

Unexpressed Objects in Russian*

Marjorie J. McShane

Abstract. This paper adduces evidence for “first internal argument” as an independent syntactic entity, regardless of case-marking, by virtue of similar behavior with respect to missing-object potential. Implications are explored for the machine translation into English of Russian sentences containing unexpressed objects, particularly for cases in which there is a mismatch for non-expression of the object in the two languages.

1. Introduction

A fundamental property of natural language is that not everything that can be said need be said. In fact, the topic of missing but understood elements covers a vast territory and poses some of the most complex questions in all of linguistics. Many approaches to unexpressed language information focus on one or another well-defined phenomenon in an isolated framework: for example, generative syntactic theories do not include full-fledged semantic or pragmatic components, so their treatment of unexpressed elements takes into consideration only a subset of the relevant contributing factors. Modularized study of a complex phenomenon tacitly promises that once the problems of each module are resolved, the larger phenomenon can be approached in bottom-up fashion. This being the case, modularization is a valid organizational principle. However, although theoretical paradigms permit strictly modularized work, practical applications often do not. A case in point is natural language processing (NLP), where the input is unrestricted text that does not offer the buffer of narrow topic definition. Simultaneously tackling many aspects of a complex phenomenon like missing elements can be daunting; however, it can also lead to insights inaccessible within a modularized approach. This paper explores a number of findings gained from concur-

* Sincere thanks to Sophia Lubensky, Ron Zacharski, and two anonymous reviewers from *JSL* for their helpful commentary on drafts of this paper. And to Leonard Babby for teaching me far more than case in Russian.

rent study of the range of contexts in which the first internal argument of Russian verbs can be unexpressed. It suggests generalizations that not only can be exploited in NLP, but can also be incorporated into theoretical work on case, ellipsis, lexicography, and semantics.

1.1. Case and the Scope of Inquiry

This study focuses on the subset of Russian objects that are direct-object like in function, meaning that they occur as the first internal argument of the verb, are bare NPs (i.e., not embedded within a PP), and are directly acted upon by the verb in a semantic sense. Objects that meet these requirements include those that are case-marked configurational ACC, configurational GEN induced by negation or quantification, and lexical GEN, DAT, and INST. These types of objects can be argued to form a natural group within the theory of Russian case developed largely by Leonard Babby (1980, 1984, 1991, 1994a, 1994b). According to Babby's classification, there are three sources of case-marking in Russian: **configurational**, **lexical**, and **semantic**.

Configurational case is assigned based on an NP's syntactic position: NOM for external arguments; ACC for direct objects that are not lexically case-marked and do not fall under the scope of negation or quantification; GEN for adnominal NPs and NPs that would have been configurational ACC had they not fallen under the scope of negation or quantification; DAT for recipients and experiencers; INST for certain types of predicate nominals. The large majority of Russian verbs take a configurational ACC first internal argument.¹

Lexical (quirky) case is unpredictable, semantically vacuous case-marking imposed by lexical-case-assigning verbs and prepositions. The only lexical cases that can be assigned by verbs to bare NPs are GEN, DAT and INST. (PREP is excluded since it only occurs within PPs.)

¹ Franks (1995: 53) refers to ACC case as "the least marked case" in Russian—the case that is assigned by a verb (or preposition) in the absence of feature specifications. Evidence for this assertion includes the following: in most declensional patterns, the ACC is like the NOM or the GEN in form; quantifiers often appear in fixed forms that look like the ACC case; when a preposition assigns two different cases with related meanings, one of these cases is always the ACC—e.g., *v gorod_{ACC}* 'to the city' ~ *v gorode_{PREP}* 'in the city' (Franks: 1995: 53). See Bailyn 1995 for a discussion of configurational GEN, DAT, and INST.

Semantic case derives from the semantic function of an NP in context, e.g., *lesom_{INSTR}* ‘through the woods’ in *On idet lesom* ‘He is walking through the woods’.

Configurational ACC, configurational GEN (of negation and quantification), and lexical oblique case form a natural group in the sense that they all appear on objects that are directly affected by the verbal action and might be called direct objects if that term were not so controversial.² This is not to say that objects with these types of case-marking are identical when subjected to theoretical analysis—quite the contrary. For example, configurational and lexical case are assigned differently; NPs with configurational case have been argued to occupy a different place in the X-bar structure than NPs with lexical case (Bailyn 1995); and objects with GEN-of-negation case-marking can carry semantic nuances that configurational ACC and lexical oblique objects lack. However, all of these objects have a similar semantic role and syntactic function, and they show similar potential to be missing at surface structure, as will be described throughout this paper.

The following types of oblique NPs are not semantically or functionally direct-object like and therefore will not be discussed here:

- DAT recipients in 3-place predicates and DAT experiencers in impersonal constructions (see, e.g., Komar 1999).
- Configurational INST complements of the copula, used in sentences like *On stal vračom_{INSTR}* ‘He became a doctor’. These do not have referential function and are therefore not subject to ellipsis.³
- Oblique objects of prepositions. Although some oblique objects of prepositions are semantically direct-object-like (e.g., *ženit’sja na kom* ‘to marry s.o.’), they are structurally embedded and raise tangential issues that extend beyond the scope of inquiry.
- INST NPs that reflect what Chvany (1996b: 163) calls “reduced transitivity”, as in the pair *švyrjat’ kamni_{ACC}* / *švyrjat’ kamnijami_{INSTR}* ‘throw rocks’. When the object in such pairs has canonical ACC case-marking, it is presented as a distinct participant in the action, whereas when it has INST case-marking it is presented “either as part of the agent or of the agent’s action”

² Hereafter, configurational GEN will refer only to GEN of negation and quantification; adnominal GEN plays no role in this analysis.

³ See Rothstein (1980: 79–82) for a discussion of NP reference.

(1996b: 163). INST objects of this type have non-direct-object-like properties that exclude them from the scope of this investigation.

- NPs with semantic case-marking, since most of these function as adverbials.

