: učastvovat' ‘participate’ [v ~ax]
$Func_0$: 'imet' mesto' ‘take place’; proizojti ‘occur’, slučit'sja ‘happen’
$IncepFunc_0$: vozniknut' ‘surge’, načat'sja ‘begin’
[Magn + $IncepFunc_0$] : vspyxnut' ‘flare up’, razrazit'sja ‘break out’
$FinFunc_0$: zakončit'sja ‘end’, **coll.** končit'sja ‘end’, prekratit'sja ‘cease’, utixnut' ‘calm down’
$ContFunc_0$: prodolžat'sja**1** ‘continue’ | IMPF only; prodolžat'sja**2** ‘continue’, dlit'sja ‘last’

	[NUM + $N_{\text{'time'}}$] | IMPF only
$CausFunc_{\theta}$	: sejat' 'sow' [~i] | B. are numerous; IMPF only
$Caus_{1}Func_{\theta}$	: ustroit' 'organize', učinit' '≈ do' [~i]; činit' '≈ do' [~i] | B. are numerous
$LiquFunc_{\theta}$	: likvidirovat' 'liquidate', podavit' 'crush' [~i]
dispersing Xs, $LiquFunc_{\theta}$	: razognat' 'disperse' [~i]
$LiquIncepFunc_{\theta}$	: predotvratit' 'prevent', preseč' 'nip in the bud' [~i]
[Magn + Involv]	: sotrjasat' 'shake' [N = S_{ACC}] | only IMPF, B. are numerous
IncepInvolv	: oxvatit' 'engulf' [N = S_{ACC}], perekinut'sja, rasprostranit'sja 'expand to' [na N = S_{ACC}] | B. are numerous
[Magn + IncepInvolv]	: zaxlestnut' 'overwhelm' [N = S_{ACC}] | B. are numerous
happening on the streets of a city/town	: uličnye 'street' [~i]
causing1 many victims	: krovavye 'bloody'

V xode besporjadkov v Gaiti pogib reportër Radio Mega Neèmi Žozef. | *Iz novostej 2030 goda: V Moskve vspyxnuli besporjadki v Russkom Kvartale.* | *V massovyx besporjadkax v Kazaxstane učastvovalo bolee tysjači čelovek.* | *Ot uličnyx besporjadkov postradali svyše 100 čelovek i okolo polusotni magazinov.* | *OAE provociruet besporjadki i stolknovenija v Irake.* | *Besporjadki na rasovoj počve načalis' v konce maja 2020 goda.*

5.1.2 *Comments on the lexical entry of* BESPORJADKI *'disturbance'*

① *Individualization of the lexeme* BESPORJADKI

The lexeme BESPORJADKI is a *plurale tantum*: it exists only in the plural form ("pl!"), although it can denote a single individual event as well. It features an interesting particularity: just as other lexemes that denote individual events, it can be quantified: *mnogočislennye* ⟨*povtorjajuščiesja*, *častye*⟩ *besporjadki* 'numerous ⟨ repeated, frequent ⟩ disturbances', but it cannot be combined with a numeral: **pjat'* ⟨*odinnadcat'*⟩ *besporjadkov* '5 ⟨ 11 ⟩ disturbances'. However, some other *pluralia tantum* that also denote events combine with numerals without problem: *pjat'* ⟨*odinnadcat'*⟩ *poxoron* ⟨*pominok*⟩ '5 ⟨ 11 ⟩ funerals ⟨ memorial services ⟩ '. Russian does not allow its speakers to number BESPORJADKI! Thus, BESPORJADKI represents a subclass of uncountable nouns: individualizable uncountable nouns. (The denotations of substances – VOZDUX 'air', VODA 'water', DROŽŽI$_{(pl!)}$ 'yeast' – are non-individualizable and uncountable.) Individualizable uncountable nouns denote events that cannot be characterized by a number (**pjat' besporjadkov* 'five disturbances'), but can be distinguished and identified as individuals (in the logical sense): *včerašnie* ⟨*prošlogodnie, pervye* ⟩ *besporjadki* 'yesterday's ⟨ last year's, first ⟩ disturbance', *čislo besporjadkov* 'the number of disturbances', *bol'šinstvo besporjadkov* 'the majority of disturbances', *odni besporjadki za drugimi* 'some disturbance after the other', *besporjadki na vokzale 1-go aprelja* 'disturbance at the station on April 1st'. We know of a few other nouns with such properties (*plu-*

ralia tantum as well): VOLNENIJA$_{(pl!)}$ 'unrest' (**pjat' volnenij* 'five unrests' vs. *mnogočislennye volnenija* 'numerous unrests'), PRENIJA$_{(pl!)}$ 'debate', BDENIJA$_{(pl!)}$ 'vigil', POSIDELKI$_{(pl!)}$ 'friendly gathering', VYBORY$_{(pl!)}$ 'election', POD″ËMNYE$_{(pl!)}$ 'travel expenses', KOMANDIROVOČNYE$_{(pl!)}$ 'mission expenses', …

② *Generic component of the semanteme 'besporjadki = disturbance'*

Even if 'besporjadki = [a] disturbance' consists of 'dejstvija = actions', one cannot maintain that 'besporjadki' are a kind of action: in Russian, *dejstvija* 'actions' *soveršajutsja* 'are performed' or *vedutsja* 'are led' (but neither **proisxodjat* 'occur' nor **slučajutsja* 'happen'), while *besporjadki*, on the contrary, *proisxodjat* or *slučajutsja* (but neither **soveršajutsja* nor **vedutsja*). Therefore, from the Russian language viewpoint, BESPORJADKI denotes a kind of event. Moreover, "besporjadki" constitutes a rather narrow subclass of events: this is an event that breaches public order, that is, violates laws and norms valid in the respective society. Therefore, the semanteme 'narušenie obščestvennogo porjadka = breach of public order' is chosen as the generic component of the lexicographic definition of BESPORJADKI. (The generic component of the semanteme 'narušenie obščestvennogo porjadka' is, as indicated above, 'sobytie = event'.)

③ *Disjunctions in the definition of the lexeme BESPORJADKI*

The lexeme BESPORJADKI can describe three different situations of unlawful actions:

1. people Xs perform aggressive actions**II.2** against other people or their possessions (e.g., a pogrom);
2. Xs perform aggressive actions**II.2** against each other (e.g., clashes between youth gangs); and
3. Xs perform actions**II.3** (e.g., a meeting or a strike).

For instance: *Stranu sotrjasali besporjadki – pogromy, styčki molodëžnyx band, zabastovki, massovye mitingi* 'The country was shaken by disturbances: pogroms, clashes between youth gangs, strikes, mass meetings'.

This fact is accounted for in the definition of BESPORJADKI by means of the corresponding disjunctions.

④ *Split actantial variables Z^1 and Z^2 in the definition of the lexeme BESPORJADKI*

The cause of *besporjadki* 'disturbance' can either be a specific fact/entity (Z^1: *besporjadki iz-za gibeli policejskogo ⟨iz-za zemel'nyx nadelov⟩* 'disturbance because of the death of a policeman' ⟨ 'because of land plots' ⟩), or some unnamed facts from a particular domain (Z^2: *religioznye besporjadki* 'religious disturbance', *besporjadki na počve ètničeskoj vraždy* 'disturbance on the soil of ethnic strife').