1.2. Sources of Missing Objects in Russian

The four types of missing-object phenomena to be discussed here are: (i) object ellipsis per se, defined as the non-expression of an obligatory (i.e., subcategorized-for) object that has a definite referent (antecedent) accessible in the linguistic or real-world context; (ii) non-selection of an object that is optionally selected by the verb; (iii) non-expression of an obligatory object that refers to generalized humans; (iv) non-expression of an obligatory object in a modal clause.⁴ All of these missing-object phenomena, with the exception of omission under modality, have received significant attention in the literature (cited in the relevant subsections); however, they have not been discussed as facets of a single complex problem. They **become** a single problem, however, in the computational environment—i.e., when one tries to program a machine to take the correct action when it is unable to find a filler for an element in the verb's subcategorization frame. Moreover, real-world examples show that there is extensive overlap between these phenomena, meaning that the topic to be explored is not simply four distinct phenomena, but also the five types of hybrid contexts that can arise from their overlap.

1.3. The Computational Connection

The current work was driven by computational necessity—more specifically, the demands of Russian-English machine translation (MT). Exploring missing-object phenomena from the computational perspective led to the following observations: (i) that objects with different case-markings have similar potential to be unexpressed; (ii) that computational lexicons must be significantly expanded to cover miss-

⁴ I do not consider objectless clauses headed by a derived *-sja* verb because intransitivization via *-sja* is a morpholexical process that removes the direct-object position from the verb's subcategorization frame; thus, derived *-sja* verbs obligatorily lack a direct object.

ing-object phenomena; (iii) that translation problems related to missing objects face not only MT systems, but humans trying to establish exact interlingual correspondences. Let us consider these issues in turn.

A prototypical MT system works as follows: first it parses the source-language text, encoding the meaning in a feature structure, then it uses that feature structure to generate a target-language translation.⁵ Building a parser or a generator involves writing language-specific phrase structure rules appended by feature constraints. A guideline for writing such rules is to write as few rules as possible to cover the necessary phenomena, thus making the computational grammar not only more efficient at run-time but also easier to debug and expand. It is for this reason that one seeks to bunch, rather than split, case behavior, maximally incorporating generalizations like “first internal arguments with any case-marking can be missing in context X”.

Another tenet of NLP is that bigger, more fully specified lexicons lead to better results (except, perhaps, as applied to sense splitting, which can become unmanageable). That is, a lexicon equipped with syntactic and semantic selectional restrictions, typical collocations, phrasals and idioms, an onomasticon, and access to an ontology should perform far better than a lexicon lacking such information. The Russian data presented here strongly suggest the need to advance the state of the art in lexicography (especially computational lexicography) by painstakingly describing the status of every object selected by every verb in every language.

The final insight gained from the computational perspective involves translation. When translating from a language like Russian, where objects are liberally omitted, to a language like English, which tightly restricts object omission, the task of filling in the blanks can pose problems. This lack of interlingual parallelism, which arises not only for machines but for humans, is easily overlooked until formalization becomes necessary. The Russian-English comparisons throughout the paper will highlight the non-trivial nature of this task as well as some notable sources of Russian ambiguity that have, to my knowledge, not been discussed in the literature.

⁵ This highly simplified description is intended only for purposes of orientation, not to reflect the complex and diverse nature of MT system architecture.

1.4. The Theoretical/Descriptive Connection

This study contributes to a number of subdisciplines of theoretical and descriptive linguistics, including case theory, ellipsis, lexicography, and semantics.

Much work on case in Russian has centered around finding similarities and differences among the morphological, syntactic, and semantic properties of NPs with different types of case-marking. For example, Jakobson (1958/1984) divides cases based on semantics, classifying NOM, ACC, and GEN as **central** cases, and DAT, PREP and INST as **peripheral** ones. Chvany (1986, 1996a) further supports the **central** nature of the GEN case, noting that it shares forms with the animate ACC, is highly differentiated morphologically (unlike the other oblique cases), is far more frequent than the other oblique cases, and alternates with the ACC in the negative. Bailyn (1995) argues that objects with configurational ACC and configurational GEN (of negation and quantification) case-marking occupy a different X-bar position than lexically case-marked objects. Fowler (1996) shows that lexical GEN and INST—but not DAT—are similar to configurational ACC in their ability to become subjects under passivization.⁶ These sample analyses show the trend toward using **same/different** as a focal point of the study of case in Russian.

The current study contributes to the SAME body of work by showing that configurational ACC, configurational GEN, and lexical oblique objects can be missing in many of the same contexts. It provides evidence for the existence of **first internal argument** as an independent syntactic entity that encompasses NPs with different case-markings, in a similar way that X-bar⁷ theory encompasses entities belonging to different lexical categories.

⁶ Fowler (1996: 538) writes: “Genitive and Instrumental complements of verbs which assign oblique lexical case are eligible for passivization because in syntactic terms, they are actually direct objects, with a superficial veneer of oblique morphological case”.

⁷ Some of the central issues of case theory as developed in the Chomskian framework—like the Case Filter and rules to resolve case conflicts (developed in Babby 1984)—are not discussed here as they do not apply to phonologically unrealized objects. A question that might occur to those working within the Chomskian framework is how the Projection Principle can be satisfied for missing objects that would have had lexical case-marking. The Projection Principle stipulates that subcategorization and theta-marking properties of a

The current analysis also contributes to the cross-linguistic work on ellipsis. Most of the literature on object ellipsis *per se* (as contrasted with other sources of missing objects) focuses on garden-variety, configurationally case-marked objects—both in languages that realize case morphologically and in those that do not.⁸ This raises the question ‘what of oblique objects?’. The collection of data presented below answers that question, at least for a subset of contexts in Russian.

The connection to lexicography can be described simply: until lexicographers take a far more rigorous approach to delineating subcategorization frames for verbs, lexicons will fail to support both the advanced needs of human users and computational applications. In the sections to follow, I will make specific proposals for expanding the object-related information found in comprehensive lexicons.

Finally, this investigation will reveal complexities in pinpointing a semantic interpretation for some missing-object contexts in Russian, as well as issues raised when trying to translate such contexts into English.