⑤ *The semantic components 'odnovremenno = simultaneously' and 'v odnom meste = in one spot'*

The components 'odnovremenno = simultaneously' and 'v odnom meste = 'in one spot' are used here in a special sense – in order to express what is known as "unity of time and space." These components do not denote a chronological and geographical location, but serve for the Speaker to identify one individual event, which necessarily has a clear time and space frame.

⑥ *Actant 'X' ⇔* **I** *of the lexeme* BESPORJADKI

The phrase filling in the Sem- and DSynt-actantial slots 'X' ⇔ **I** of BESPORJADKI denotes a group of people that can have a common characteristic: age, social status, profession, religion, ethnicity, etc. With BESPORJADKI this group can be specified by the indication of the typical place where Xs are ordinarily found*:* *besporjadki v školax* 'disturbances in schools' ~ Xs are students; *besporjadki na zavodax* 'disturbances in factories' ~ Xs are workers; *besporjadki na rudnikax* 'disturbances in mines' ~ Xs are miners; etc. The typical place of Xs is described by the lexical function S_{loc}(L('X')), and the localization in this place, by the function Loc_{in}; this explains the line $C_{I.3}$ in the government pattern of BESPORJADKI. In a number of cases, the relative adjective of S_{loc}(L('X')) can also be used: *škol'nye* ⟨*tjuremnye*⟩ *besporjadki* 'school ⟨ prison ⟩ disturbances' (line $C_{I.4}$).

5.2 STOLKNOVENIE 'clash$_{(N)}$'

5.2.1 *Lexical entry of STOLKNOVENIE 'clash$_{(N)}$'*

STOLKNOVENIE 'clash$_{(N)}$', noun

Lexicographic definition

– Textual lexicographic definition

'stolknovenie meždu X-ami i Y-ami iz-za Z-a' = 'clash between Xs and Ys because of Z' =

'konflikt**II.2** meždu X-ami i Y-ami iz-za Z-a, sostojaščij v tom, čto

'conflict between Xs and Ys because of Z consisting in that

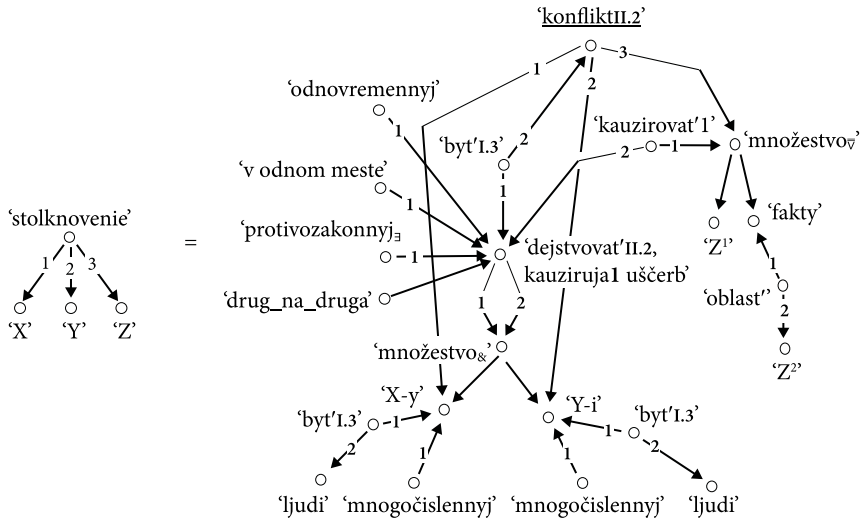
mnogočislennye ljudi X i mnogočislennye ljudi Y odnovremenno v odnom meste dejstvujut**II.2,** kauziruja**1** uščerb, drug_na_druga, pričëm dejstvija**II.2** X-ov i/ili Y-ov protivozakonny i kauzirovany**1** libo Z^1-om, libo faktami iz oblasti Z^2'

numerous people Xs and numerous people Ys simultaneously in one spot act**II.2**, causing**1** damage, upon_each_other, the Xs' and/or Ys' actions**II.2** being unlawful and caused**1** either by Z^1 or by facts from domain Z^2'

– The scheme of the textual lexicographic definition

'stolknovenie meždu X-ami i Y-ami iz-za Z-a' =
'konflikt**II.2** meždu X-ami i Y-ami iz-za Z-a, sostojaščij v tom, čto
X-y i Y-i dejstvujut**II.2** drug_na_druga, kauziruja**1** uščerb:
X-y i Y-i – ljudi i mnogočislenny;
dejstvija**II.2** X-ov i dejstvija**II.2** Y-ov odnovremenny;
dejstvija**II.2** X-ov i dejstvija**II.2** Y-ov – v odnom meste;
dejstvija**II.2** X-ov i/ili dejstvija**II.2** Y-ov protivozakonny;
dejstvija**II.2** X-ov i/ili dejstvija**II.2** Y-ov kauzirovany**1** libo Z^1-om, libo faktami iz oblasti Z^2'

– Formal lexicographic definition



Government pattern

'X' ⇔ I	'Y' ⇔ II	'Z' ⇔ III
1. S'_{GEN} *i* S''_{GEN} 2. *meždu* S'_{INSTR} *i* S''_{INSTR} 3. *meždu* $S_{PL, INSTR}$		1. *iz-za* S_{GEN} 2. *na počve* S_{GEN} 3. *na* $A_{\theta}(S)$ *počve*
4. S_{GEN}	4. *s* S_{INSTR}	4. $A_{\theta}(S)$

1. **If** 'Y' are 'forces of order', **then** $C_{II} = C_{II.4}$.
   [*stolknovenie demonstrantov s policiej* vs. **stolknovenie policii s demonstrantami*
   'clash of the demonstrators with police' vs. *'clash of police with the demonstrators']
2. **If** 'Z' = 'Z²', **then** $C_{III} \neq C_{III.1}$.
3. **Desirable**: at least one actant of STOLKNOVENIE is expressed or STOLKNOVENIE has a dependent lexical-functional adjective.