In sum, this investigation slices up linguistic reality rather differently than is traditionally done, dealing with a large chunk of missing-object phenomena at a single go and seeking practical approaches to the complex issues found therein.

lexical item must be satisfied at every level of representation; but lexical case-marking cannot be realized at surface structure if the argument it applies to is missing. However, Fowler (1996: 543) suggests an interpretation of the Projection Principle that circumvents such violation. In explaining why it is possible to passivize INST and GEN verbs (and thus not realize the lexical case of their objects) he writes: “lexical case-marking does not state that the verb must assign Instrumental or Genitive case to its object; rather, it states that the object, if it is case-marked, must not appear in the structural default Accusative case.” Extending this interpretation to missing-object phenomena, we can say that if a lexically case-marked object is overt, it must not appear in the ACC case. This does not stipulate that the object must be overt to begin with.

⁸ For treatments of object ellipsis see, e.g., Cole 1987, Farrell 1990, Franks 1995, Huang 1984, Law 1993, Lillo-Martin 1991, Massam and Roberge 1989, McShane 1998, 1999a, 1999b, 2002, Nilsson 1982, Padučeva and Ljaščenko 1973, Raposo 1986, Roberge 1990, Rögnvaldsson 1990, Saloni 1974, and Zemskaja 1987.

1.5. Organization

The remainder of the paper is organized as follows. Section 2 presents two types of discourse-affected contexts that permit the ellipsis per se of objects; section 3 discusses three types of missing-object phenomena that are clearly not elliptical; and section 4 turns to hybrid missing-object phenomena—i.e., instances of missing objects whose source cannot be precisely determined. All of these missing-object phenomena are shown to affect ACC and contain oblique objects similarly. Section 5 concludes the paper.⁹

2. Ellipsis

Ellipsis is the non-expression of a category whose referent is recoverable from the context. In Russian, the antecedent of an elided object can be syntactically overt or extra-linguistically understood. Object ellipsis is widely possible in Russian but is not permitted in English except in stylistically marked registers, like recipe contexts (*Open cake mix, pour [e] in bowl*), stage directions (*Approaches door, unlocks [e]*,

⁹ The examples in this paper were gathered from three sources: my card file of literary citations, the Tübingen Russian corpora, and native-speaker informants. Card-file examples are referenced by the author's last name (plus a number to disambiguate, if more than one work is cited) and the page number. Corpus examples are referenced by the first initial and last name of the author or the name of the periodical (further information and context can be found by accessing these examples from the corpus at <http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html>). Examples provided by informants show no reference. Examples from unedited speech are not included because they tend to include extensive pragmatic influences and production errors that would unduly complicate the current study. A number of conventions are used in the examples. Objects that are elided are indicated by [e], which is marked for case in the gloss. Objects that are missing due to non-elliptical or hybrid sources are indicated by Ø, also marked for case in the gloss. Since Russian has so-called free word order, the placement of pronouns is often variable, making the placement of [e] and Ø somewhat arbitrary in the examples (native speakers suggested their placement in each example). Glosses of reported speech are in double quotes, whereas glosses of narration are in single quotes. Antecedents are emboldened, as are all missing categories. Translation variants are shown as {variant a / variant b}.

*opens [e]), and other types of telegraphic language (*Dinner in fridge, heat [e] and eat [e]*).¹⁰*

Configurational ACC objects are widely elided in Russian, and examples like the following abound.

- (1) «Včera napolili žandarmov, svjazali [e],
yesterday got-drunk_{PL} gendarmes_{ACC} tied-up_{PL} [e]_{ACC}
sprjatali [e], priexali.»¹¹
hid_{PL} [e]_{ACC} came_{PL}
“Yesterday we got the gendarmes drunk, tied them up, hid
them, and came here.” (Švarc 1: 81)

For more discussion and examples of ACC object ellipsis in Russian see McShane 1998, 1999a, 1999b, 2002. The goal here is not to repeat those studies, but to compare the behavior of ACC and oblique objects in contexts in which both are found.

However, clear-cut instances of oblique-object ellipsis are not easy to find because in order to ensure that a missing object is due to ellipsis, one must rule out the possibility of object non-selection, non-expression due to modality, and non-expression due to a generalized human referent. Considering the small inventory of lexical case-marking verbs, and the correspondingly limited examples one finds in texts and corpora, these are stiff restrictions.¹² However, there are at least two types of contexts in which the ellipsis of both ACC and oblique objects is well attested: repetition structures, and structures with an extra-linguistic antecedent—both of which are significantly affected by discourse factors.

¹⁰ For a discussion of missing objects in English recipe contexts, see Massam and Roberge 1989. In addition, note that although examples like *That horse kicks!* exist in English, they are extremely few and cannot be considered a productive process.

¹¹ The context indicates that the plural verbs are 1st person (*we*).

¹² In fact, outside of the two configurations discussed below, I found practically no examples that unquestionably show the ellipsis of lexical GEN or INST objects. I found quite a number of DAT examples, but preliminary analysis led to a host of questions regarding the nature and function of the DAT case. I postpone this topic, with due regard for a reviewer’s opinion that pursuing theoretical aspects of it here would have been indicated.

2.1. Object Ellipsis in Repetition Structures

Repetition structures are what I call contexts in which two consecutive clauses contain the same verb selecting the same object(s). In almost all Russian repetition structures, the object in the second clause not only can, but should, be elided, regardless of its case-marking. The preference for object ellipsis derives from the function of repetition structures: to focus on the verb and defocus its argument(s). In repetition structures, one can repeat the verb alone for emphasis (2–3), repeat it with a modifier (4–5), or repeat it in a different tense and/or mood (6–7). The first example of each subtype has an ACC object, whereas the second has an oblique object.

- (2) Krasnoe nebo, uže načinaet vosxodit' luna, i ja
 red sky already starts rise_{INFIN} moon and I
 gnala lošad', gnala [e].
 drove horse_{ACC} drove [e]_{ACC}

'The sky was red, the moon on the rise, and I drove that horse,
 drove it hard.' (Čexov 2: 396)

- (3) «Čto šapka-nevidimka». — «Prelestno!» — «Kak tol'ko
 this-is hat-invisibilizer great as soon as
 vy ee nadenete, tak i isčeznete,
 you_{NOM} it_{ACC} put-on then PART will-disappear
 i bednyj master voveki ne uznaet,
 and poor mastercraftsman_{NOM} forever NEG will-know
 idet ona vam ili net. Berite, tol'ko ne
 suits it_{NOM} you_{DAT} or not take_{IMPER} only NEG
 primerjajte pri mne. Ja ètogo ne perenesu!
 try-on_{IMPER} in-front-of me_{DAT} I_{NOM} that_{GEN} NEG will-
 survive
 Net, ne perenesu [e]!»
 no NEG will-survive [e]_{GEN}

“This is a magic hat that makes you invisible.” “Great!” “As
 soon as you put it on, you disappear, and the poor hatmaker
 never knows if it looks good on you or not. Take it, just don't try
 it on in front of me. I couldn't stand it! I just couldn't stand it!”