$C_{I+II.1}$: *stolknovenie čečencev i osetin* 'clash of Chechens and Ossetians'
$C_{I+II.2}$: *stolknovenie meždu čečencami i osetinami* 'clash between Chechens and Ossetians'
$C_{I+II.3}$: *stolknovenie meždu žiteljami sosednix ostrovov* 'clash between the inhabitants of neighboring islands'
$C_{I.4} + C_{II.4}$: *stolknovenie čečencev s osetinami* 'clash of Chechens with Ossetians'
$C_{III.1}$: *stolknovenie iz-za pastbišč* 'clash because of pastures'
$C_{III.2}$: *stolknovenie na počve mežnacional'noj vraždy* 'clash on the soil of interethnic strife'
$C_{III.3,\,4}$: *stolknovenie na religioznoj počve* 'clash on the religious soil', *religioznoe stolknovenie* 'religious clash'
$C_{I+II.3}$ + $C_{III.2}$: *stolknovenie meždu žiteljami sosednix ostrovov na počve èkonomičeskix raznoglasij* 'clash between the inhabitants of neighboring islands on the soil of economic contradictions'

Lexical functions

Syn_{n} : besporjadki 'disturbance'; draka 'fistfight', potasovka 'brawl', styčka ≈ 'clash$_{(N)}$'
Gener : konflikt**II.2**
$S_{1/2}$: učastnik/-ca 'participant' [~ja]
S_{1+2} : protivniki 'adversaries'
Magn : krupnoe 'large', masštabnoe 'large-scale'
$Magn^{quant}_{1+2}$: massovoe 'mass-'
Loc_{1n} : v xode 'in the course of' [~ja]
$Func_{0}$: proizojti | **only** PERF; imet' mesto, proizojti | S. – PL
$IncepFunc_{0}$: načat'sja 'begin'; vozniknut' 'appear' | S. – PL
[Magn + $IncepFunc_{0}$] : vspyxnut' 'flare up', razrazit'sja 'break out' | S. – PL
$FinFunc_{0}$: zakončit'sja 'end', **coll.** končit'sja 'end', prekratit'sja 'cease'
$ContFunc_{0}$: prodolžat'sja**1**'continue'; prodolžat'sja**2** 'continue', dlit'sja 'last' [NUM + $N_{'time'}$]
Involv : oxvatit' 'engulf' [N = S_{ACC}], perekinut'sja 'spill over', rasprostranit'sja 'expand' [na N = S_{ACC}] | S. – PL
$Caus_{1}Func_{0}$: ustroit' 'organize' [~e]; učinit' ≈ 'do' [~ja] | S. – PL
$LiquFunc_{0}$: likvidirovat' 'liquidate', podavit' 'crush' [~ja] | S. – PL
$LiquIncepFunc_{0}$: predotvratit' 'prevent' [~e]; preseč' 'nip in the bud' [~ja] | S. – PL
happening on a street : uličnoe 'street' [~e]
causing**1** many victims : krovavoe 'bloody'

```
Xs and Ys are using
            weapons   : vooružënnoe 'armed'
```

Včera v gorode proizošlo pjat' stolknovenij meždu čečencami i osetinami. | Stolknovenija na religioznoj počve v centre Kaira paralizovali dviženie transporta. | V xode ètogo stolknovenija pogiblo šest' čelovek. | V Sirii vspyxnuli stolknovenija meždu dvumja protureckimi gruppirovkami. | Policija dolžna podavljat' stolknovenija meždu različnymi gruppami immigrantov. | 13-go oktjabrja v Pridnestrov'e proizošlo ètničeskoe stolknovenie. | Sotni migrantov postradali pri stolknovenijax s makedonskoj policiej.

5.2.2 *Comments on the lexical entry of* STOLKNOVENIE *'clash'*

① *Generic component of the lexicographic definition of the lexeme* STOLKNOVENIE *'clash$_{(N)}$': 'konflikt**II.2** = conflict**II.2**'*

'Konflikt**II.2** = conflict**II.2**' denotes an event consisting of reciprocal violent actions of two opposing parties (for other lexemes of the KONFLIKT vocable, see Appendix). The meaning of 'konflikt**II.2** = conflict**II.2**' is wider, that is, poorer, than the meaning of 'stolknovenie = clash$_{(N)}$': 'conflict**II.2**' does not presuppose either the unity of time and space of the actions under discussion, or their unlawful character. In other words, 'stolknovenie = clash$_{(N)}$' is a particular case of 'konflikt**II.2** = conflict**II.2**'.

② *Split actantial variables Z¹ and Z² in the definition of the lexeme* STOLKNOVENIE

The splitting of the variable Z is here of the same nature as with the lexeme BESPORJADKI. The cause of a clash can also be either a specific fact/entity (Z¹: *stolknovenija iz-za policejskogo proizvola ⟨iz-za zemel'nyx nadelov⟩* 'clashes because of police brutality' ⟨ 'because of land plots' ⟩ , or some unnamed facts from a particular domain (Z²: *religioznoe stolknovenie* 'religious clash', *stolknovenija na počve ètničeskoj vraždy* 'clashes on the soil of religious strife').

③ *Lexical functions with a key noun in the plural*

Some LFs in the entries of the lexemes BESPORJADKI and STOLKNOVENIE carry constraints on the grammatical number of the keyword: "B. are numerous" and "S. – PL." (Since BESPORJADKI is a *plurale tantum*, it is impossible to indicate plurality with this noun simply by the marker "PL.") Thus, one *stolknovenie* 'clash' cannot **oxvatit'* 'engulf' [$N = S_{ACC}$], **perekinut'sja* 'spill over' or **rasprostranit'sja* 'expand' [*na* $N = S_{ACC}$]: these verbs combine only with the plural forms of this lexeme. This phenomenon is quite systematic and has to be described in a general

form: the plural modifies the semantism of the keyword noun and thus entails some modifications in its lexical cooccurrence.

5.3 Comparison of the semantemes 'besporjadki = disturbance' and 'stolknovenie = clash'

The meanings of the lexemes BESPORJADKI and STOLKNOVENIE are sufficiently close for them to be considered semantically quasi-equivalent; they are intersecting synonyms ($Syn_{\cap}$). They feature the following semantic differences.

1) They have different generic components: BESPORJADKI has 'narušenie obščestvennogo porjadka = breach of public order', and STOLKNOVENIE has 'konflikt**II.2** = conflict**II.2**'.
2) The difference of the generic components entails the difference in the number and nature of actants. The lexeme BESPORJADKI does not presuppose two opposing parties and therefore it has only one Actor: actant X; on the contrary, STOLKNOVENIE happens necessarily between two adversaries, so that the lexeme has two Actors: actants X and Y.
3) The lexicographic definition of BESPORJADKI contains a triple disjunction, see 5.1.2, Comment ③, p. 115; the definition of STOLKNOVENIE includes only one disjunction.
4) Both in BESPORJADKI and STOLKNOVENIE, Xs are numerous. But the number of the targets of Xs' action**II.2** in BESPORJADKI is irrelevant: BESPORJADKI can denote an attack by a mob on a few people. STOLKNOVENIE, however, presupposes relative equality in number of Xs and Ys.
5) BESPORJADKI implies the unlawfulness of **all** Xs' actions; in STOLKNOVENIE one of the two sides – the forces of order – can act lawfully. Therefore, the semanteme 'protivozakonnyj = unlawful' in the lexicographic definition of STOLKNOVENIE has the subscript $_{\exists}$ (see Appendix).

6. Transition from SemR (5) to the DSynt-representations of sentences (2) and (3)

The transition from SemR (5) the to DSyntRs of sentences (2) and (3) is carried out in two steps:

– The transition from the Sem-structure in (5) to the DSynt-structures of sentences (2) and (3).

– The transition from the Sem-Comm-structure in (5) to the DSynt-Comm-structures of sentences (2) and (3); this transition is rather simple and will not be considered in this paper.