(Švarc 2: 305)

- (4) «Ja ljublju **mat'**, sil'no ljublju [e] ... »
 I love **mother_{ACC}** strongly love [e]_{ACC}
 “I love my mother, I love her deeply...” (Čexov 2: 395)
- (5) Bylo vidno, čto gromkaja muzyka za stenoj mešala
 was visible that loud music_{NOM} behind wall bothered
babuške, očen' daže mešala [e].
grandma_{DAT} very-much even bothered [e]_{DAT}
 ‘It was clear that the loud music on the other side of the wall
 was bothering grandma — *really* bothering her.’
- (6) [The speaker warns the listener that everyone is mad at him for
 falling in love with the princess]
 «Vse oni gotovy s'est' **tebja**, i s'eli by [e]
 all they ready eat_{INFIN} you_{ACC} and ate CONDIT [e]_{ACC}
 segodnja že, esli by ne ja.»
 today PART if CONDIT NEG I
 “They're all ready to eat you alive, and they'd do it too, if it
 weren't for me.” (Švarc 4: 236)
- (7) «Ja dolžen vam ob"jasnit' svoi čuvstva, te,
 I must you_{DAT} explain_{INFIN} self's feelings_{ACC} those_{NOM}
 kotorye rukovodili **mnoj** i budut rukovodit' [e],
 which guided **me_{INSTR}** and will guide_{INFIN} [e]_{INSTR}
 čtoby vy ne zabluždalis' otnositel'no menja.»
 so-that you_{NOM} NEG got-led-astray with-respect-to me
 “I must explain my feelings to you, those that guided me and
 will continue to guide me, so that you don't harbor any
 misconceptions about me.” (Tolstoj 1: 531)

The next example shows a variation on this theme: the verbs have the same stem but different prefixes that convey semantic nuances as well as aspectual differences; however, the pragmatically-driven ellipsis is not affected by this slight lexical mismatch.

- (8) [A children's author is disagreeing with his publisher, who had just waved his arms to drive home his point]
- Ja snačala ne razmaxival **rukami**, no potom rasserdilsja
I at-first NEG waved **arms_{INSTR}** but then got-mad
i tože zamaxal [e]: «Ja sovsem ne xoču,
and also started-to-wave [e]_{INSTR} I totally NEG want
čtoby knižka ostalas' bez kartinok!»
that book_{NOM} remained without pictures
'At first I didn't wave my arms, but then I got mad and started waving mine too: "But I don't want my book to be without pictures!"'
- (Šinov: 12)

2.2. Ellipsis in Contexts with an Extra-Linguistic Antecedent

The second type of context in which at least configurational ACC and lexical DAT objects can liberally be elided is when the object has an extra-linguistic antecedent. (INST and GEN objects are excluded from this generalization only due to a lack of examples; there is no reason to believe they should resist ellipsis in such contexts.) Extra-linguistic antecedents stand in contrast to syntactic antecedents, which can license object ellipsis only in restricted configurations. In most cases, when ACC-object ellipsis is syntactically licensed, its antecedent is another ACC object located in the preceding clause, and the two clauses must have coreferential subjects.¹³ Thus, it is **not** the case that any previous mention of a person/thing at any point in the previous discourse makes that person/thing a licit syntactic licenser of object ellipsis.

Three discourse factors promote object ellipsis with an extra-linguistic antecedent:

- **The object is a 1st or 2nd person pronoun (*me/us* or *you*).** Participants in the discourse have particularly prominent status and can therefore be readily elided.¹⁴
- **The object is prominent in the real-world context.** Of course, all 1st and 2nd persons are prominent in the real-world con-

¹³ This generalization covers the most common cases, but there are plenty of exceptions, described in McShane 1998, 1999a, 1999b, 2002.

¹⁴ See Yokoyama 1986: 31–32 and Firbas 1992: 24 for discussions of the centrality of the speaker and hearer in the speech event.

text, but so are 3rd person objects that can be seen, heard, etc., by the interlocutors. These too can be assumed as part of currently active knowledge shared by the speaker and listener and, as such, can often be elided.

- **The selecting verb is in the imperative or interrogative mood.** Both of these moods require action on the part of the listener and assume at least a momentarily close speaker-listener relationship that promotes ellipsis whether the antecedent is syntactic or extra-linguistic.

The examples below show the ellipsis of configurational ACC and lexical DAT objects supported by various combinations of the above-mentioned factors. Examples (9–14) all have one of the interlocutors as the elided object. The examples represent pairs that differ only with respect to whether the object in question has ACC or DAT case-marking. Example (15) shows pragmatically licensed ellipsis of a 3rd person object.¹⁵

- (9) ANTECEDENT–PARTICIPANT; VERB–IMPERATIVE; OBJECT–ACC

«Lošadi moi stojat u kalitki. Ne provožajte [e],
 horses_{NOM} my_{NOM} stand at gate NEG escort_{IMPER} [e]_{ACC}
 sama dojdu... »
 self_{NOM} will-go

“My horses are at the gate. Don’t walk me out, I’ll go myself.””
 (Čexov 2: 440)

- (10) ANTECEDENT–PARTICIPANT; VERB–IMPERATIVE; OBJECT–DAT

[The stepmother, after looking over the dresses Zoluška made]

«U nas net osnovanij otvergat’ tvoju rabotu. Pomogi [e]
 at us no basis reject_{INFIN} your work_{ACC} help_{IMPER} [e]_{DAT}
 odet’sja» (Švarc 3: 541).
 get-dressed_{INFIN}

“We have no reason to reject your work. Help us get dressed.””

¹⁵ For more examples of ACC ellipsis with an extra-linguistic antecedent, see McShane 1998, 2002.

- (15) ANTECEDENT–3RD PERSON/THING; VERB–IMPERATIVE;
OBJECT–ACC
- Utrom vozle pravlenija k Goške
in-the-morning outside-of government-building to Goshka
podošel Ivan i, protjanuv svoj znamenityj
approached Ivan and having-held-out self's famous
pjatak, skazal zastenčivo: «Na, voz'mi [e].»
five-kopeck-coin said diffidently here take_{IMPER} [e]_{ACC}
'In the morning, outside of the government building, Ivan
walked up to Goshka and, holding out his famous five-kopeck
coin, said diffidently, "Here, take it."'
(Vojnovič 1: 39)

2.3. Conclusions About Object Ellipsis

In certain types of discourse-affected contexts—namely, repetition structures and contexts with an extra-linguistic antecedent—ellipsis potential extends to oblique objects as well as ACC ones. The descriptive/theoretical generalization is that the surface differences in case-markings are neutralized by the discourse factors responsible for licensing the object ellipsis. The computational implications are that formal grammar rules covering these elliptical contexts can be written in case-neutral terms, referring only to object status. Informal prose renderings of parsing rules for a Russian computational grammar might read as follows: If the same verb (regardless of tense, mood, modifiers, etc.) occurs in two consecutive clauses and all arguments are coreferential, expect the first internal argument in the second clause to be elided; if an obligatory first internal argument is absent in direct speech, retrieve the referent from the context (selected according to further rules).¹⁶ Thus, objects with similar elliptical properties provide generalizations applicable to both the descriptive/ theoretical and computational realms.

¹⁶ These rules must be amended when other sources of missing objects are incorporated.

3. Non-Elliptical Sources of Missing Objects

In some contexts, a missing object clearly derives from a non-elliptical source. The three primary sources of non-elliptical missing objects in Russian are non-selection, non-realization triggered by modality, and non-expression due to a generalized human referent—discussed in turn below.

3.1. Object Non-Selection

Traditionally, verbs that optionally select a direct object are classified as optionally transitive. However, this term is not ideal for our purposes because there is debate regarding whether verbs imposing lexical case-marking should be called transitive to begin with. Therefore, despite some awkwardness, I will use the term optional-object verbs instead. Optional-object verbs express a process-oriented action or state. Determining whether a verb optionally or obligatorily selects an object has nothing to do with case-marking—semantics is the key. Since verbal semantics cuts across human language, there is much cross-linguistic stability in the inventory of optional-object verbs: e.g., *čitat'* (NP_{ACC}) 'to read (NP)', *pet'* (NP_{ACC}) 'to sing (NP)'. Determining whether a verb obligatorily or optionally selects an object is fraught with many of the same complexities as distinguishing between arguments and adjuncts.¹⁷ But putting aside difficult cases, we can say uncontroversially that highly process-oriented verbs like *čitat'* 'to read' and *zanimat'sja* 'to study' optionally select an object in many languages, as shown in (16).

- (16) Tanja čitaet Ø / zanimaetsja Ø po utram.
 Tanya reads Ø_{ACC} / studies Ø_{INSTR} in mornings
 'Tanya reads/studies in the morning.'

Even some verbs that are less purely process-oriented permit object non-selection in both Russian and English—e.g., *protivorečit'* 'to contradict'.

¹⁷ For discussions of valency see, e.g., Günther 1978 and Somers 1984.

- (17) [Oni] primirilis'. Oni ne soprotivljajutsja, ne protivorečat
 [they] yielded they NEG resist NEG object
 \emptyset , ne vydvigajut nikakix svoix vstrečnyx predloženij,
 \emptyset_{DAT} NEG put-forth any self's counter proposals_{GEN}
 oni delajut to, čto im veljat [...]
 they do that which them_{DAT} order_{3.PL}
 'They have yielded. They are not resisting or objecting or
 putting forward any of their own counter proposals. They are
 just doing what they're told ...' (M. Kočnev)

Finally, there are a considerable number of verbs that have different subcategorization frames in Russian and English, showing that the status of an object is not entirely semantically determined. For example, the Russian verbs *mešat'* and *verit'* permit object non-selection, whereas their English equivalents *bother* and *believe* do not.

- (18) Navernoe, koe-čto iz nazvannogo možno bylo
 probably something of aforementioned possible_{IMPERS} was
 by rešit' na meste. Odnako mešajut \emptyset
 CONDIT decide_{INFIN} on spot but bother_{3.PL} \emptyset_{DAT}
 stereotipy myšlenija, skazyvaetsja deficit
 stereotypes_{NOM} thought_{GEN} is-visible deficit_{NOM}
 samostojatel'nosti.
 independence_{GEN}
 'Some of the abovementioned things could probably have been
 decided on the spot. But stereotyped thinking got in the way, a
 lack of independence was evident.' (*Pravda*, 88-08-29)
- (19) Večerom, vzjav v postel' pljuševogo zajca,
 evening_{INSTR} having-taken into bed stuffed hare_{ACC}
 on rasskazyval emu pro svoju buduščuju žizn' [...]
 he told him about self's future life
 Zajac veril \emptyset .
 hare_{NOM} believed \emptyset_{DAT}
 'In the evening, having taken his stuffed hare to bed with him,
 he told him about his future life ... The hare believed
 {him/what he said/it}.' (T. Tolstaja)

Such mismatches in subcategorization are particularly problematic for Russian-English MT because the parser will create a valid feature structure with no object, but the English generator will be unable to produce a grammatical English sentence without filling the slot for the verb's first internal argument. A solution to this problem, I propose, is to supplement the lexical entries for languages like English with **objectless workarounds**—turns of phrase that do not require making reference to the object. English *bother*, for instance, can be replaced by *get in the way*, and English *believe* can be replaced by *believe it*, in which *it* refers to the whole situation rather than a specific fact or statement. Such workarounds could be listed in a separate field in the lexical entry and accessed when English syntax requires an object that is not overtly provided in the Russian source. Of course, any prefabricated workaround will fail in some contexts; moreover, the task of determining which English verbs require workarounds and creating sufficiently generalized variants will not be an exact science. However, this type of lexicon supplementation would be a step toward bridging the chasm between Russian and English syntactic structures.

3.2. Object Non-Expression Triggered by Modality

Superimposing modality on a verb can relax the necessity of overtly specifying its object, even in languages like English, where objectless usage is relatively restricted.