6.1 Transition from the Sem-structure in (5) to the DSynt-structures of sentences (2) and (3)

The transition from a SemS to a DSyntS requires two operations, which are intimately intertwined:

– **Arborization**, which synthesizes, for a given Sem-network, a corresponding DSynt-tree. First, an entry node is selected in the Sem-network; this node determines the top node of the tree. Then for each Sem-relation in the Sem-network, the corresponding DSynt-relation is selected; these DSynt-relations build up the DSynt-tree.
– **Lexicalization**, which includes lexicalization proper and morphologization.
  – Lexicalization proper: for a configuration of semantemes '$\tilde{\sigma}$' in the starting Sem-network a corresponding lexical unit is selected[5] and put on the appropriate node in the DSynt-tree. This is done by means of matching the lexicographic definitions against the starting Sem-network:
    a. The semanteme 'σ_{def}' of the lexicographic definition under processing is matched to the corresponding semanteme 'σ_{SemR}' in the starting SemR. If there is no perfect match, the semantic decomposition of both semantemes should be recurred to – in order to verify the possibility of an admissible partial match (see 6.3, Item 5, p. 124).
    b. An actantial variable of 'σ_{def}', for instance, X in the configuration 'σ_{def}–**i**→X', is matched to the corresponding actant of 'σ_{SemR}'; all semantic constraints imposed on X are checked.
    We are not in a position to describe here the procedure of matching the lexicographic definitions on the SemS in any detail.
  – Morphologization: for a particular configuration of semantemes in the Sem-network the corresponding grammeme is selected and attached to its lexical unit in the DSynt-tree.

5. In point of fact, the configuration of semantemes '$\tilde{\sigma}$' under processing is first replaced by a semanteme 'σ' such that '$\tilde{\sigma}$' = 'σ'; it is for 'σ' that the lexical unit L('σ') is selected and put into the DSynt-tree.

Arborization and lexicalization are performed, of course, not consecutively (one after the other), but in an interspersed manner. Their application is illustrated by the example of transition from SemR (5) to the DSyntRs of sentences (2) and (3).

In conformity with the rules of establishing the entry node in a SemR (Iordanskaja 1990; Mel'čuk 2001: 38–48),[6] there are two choices for the entry node in SemR (5) – that is, for the node that determines the top node of the DSynt-structure being synthesized:

– Either the semanteme 'skauzirovat'**1** = have caused**1**' (which is the Comm[unicative]-dominant node of the Sem-Rheme$_1$, that is, of the whole SemR).
– Or the semanteme 'pogibnut' ≈ have died' (which is the Comm-dominant node of the Sem-Rheme$_2$'s Focus).

The first choice leads to sentences (2), and the second, to sentences (3).

6.2 Lexicalization principles

Let us remind the reader of the general principles of lexicalization in the Meaning-Text approach, formulated in Polguère (1990).

---

6. Here is a concise presentation of these rules.

Choosing the candidates for the entry node in the SemR

A node 'σ' in the SemR 'S' can be the entry node, if and only if 'σ' satisfies at least one of the three following conditions:

1. either 'σ' is a normal predicate and the Comm-dominant node of the Sem-Rheme or the Sem-Theme of 'S';
2. or 'σ' is a mental predicate (that expresses an opinion or an estimate) and the Comm-dominant node of one of the Sem-Specifiers;
3. or 'σ' is a quasi-predicate, the Comm-dominant node of the Sem-Rheme and the only candidate for the entry node.

Choosing between the candidates 'σ$_1$' and 'σ$_2$' for the entry node in the SemR

1. If one from the semantemes 'σ$_1$' or 'σ$_2$' has a verbal expression in **L**, while the other has no such expression, the first one is selected.
2. If both semantemes 'σ$_1$' and 'σ$_2$' have/do not have a verbal expression, the one that is part of the Sem-Rheme is selected.

"Verbalizing" a non-verbal top node in the DSynt-tree

If the selected entry node 'σ' has in **L** only non-verbal expression L('σ'), the expression L('σ') is supplied with one of the semantically empty lexical-functional verbs (from the Oper$_i$ family).

1. The Rhematic and the Thematic communicative areas of the starting Sem-structure must be lexicalized separately. In other words, a subnetwork of the Sem-structure to be realized by one lexical unit must be completely contained either in the Sem-Rheme or in the Sem-Theme.
2. The preservation of the initial Sem-Communicative structure requires that the top node of the DSynt-tree that implements the Rhematic or the Thematic communicative area be filled in with a lexeme whose definition's generic component coincides with the Comm-dominant node of this area.
3. If there are several lexemes whose definitions match subnetworks of the starting SemR, the preference is given to the lexeme that ensures the maximal matching. (The notion of "maximal matching" needs, of course, a rigorous definition.)

These principles are necessary, but they are not sufficient in the general – and linguistically quite habitual – case of lexicalization. Namely, if the language lacks a lexeme whose definition matches the SemR perfectly (that is, each 'σ_{def}' matches a 'σ_{SemR}'), a **partial matching** becomes unavoidable. A question arises: Which partial matchings are acceptable and which are not? The present paper deals with partial matchings, and we propose some answers, formulated as four additional lexicalization principles (Section 7).

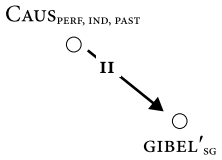
6.3 Constructing the DSynt-structure of sentence (2a) (*besporjadki priveli k gibeli*)

The transition from SemR (5) to the DSynt-structure of sentence (2a) is carried out in eight steps. The first four perform the lexicalization of the rhematic part of the SemR, and the next four, the lexicalization of the thematic part. Semantemes and configurations of semantemes receive their linguistic expressions – lexemes or grammemes.

Processing the Sem-Rheme$_2$

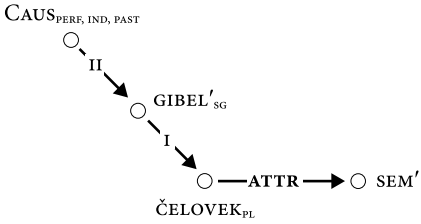
1. The entry node of SemR (5), that is, the semanteme 'skauzirovat′**1** = have caused**1**', gives rise to the verbal lexical function Caus, one of the possible realizations of 'skauzirovat′**1**'. The deep lexeme Caus is put on the top node of the DSynt-tree to be synthesized.
2. The grammeme PERF, attached to the verb Caus, comes from the semanteme 'zaveršënnyj = completed', and the grammeme IND, from the indication "neutral statement" in the Rhetorical Structure of the starting SemR. The semanteme configuration 'do momenta reči = before the moment of speech', bearing on the semanteme 'skauzirovat′**1**', gives the grammeme PAST. The result is

3. The semanteme 'pogibnut′ ≈ have died' (Sem-actant **2** of 'skauzirovat′**1** = have caused**1**') has as one of its possible implementations the noun GIBEL′ ≈ 'death' (*singulare tantum*):



The semanteme 'pogibnut′ ≈ have died' can also be implemented by the verb POGIBNUT′, which would give … *priveli k tomu, čto pogibli*… 'led to the.fact that … died'.