- (20) a. *On ljubit.
he loves / is loving.
 - b. Ja znaju, čto ty umeeš' ljubit'.
I know that you know how to love.
- (21) Džindži byl zedorovyj i sil'nyj, umel xotet'
Dzhindzhi was healthy and strong knew-how want_{INFIN}
 \emptyset i točno znal, čego xočet.
 \emptyset _{ACC/GEN} and exactly knew what_{GEN} wants
'Dzhindzhi was healthy and strong, he knew how to want and
knew exactly what he wanted.' (Tokareva: 34)

As (20–21) show, both Russian and English have an object relaxation rule that applies to some portion of verbs that would normally

require an object except when used in a modal clause.¹⁸ However, the relevant subset of verbs is smaller in English than in Russian (as will be shown in the hybrid examples to come). The resulting potential for interlingual mismatch means that each verb in the Russian and English lexicons must be explicitly tagged as \pm [modal-induced object non-expression]. If both the Russian verb and its English translation permit modal-induced non-expression—as would be the case for *ljubit'* / *love*—translation is trivial, since neither side requires an object in modal contexts. If, however, the Russian verb permits modal-induced non-expression but the English equivalent does not, we are faced with having to fill in a blank on the English side. This is yet another place where the notion of objectless workaround might be exploited to expand the English lexicon in preparation for structural mismatches encountered in MT.

The possibility of object non-expression in modal structures derives in large part from the semantic function of the subset of modals in question—i.e., those expressing ability, liking, etc. These modal elements focus on one's ability or desire to carry out an action, with the object of the action being virtually irrelevant. As such, the semantics of the modal construction in combination with focus-related pragmatic forces override the lexico-syntactic necessity of overtly expressing the object of the main verb.

3.3. Non-Expression of Generalized Human Objects

Generalized human objects are commonly unexpressed in Russian, irrespective of their case-marking.¹⁹

¹⁸ In both languages, and perhaps even in all languages, highly object-oriented verbs reject object omission under modality on semantic grounds: e.g., *predpočitat'* (NP_{ACC}) 'prefer', *podvergat'sja* (NP_{DAT}) 'undergo'.

¹⁹ This is the objective correlate of the indefinite-personal construction: *Postučali_{3.PL} v dver'* 'There was a knock at the door'. Note that the non-expression of generalized human objects is possible even in some languages that do not permit object ellipsis, like Italian (Rizzi 1986).

- (22) Gore nikogda ne ubivaet \emptyset .
 grief never NEG kills $\emptyset_{ACC/GEN}$ ²⁰
 'Grief never kills anyone.' (Tolstoy 2: 101)

This is not ellipsis because there is no contextually bound referent, nor is it object non-selection because the understood human object is felt to play a role in the sentence. Unexpressed generalized human objects in Russian can refer to all of humanity, some contextually implied subset of humanity (e.g., women, children), or some specific person viewed as a representative of all of humanity (similar to the generalized and unexpressed use of *ty/vy* in Russian).²¹

There are both language-internal and cross-linguistic diagnostics suggesting that a missing object in Russian is due to generalized human non-expression. The Russian-internal diagnostic is: the verb requires an object; that object is human; there is no contextually available human referent; the clause is not modal. The cross-linguistic diagnostic is that an English translation of the sentence requires some type of sentence restructuring, like: (i) inserting *people, a person*, generalized *you*, or (in negated clauses) *anyone* as the object; (ii) passivizing the clause; or (iii) using a context-specific translation. (Recall that in most contexts English does not permit generalized human objects to go unexpressed, with the exception of lexically restricted examples like '*Your son bites!*'). The following examples show Russian contexts containing an unexpressed generalized human referent and their structurally non-parallel English translations.

- (23) «Vam ponadobilis' velikany... Oni tol'ko v skazkax
 you_{DAT} were-necessary giants_{NOM} they only in fairytales
 xoroši, a tak oni pugajut $\emptyset»$.
 good but otherwise they scare \emptyset_{ACC}
 "You needed giants... It's only in fairytales that they're good; in
 real life they {scare people/are scary}." (Čexov 1: 580)

²⁰ The verb *ubivat'* subcategorizes for an ACC object, but in this context would impose GEN of negation on its object.

²¹ Example (26) shows the subset 'women'. For more examples of this type, see McShane 1998 and 2002.

- (24) «Ja ostanus' zdes', ja budu borot'sja, drat'sja, kusat'sja, no
 I will-stay here I will fight struggle bite but
 unižat' moe čelovečeskoe dostoinstvo ne pozvolju Ø.»
 degrade my human dignity NEG will-permit Ø_{DAT}
 “I’m staying here. I’ll fight, struggle, bite, but {I won’t permit
 my human dignity to be degraded/I won’t let anyone strip me
 of my human dignity}.” (Vojnović 2: 94)

Selecting an appropriate English equivalent for generalized humans in contexts such as these is relatively simple for human translators but presents significant challenges to an MT system because the potential variants are often not interchangeable. Consider again examples (22) through (24), translated below using all the suggested variants. My judgments of the translations use the permissive standards that must be adopted considering the current level of MT performance.

(22) permits most of the basic variants:

Grief never kills {a person / people / anyone}.
 Nobody is ever killed by grief.
 ? Grief never kills you. [but *Grief can never kill you* is fine]

(23) permits only some variants:

...in real life they only scare {people / you}.
 ...in real life people are only scared by them
 ...in real life they only scare a person.

(24) permits only some variants:

... I won’t permit my human dignity to be degraded.
 ... I won’t permit anyone to strip me of my human dignity.
 *... I won’t permit you to strip me of my human dignity [loses
 generalization]
 *... I won’t permit {a person / people} to strip me of my human
 dignity.

Of course, in focusing on object realization in these examples, I am prematurely assuming that: (i) the parser can distinguish instances of non-expressed generalized human objects from other sources of miss-

ing objects; and (ii) other translation hurdles, like correctly passivizing a clause, are solved. These challenges are not, in fact, fully resolved; but one thing at a time.

To summarize, the main points about generalized human non-expression are as follows. Russian permits generalized human objects to be unexpressed whereas English does not. This means that translating into English requires using one of the many English expressions that convey generalized human meaning. Determining the heuristics for automated selection of the best among these for any given context remains a research issue.