4. The configuration of semantemes 'sem′–**1**→ljudi = seven people' is lexicalized in a trivial way:



Processing the Sem-Theme$_2$

5. The semantic configuration 'event consisting in unlawful simultaneous actions**II.2** of numerous people causing**1** damage that happen in one place', which is seen in SemR (5), is part of the lexicographic definitions of several Russian lexemes (BESPORJADKI, BUNT, VOLNENIJA, MJATEŽ, STOLKNOVENIE, …); here the lexeme BESPORJADKI will be considered. The lexicographic definition of this lexeme is matched to the starting SemR in such a way that the semantemes of the definition are superimposed on the semantemes occupying the nodes of the SemR. The process meets three challenges:

– Generic (= central) component of the lexicographic definition of BESPORJADKI is 'narušenie obščestvennogo porjadka = breach of public order'; it is absent from SemR (5). However, its own generic component – 'sobytie = event' – is found in the SemR; thus, we have here a partial matching. Matchings of this type are considered legitimate, in spite of the fact that the sentence

produced adds meaning to the meaning presented in (5) – namely, the information about the inclusion of the given event into the class "breach of public order."

The phenomenon described here corresponds, as we believe, to the actual process of speaking. Selecting lexical units during lexicalization, the Speaker quite often enriches or impoverishes the starting meaning.

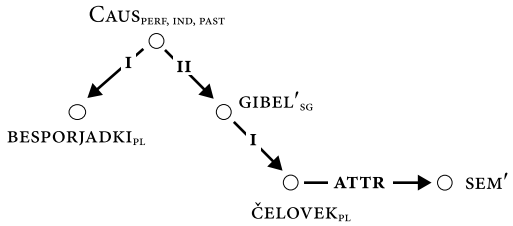
A question naturally arises: What is the information that can be added to or lost from the sentence produced under lexicalization? In other words, what is the allowed degree of approximation of this operation? To answer this question, a special investigation is needed, and a very serious one. For the time being, we have to limit ourselves to the following particular remark concerning the synthesis of sentence (2a). Adding the semantic component 'be**I.3** a breach of public order' to the meaning presented by the starting SemR is allowed, because:

1. it contradicts nothing expressed in (5);
2. it does not change the presentation of the situation SIT under consideration, since it expresses a subjective opinion of the Speaker about the classification of SIT;
3. it does not carry an axiological evaluation ("positive ~ negative"): it simply includes SIT into a particular class of events. (This is in fact a formulation of a new principle of lexicalization – see Section 7.)

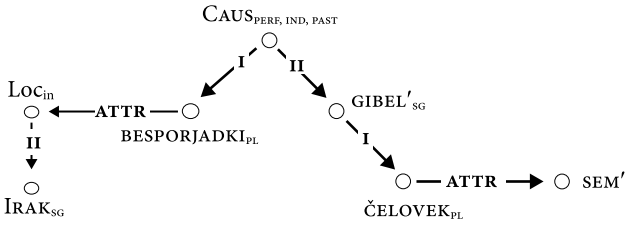
– As stated above, the definition of BESPORJADKI contains the following disjunctions:
{either {Xs' actions**II.2** against non-Xs **or** against non-Xs' possessions}$_1$
or {Xs' actions**II.2** against each_other}$_2$}
or {Xs′ actions**II.3**}$_3$

SemR (5) presents only the disjunct **2:** actions**II.2** of two groups of people against each other. This, however, does not prevent the Speaker from using the lexeme BESPORJADKI: disjunctions in lexicographic definitions are there in order to allow the use of the given lexeme in any of disjunctive cases.

– The semantic component 'is caused**1** by Z^1 or …' in the definition of BESPORJADKI has no correspondence in SemR (5). This is not an obstacle either: it is a standard case of a meaning that, while foreseen by BESPORJADKI as a non-obligatory actant, remains unexpressed.
To sum up: all three complications do not block the matching of the definition of BESPORJADKI to SemR (5). The result is the following DSynt-subtree:



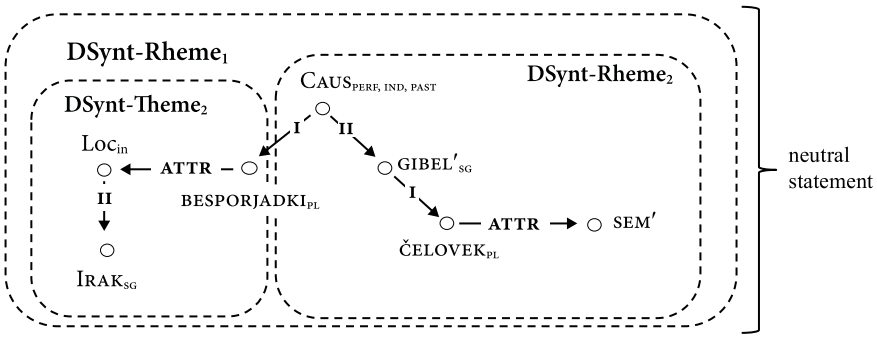
6. The semanteme 'bolee odnogo = more than one', which characterizes the Comm-dominant node 'sobytie = event', is realized by the grammeme PL – to be attached to the lexeme BESPORJADKI. However, this lexeme is a *plurale tantum* (it has the grammeme PL in its lexical entry), so that nothing happens. (In sentences (3) the grammeme PL is attached to STOLKNOVENIE to give a "normal" plural: *stolknovenija*.)
7. The processing of the configuration 'localized–**2**→Iraq' is trivial: the semanteme 'localized' corresponds, among other things, to the lexical function $\mathrm{Loc}_{\mathrm{in}}$; as a result, we obtain:



In the SSynt-structure, $\mathrm{Loc}_{\mathrm{in}}$(IRAK) is implemented as the preposition v 'in' governing the locative.

Constructing the DSynt-Communicative and the DSynt-Prosodic structures for sentence (2a) is rather trivial; the final DSyntR of (2a) is as follows:

(6)

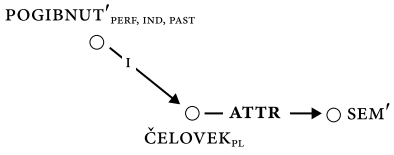


6.4 Constructing the DSynt-structure of sentence (3a) (*v rezul'tate besporjadkov pogibli*)

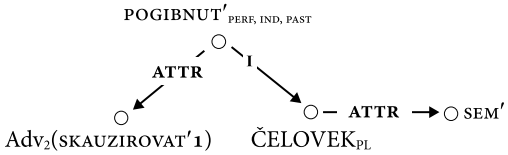
The transition from SemR (5) to the DSyntS of sentence (3a) differs from the transition to the DSyntS of sentence (2a) by the lexicalization and arborization only of the rhematic area of SemR (5); the processing of the thematic area of SemR (5) is the same as for sentence (2a).

1. The lexicalization and arborization of the entry node 'pogibnut′ ≈ have died' of SemR (5) consists in selecting the verb POGIBNUT′$_{PERF}$ and putting it at the top node of the DSynt-tree; the configuration 'do momenta reči = before the moment of speech', which characterizes the semanteme 'pogibnut′', produces the grammeme PAST attached to the verb; the grammeme IND is derived from the indication "neutral statement" in the Rhetorical Structure of the starting SemR. The result:

2. The configuration of semantemes 'sem′–**1**→ljudi = seven people' gives, as in sentence (2a), an obvious result:



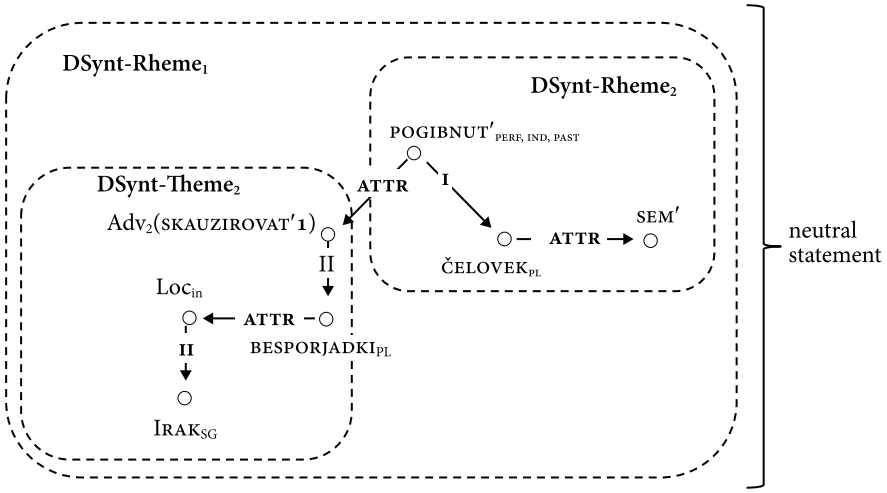
3. The lexicalization and arborization of the semantic node 'skauzirovat′**1** = have caused**1**', one of whose actants is 'pogibnut′': since the verb POGIBNUT′ is already selected as the top node of the DSynt-tree, the inversion of dependencies, or "head-switching," becomes necessary. The lexical unit that expresses the semanteme 'cause**1**' must be a DSynt-attribute of the verb POGIBNUT′, i.e. an adverb; we can select the lexical function Adv, or, more precisely, Adv$_2$, because the meaning to be expressed is 'being caused**1**':



At the SSynt-level the LF Adv$_2$(SKAUZIROVAT′**1**) is realized as the idiom ⌜V REZUL′TATE⌝ 'as a result of'.

All other operations of the transition from the SemR to the DSyntS are the same as explained above for sentence (2a); the final result is (7):

(7)



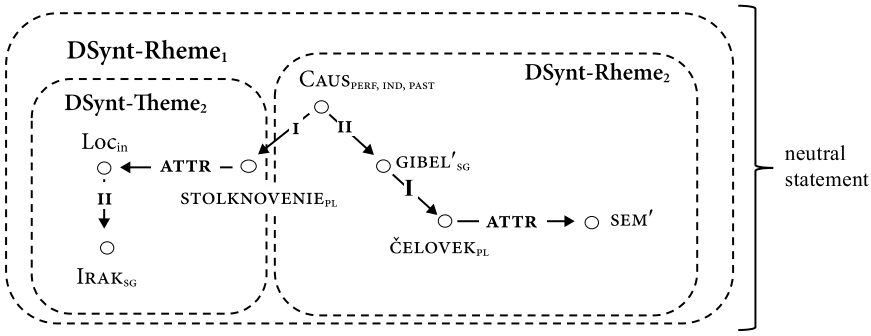
6.5 Constructing the DSynt-structure of sentence (2b) (*stolknovenija priveli k gibeli*)

Sentence (2b) uses, instead of the lexeme BESPORJADKI 'disturbance', its synonym STOLKNOVENIE 'clash$_{(N)}$'; all the other lexemes of the sentence are the same as in (2a). Therefore, we have to consider here only the matching of the STOLKNOVENIE definition.

The generic component of the semanteme 'stolknovenie = clash', that is, 'konflikt**II.2**', is matched to the Comm-dominant node of the Sem-Theme – 'sobytie = event' – partially, namely, by its own generic component, which is also 'sobytie'; but, as was indicated above, such a partial matching is considered legitimate. The other semantemes of the definition are matched practically in the same way as it was the case for BESPORJADKI.

Here is the resulting DSyntR:

(8)



This DSyntR is absolutely correct, but sentence (9), obtained from it, is stylistically not quite satisfactory, at least as the title of a piece of news:

(9) *Stolknovenija v Irake priveli k gibeli semi čelovek*
'Clashes in Iraq led to the death of seven people'.

This fact is accounted for in the lexical entry for STOLKNOVENIE: Rule 3), which follows the Government Pattern, gives an explicit indication that it is preferable to use this lexeme with at least one of its actants expressed or with one of its lexical functional adjectives. (It is quite possible that this particularity is due to the high polysemy of the vocable STOLKNOVENIE.) SemR (5) does not specify the actants of 'stolknovenie', and it is impossible to supply them without "unlawfully" enriching the starting meaning. As far as the LF-adjectives are concerned, STOLKNOVENIE has six: *krupnoe* 'large', *masštabnoe* 'large-scale', *massovoe* 'mass-', *vooružënnoe* 'armed', *krovavoe* 'bloody' and *uličnoe* 'street-'. *Krupnoe* ⟨*masštabnoe*, *massovoe*⟩ *stolknovenie* all imply a high number of Xs and Ys, and *vooružënnoe stolknovenie*, the use of weapons. However, since SemR (5) carries none of these meanings, it would be presumptuous to put one of these adjectives into the sentence. If *stolknovenie* is called *krovavoe* 'bloody', it is supposed to have many victims – yet it is not obvious that seven victims is many for Iraq. So we are left with the adjective *uličnoe* 'street-', which can be used, since SemR (5) does not specify the localization of the clashes (there are no circumstantials such as *na granice* 'at the border', *na universitetskix kampusax* 'on university campuses', etc.), while *stolknovenija* 'clashes' most often happen in open spaces. So one is allowed to add to the DSynt-tree under synthesis the adjective ULIČNYJ 'street-' as an **ATTR** to STOLKNOVENIE, which results in sentence (10):

(10) *Uličnye stolknovenija v Irake priveli k gibeli semi čelovek*
'Street clashes in Iraq led to the death of seven people'.

6.6 Constructing the DSynt-structure of sentence (3b) (*v rezul'tate stolknovenij pogibli*)

Sentence (3b) does not present new problems.