3.4. Review of Non-Elliptical Missing-Object Phenomena

There are three sources of non-elliptical missing-object phenomena in Russian: non-selection of an optional object, non-expression triggered by modality, and non-expression due to a generalized human referent. English permits only the first two, and for a smaller subset of verbs than Russian. This creates the following table of possibilities (opposite) when seeking Russian-English correspondences.

Table 1. Russian-English Correspondences
in Verbal Subcategorization

Phenomenon	Russian	English	MT action
Non-selection	permits	permits	trivial
	blocks	blocks	trivial
	permits	blocks	use objectless work-around
Modal-Induced Non-Expression	permits	permits	trivial
	blocks	blocks	trivial
	permits	blocks	use objectless work-around
Generalized-Human Non-Expression	permits	blocks	select overt reference to generalized humans
	blocks	blocks	trivial

For each case, assume we have in mind a particular Russian verb and a particular English equivalent of it—both of which are supplied with the supplementary object information suggested earlier.

Trivial MT action means that the subcategorization frames for the Russian verb and its English equivalent match, so the status of the object in the Russian parse can be directly matched by the English generator. The MT action is not trivial when the Russian permits a missing object but the English does not. In such instances, if we can diagnose the source of the missing object, we can fill in the English object slot with the appropriate pre-fabricated filler: an objectless workaround or one of the overt references to generalized humanity. The initial assumption that we can diagnose different sources of object omission is necessary to get us beyond a logjam that would otherwise halt the investigation.

The above table of correspondences is, of course, valid whether translation is done manually or by machine. However, it is in the MT environment that the mismatches pose the larger problem.

4. Hybrid Phenomena

Although missing-object phenomena in Russian can be described as ellipsis as well as three other distinct linguistic processes, there is actually plenty of grey area between them, and the specific source of a missing object in a given context can defy clear analysis. We might think of missing-object phenomena in Russian as a broad plane with four distinct nodes whose spheres of influence radiate out with unclear borders. Each node independently accounts for some missing objects, but there are expansive areas of overlap between them, as shown by the lines in Diagram 1 on the following page. Although I have not found any examples of ambiguity between modal-induced non-expression and ellipsis, there is no reason to believe such examples do not exist.

Subsections 4.1 through 4.5 describe each of the attested hybrid phenomena in turn, concentrating on data presentation and analysis. The examples in each section include objects with both ACC and oblique case-marking. Section 4.6 discusses broader issues.

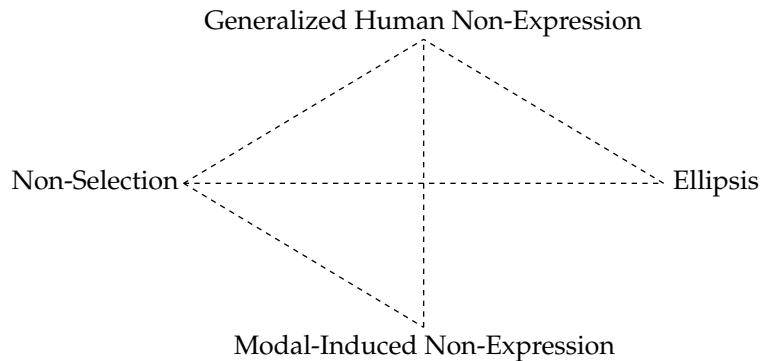


Diagram 1. Missing-Object Phenomena in Russian

4.1. Non-Selection or Generalized-Human Non-Expression?

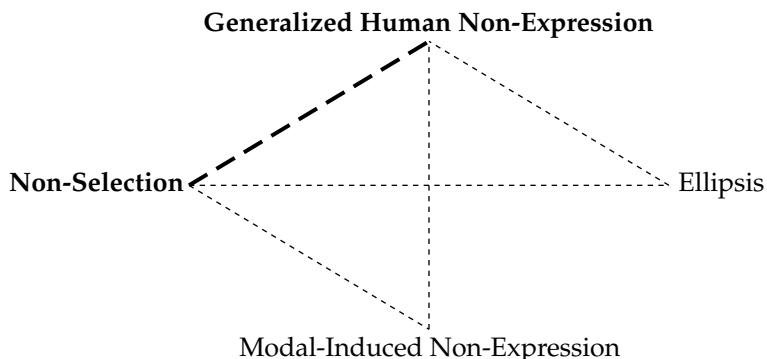


Diagram 2. Non-Selection *vs.* Generalized Human Non-Expression

The question of non-selection or generalized-human non-expression arises in Russian when the following four conditions hold: the verb optionally selects an object; that object is human; there is no contextually available human referent; the clause is not modal. In such contexts, either the object is not selected (implying slightly stronger verbal focus), or it *is* selected but not expressed because it refers to generalized humans (implying slightly stronger object focus). In Russian, there is no reason to choose between these interpretations because, in the first place, they are not incompatible, and in the second place, there is a very fine line distinguishing them to begin with. This slight ambiguity cannot, however, be preserved in English translation

because English does not permit the non-expression of generalized human objects. Therefore, in order to translate into English we must settle on one or the other interpretation of the Russian. If we lean toward object non-selection, we should either leave out the object, if the English verb permits, or else posit an objectless workaround. If we lean toward generalized-human non-expression, we should make overt reference to the object in the English translation. It must be emphasized, however, that in English, as in Russian, the two variants are often extremely similar semantically. This is because referring expressions like *people* can function as semantically impoverished place holders. For each of the examples below, two English translations are posited—one for the non-selection reading and one for the generalized-human reading. The semantic differences are discrete.

- (25) No teper'-to ugrožaet \emptyset ne nekij gipotetičeskij
 but now threatens \emptyset_{DAT} NEG some hypothetical
 gnev božij, my gibel'no opasny sami
 fury_{NOM} God's_{NOM} we_{NOM} deathly dangerous selves_{NOM}
 sebe svoim vozrossim moguščestvom, svoej
 selves_{DAT} self's growing power_{INSTR} our
 neobuzdannoj ènergiej, svoim neupravljaemym
 unbridled energy_{INSTR} our uncontrolled
 povedeniem.
 behavior_{INSTR}

'But now it's not some hypothetical divine fury that threatens { \emptyset / us}; we are mortally dangerous to ourselves by virtue of our growing power, our unbridled energy, our uncontrolled behavior.'