6.7 Filtering out wrong lexicalizations

The correct lexicalizations have been considered in Subsections 6.3–6.6; now we must say a few words about the blocking of wrong lexicalizations. A possibility for a wrong lexicalization of SemR (5) is due to the following fact: the definitions of the "correct" lexemes BESPORJADKI and STOLKNOVENIE include the semantic component 'event that consists of unlawful actions**II.2** by numerous people that cause**1** damage and are performed simultaneously and in one spot', and this component is also part of the definitions of such "wrong" lexemes as BUNT 'riot' and MJATEŽ 'rebellion' (the semantemes of this component are shown in the definitions by boldface); cf.:

BUNT 'riot'
'bunt X-ov protiv Y-a' = 'Xs' riot against Y' =
'**sobytie, sostojaščee v tom, čto**
'event consisting in that
mnogočislennye ljudi X, prinadležaščie k nizšemu social'nomu klassu, publično otkazavšis' povinovat'sja mestnoj vlasti Y^1 ili rasporjaženijam Y^2 vlasti Y^1, **odnovremenno v odnom meste**, ispol'zuja oružie, **protivozakonno dejstvujutII.2, kauziruja1 uščerb**, na predstavitelej Y^1-a s cel'ju likvidirovat' Y^1'
numerous people Xs that belong to a lower social class, having publicly refused to obey local authorities Y^1 or the instructions Y^2 of Y^1, simultaneously in.one.spot and using weapons unlawfully act**II.2,** causing**1** damage, upon representatives of Y^1 with the goal of liquidating Y^1'

MJATEŽ 'rebellion'
'mjatež X-ov protiv Y-a' = 'Xs' rebellion against Y' =
'**sobytie, sostojaščee v tom, čto**
'event consisting in that
mnogočislennye ljudi X, prinadležaščie k odnoj gruppirovke, publično otkazavšis' povinovat'sja vlasti Y, **odnovremenno v odnom meste**, ispol'zuja oružie, **protivozakonno dejstvujutII.2, kauziruja1 uščerb**, na predstavitelej Y-a s cel'ju likvidirovat' Y; Govorjaščij otnositsja k ètomu sobytiju otricatel'no'
numerous people Xs that belong to one grouping, having publicly refused to obey authorities Y, simultaneously in.one.spot and using weapons unlawfully act**II.2,** caus-

ing**1** damage, upon representatives of Y with the goal of liquidating Y; the Speaker evaluates this event negatively'

However, the selection of the lexeme BUNT 'riot' or MJATEŽ 'rebellion' is blocked by the Third Lexicalization Principle (maximal matching), stated in 6.2 above. Sentences (11) and (12) are absolutely correct as such, yet they cannot be accepted as the expressions of the starting meaning (5):

(11) *Bunty/Mjateži v Irake priveli k gibeli semi čelovek*
'Riots/Rebellions in Iraq led to the death of seven people'.

(12) *V rezul'tate buntov/mjatežej v Irake pogibli sem' čelovek*
'As a result of riots/rebellions in Iraq, seven people died'.

The selection of the lexeme BUNT 'riot' or MJATEŽ 'rebellion' causes essential modification of the meaning (5). On the one hand, a very important idea of reciprocity is lost: 'each_other'. On the other hand, informationally heavy semantic components are added – the orientation of Xs' actions**II.2** against the authorities; the definition of MJATEŽ 'rebellion' includes also the negative evaluation of the event by the Speaker (and we do not indicate here other discrepancies).[7] The losses/additions like these are disallowed.

7. Additional principles for the partial matchings of lexicographic definitions to the SemR

As indicated in Section 6.2, p. 123, the partial matchings of lexicographic definitions to the starting SemR require additional lexicalization principles. Here are four such principles established in the course of this investigation.

1. If, while matching the given definition to a SemR, a semanteme 'σ_{def}' coincides with no 'σ_{SemR}',
   then it is necessary to semantically decompose the semantemes involved, which can reveal the possibility of a partial matching.
2. If the Comm-dominant node '$\underline{\sigma}_{SemR}$' of the given Sem-Comm-area is part of the configuration '$\underline{\sigma}_{SemR}$←**2**–byt'**1.3** 'be' –**1**→σ'',
   then instead of '$\underline{\sigma}_{SemR}$' the node '$\sigma$'' must be taken for lexicalization.
   More specifically, in this case the generic component in the definition of the lexeme selected must coincide with 'σ''.

7. For the expression of the Speaker's negative attitude toward the event under consideration in the meaning of 'mjatež = rebellion', see Dobrovol'skij & Pöppel (2013).

3. If a definition matches the starting SemR only partially,
   then it is necessary to make sure that there is no contradiction between some semantemes of the definition and some semantemes of the SemR.
   In particular, the semantic constraints on the actants of the definition should not contradict such constraints on the corresponding actants in the SemR.
4. If a definition matches the starting SemR only partially,
   then an addition/subtraction of meaning (resulting from the use of this definition) must not be informationally relevant.

The notion of informational relevance is complex and needs a special investigation. Here, we have to limit ourselves to some particular remarks. An addition/subtraction of a semantic component 'σ' is possible/impossible in the following cases:

– If 'σ' relegates the considered fact/entity to a specific class, its addition/subtraction is possible: 'σ' does not change the description of the situation SIT.
– If 'σ' = 'each_other', its addition/subtraction is impossible, since it would lead to the description of a different situation.
– If 'σ' = 'the Speaker's evaluation of the situation', its subtraction is possible, but its addition is impossible.

Acknowledgments

The first sketch of this paper was read by Valentina Apresjan, Igor Boguslavskij, Svetlana Krylosova, Leonid Iomdin, Sébastien Marengo, Jasmina Milićević, Polina Mikhel, Alain Polguère and Leo Wanner. We would like to express here our heartfelt gratitude to these colleagues and friends for their constructive remarks.

References

Dobrovol'skij, D. & Pöppel, L. 2013. Lexical Synonymy within the Semantic Field POWER. In: Kor Chanine, I., (ed.), *Current Studies in Slavic Linguistics*, Amsterdam/Philadelphia: John Benjamins, 281–296. https://doi.org/10.1075/slcs.146.16dob

Iordanskaja, L. 1990. Ot semantičeskoj seti k glubinno-sintaksičeskomu derevu: pravila naxoždenija veršiny dereva [From a Semantic Network to a Deep-syntactic Tree: Rules for the Determination of the Tree Top Node]. In Saloni, Z. (ed.), *Metody formalne w opisie języków słowiańskich* [Festschrift Ju. Apresjan], Białystok: Filija Uniwersyteta Warszawskiego w Białymstoku, 33–46.

Iordanskaja, L. & Polguère, A. 1988. Semantic Processing for Text Generation. In *Proceedings of International Computer Science Conference '88*, Hong Kong, 310–318.

Lareau, F., Lambrey, F., Dubinskaitė, I., Galarreta-Piquette, D. & Nejat, M. 2018. GenDR: A Generic Deep Realizer with Complex Lexicalization. In: *Proceedings of the 11th edition of the Language Resources and Evaluation Conference (LREC)*, Miyazaki (Japan), European Language Resources Association, 3018–3025.

Lareau, F. & Wanner, L. 2007. Towards a Generic Multilingual Dependency Grammar for Text Generation. In Holloway King, T. & Bender, E.M. (eds), *Proceedings of the Grammar Engineering Across Frameworks 2007 Workshop* [CSLI Publications, http://csli-publications.stanford.edu], 203–223.

Mel'čuk, I. 2001. *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/slcs.57

Mel'čuk, I. 2012–2015. *Semantics. From Meaning to Text*. Vols *1–3*. Amsterdam/Philadelphia: John Benjamins.

Mel'čuk, I. 2016. *Language: From Meaning to Text*. Moskva/Boston: Academic Studies Press. https://doi.org/10.1515/9781618114570

Mel'čuk, I. & Polguère, A. 2018. Theory and Practice of Lexicographic Definition. *Journal of Cognitive Science*, *19*: 4, 417–470. https://doi.org/10.17791/jcs.2018.19.4.417

Milićević, J. 2007a. *La paraphrase. Modélisation de la paraphrase langagière*. Berne: Peter Lang.

Milićević, J. 2007b. Semantic Equivalence Rules in Meaning-Text Paraphrasing. In: Wanner, L., ed., *Selected Lexical and Grammatical Issues in Meaning-Text Theory*, Amsterdam/Philadelphia: John Benjamins, 267–297. https://doi.org/10.1075/slcs.84.09mil

Milićević, J. 2021. Modélisation de la paraphrase sémantique approximative. In: Marengo, S., dir., *La Théorie Sens-Texte. Concepts-clés et applications*, coll. « Dixit Grammatica », Paris: L'Harmattan, 51–74.

Polguère, A. 1990. *Structuration et mise en jeu procédurale d'un modèle linguistique décla-ratif dans un cadre de génération de texte*. Montréal: Université de Montréal [thèse de doctorat].

Sowa, J. F. 1976. Conceptual Graphs for a Data Base Interface. *IBM Journal of Research and Development*, *20*: 4, 336–357. https://doi.org/10.1147/rd.204.0336

Sowa, J. F. 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley.

Appendix

This Appendix contains a list of "non-obvious" semantemes supplied with necessary comments.

'byt'I.3 = be'
'X $Ø^{BYT'I.3}$ Y' = 'X is an element of the class Y'

'actions'

'X (voz)dejstvuet**I.1** na Y' = 'X acts**I.1** upon Y' [a semantic primitive; Agent X has no will]:
Volny$_X$ *b'jut o skaly*$_Y$ 'Waves$_X$ hit on the rocks$_Y$'.
'X (voz)dejstvuet**I.2** na Y' = 'X acts**I.2** upon Y' ['X (voz)dejstvuet**I.1** na Y, čto vlečët Z – izmenenie Y-a' = 'X acts**I.1** upon Y, which entails Y's modification Z']:
Èta kislota$_X$ *rastvorjaet kal'cij*$_Y$ 'This acid dissolves calcium' = 'This acid$_X$ causes**1** the dissolution$_Z$ of calcium$_Y$'.
'X (voz)dejstvuet**II.1** na Y W-om' = 'X acts**II.1** upon Y with W' [a semantic primitive; Agent X has a will]:
Ivan$_X$ *tolknul dver'*$_Y$ *nogoj*$_W$, *no ona ne poddalas'* 'Ivan$_X$ pushed the door$_Y$ with his foot$_W$, but it did not cede'.
'X (voz)dejstvuet**II.2** na Y W-om' = 'X acts**II.2** upon Y with W'
['X (voz)dejstvuet**II.1** na Y W-om, čto vlečët Z – izmenenie Y-a' = 'X acts**II.1** upon Y with W, which entails Y's modification Z']:
Ivan$_X$ *raspilil brevno*$_Y$ *novoj piloj*$_W$ 'Ivan$_X$ sawed the log$_Y$ with a new saw$_W$' =
'Ivan$_X$ caused**2**, by means of a new saw$_W$, that the log$_Y$ is sawed$_Z$'.
'X dejstvuet**II.3**' = 'X acts**II.3** Z' [a semantic primitive; Agent X has a will; action**II.3** has neither object, nor instrument]:
Ivan$_X$ *bežit* ⟨*kričit*, *vstaët*⟩ 'Ivan runs ⟨ shouts, stands up ⟩'.

'kauzacii = causations'

'X kauziruet**1** Z(Y) = 'X (voz)dejstvuet**I.2** na Y, čto vlečët Z(Y)
'X causes**1** Z(Y)' = 'X acts**I.2** upon Y, which entails Z(Y)' [Agent X has no will; X is the cause of Z(Y)]
Spirt$_X$ *kills mikroby*$_Y$ 'Spirit kills microbes' = 'Spirit$_X$ causes**1** the death$_Z$ of microbes$_Y$'.
'X kauziruet**2** Z(Y)' = 'X (voz)dejstvuet**II.2** na Y W-om, čto kauziruet**1** Z(Y)' =
'X causes**2**, by means of W, Z(Y) = 'X acts**II.2** upon Y with W, which causes**1** Z(Y)' [Agent X has a will; X is the causer of Z(Y)]
Ivan$_X$ *ubil Petra*$_Y$ *nožom*$_W$ 'Ivan killed Pëtr with a knife' = 'Ivan$_X$ caused**2** , by means of a knife$_W$, the death$_Z$ of Peter$_Y$'.

'konflikt = conflict'

Non-violent conflicts
'konflikt**I.1**' : state (*Direktor i naš šef naxodjatsja v konflikte* 'The director and our boss are in conflict'. | *dlitel'nyj semejnyj konflikt* 'prolonged family conflict' | *konflikt ètix naučnyx škol* 'the conflict between these scientific schools')

'konflikt**II.2**' : event (*Na sobranii proizošël konflikt meždu direktorom i našim šefom* 'During the meeting, a conflict happened between the director and our boss.')

Violent conflicts

'konflikt**II.1**' : state (*Armenija i Azerbajdžan uže davno naxodjatsja v voennom konflikte* 'Armenia and Azerbaijan have long been in a military conflict.')

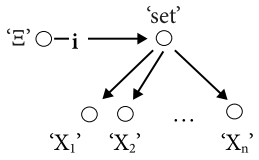
'konflikt**III.2**' : event (*Na granice Armenii i Azerbajdžana proizošël očerednoj voennyj konflikt* 'The border between Armenia and Azerbaijan saw another military conflict.')

'numerous'

'X mnogočislennyj = X is numerous' ≈ 'There are 7 or more Xs'

'set$_{\&}$', 'set$_{\vee}$' and 'set$_{\bar{\vee}}$'

The semantemes 'množestvo$_{\&}$ = set$_{\&}$' (logical conjunction), 'množestvo$_{\vee}$ = set$_{\vee}$' (inclusive logical disjunction) and 'množestvo$_{\bar{\vee}}$ = set$_{\bar{\vee}}$' (exclusive logical disjunction) are convenient abbreviations: the expression



means that the predicate Ξ bears on each element of the corresponding conjunctional or disjunctional set. Thanks to this, the Ξ predicate is not repeatedly present in the SemR. (Since the elements of this type of set are logically and semantically equal, the corresponding arrows of semantic dependency need not be distinguished, i.e. they are not numbered. See Mel'čuk 2012–2015: vol. 1, Chapter 4, 4.2.)

If the predicate Ξ does not bear on all the elements of the corresponding set, this fact is indicated by the existence quantifier in the subscript: Ξ∃ means that this predicate bears at least on one of the elements.

Address for correspondence

Igor Mel'čuk
Observatoire de linguistique Sens-Texte
Université de Montréal
Canada

igor.melcuk@umontreal.ca

Co-author information

Lidija Iordanskaja
Observatoire de linguistique Sens-Texte
Université de Montréal
lidija.iordanskaja@umontreal.ca

Publication history

Date received: 19 November 2021
Date accepted: 19 March 2021