(V. Tendrjakov)

- (26) «Im, mužikam, liš' by obmanut' \emptyset , a naš
 them_{DAT} men_{DAT} only CONDIT deceive_{INFIN} \emptyset_{ACC} but our
 brat — baba — vsegda stradaet.»
 brother_{NOM} woman_{NOM} always suffers
 “For them, men, the idea is to deceive { \emptyset / us women}, and
 we're the ones who always end up suffering.” (Vojnović 1: 11)

- (27) «Čto ž, derevnja, govorjat, uspokaivaet Ø, on well PART country_{NOM} say_{3.PL} soothes Ø_{ACC} he pravil'no sdelal, čto tuda uexal.» correctly did that to-there left_{M.SG}
 "Well, they say the country {has a soothing effect/soothes a person}. He did right in going there." (Vampilov: 327)
- (28) Tam, naverxu, komandoval Ø Afonja Bronnikov.
 there upstairs commanded Ø_{INSTR} Afonja Bronnikov_{NOM}
 'There, upstairs, Afonja Bronnikov was {giving orders/ordering people around}.' (V. Rasputin)
- (29) Tan'ka zabivala v soplo ogromnuju kormovuju tykву, Tank'a shoved into nozzle huge feed_{ADJ} pumpkin_{ACC}
 a Veronika stojala rjadom i rukovodila Ø.
 whereas Veronika stood alongside and directed Ø_{INSTR}
 'Tan'ka shoved a huge pumpkin into the nozzle while Veronika stood alongside {giving orders / telling people what to do}.'
 (Tokareva: 497)

4.2. Modal-Induced or Generalized-Human Non-Expression?

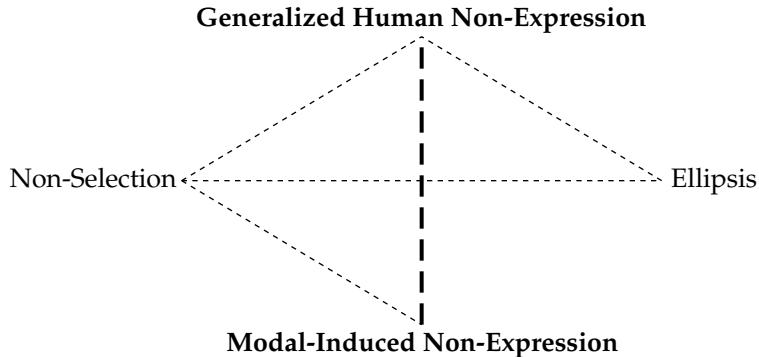


Diagram 3. Modal-Induced *vs.* Generalized Human Non-Expression

The question of modal-induced or generalized-human non-expression arises under the following conditions: the verb requires an object; that object is human; there is no contextually available human referent; the clause is modal. Since both modality and having a generalized human referent can permit the non-expression of otherwise obligatory objects

in Russian, it is impossible to select one or the other as the definitive source of the missing objects in examples like (30) through (32). Notice that (30) and (31) permit more or less acceptable English realizations of both interpretations, whereas (32) requires complete restructuring for context-related reasons.

- (30) On znal za soboj: umel nravit'sja Ø. No — he knew about self knew-how please_{INFIN} Ø_{DAT} but ljudjam opredelennogo sklada.²² people_{DAT} certain_{GEN} type_{GEN}
'He was well aware that he knew how to {please/please people}. But only a certain kind of people.' (Ju. Trifonov)
- (31) «Esli by ja byl vsegda takoj, kak teper', — podumal if CONDIT I was always such like now thought ja, — ja by ešče mog ponravit'sja Ø.» I I CONDIT still could please_{INFIN} Ø_{DAT}
"If I were always the way I am now," I thought, "I might actually be able to {? please / please people}." (Tolstoy 2: 82)
- (32) «Za takoj stil', konečno, nado ubivat' Ø.» for such style of-course necessary_{IMPERS} kill_{INFIN} Ø_{ACC}
"For style like that, a person should be killed." (Vojnović 2: 13)

The following example, although outside of the stated scope of investigation, shows two notable points: first, even **PPs** that are the first internal argument of a verb can be unexpressed and produce the ambiguity in question; second, even nouns like *talant* 'talent' can function as modals and thus permit modal-induced object non-expression both in Russian and in English.

- (33) «Samoe glavnoe, čto ostalos' v pamjati, — ego udivitel'nyj most important that remained in memory his incredible talant uxazivat' Ø.» talent woo_{INFIN} Ø_{PP-INSTR}
"The main thing that I remember is his incredible ability {to woo/to woo a girl}."

²² The appearance of the object *ljudjam* 'people' following the site of the missing object does not affect the analysis of the missing object itself.

4.3. Generalized-Human Non-Expression or Ellipsis?

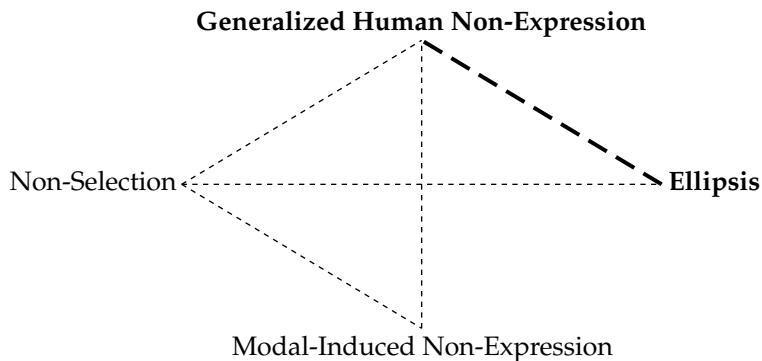


Diagram 4. Generalized Human Non-Expression *vs.* Ellipsis

The area of overlap between generalized-human non-expression and ellipsis represents contexts in which a specific human referent is available but the statement can be understood to apply more broadly as well. That is, a Russian speaker can assert something about a specific person and simultaneously generalize that statement to anyone else who might find himself in the same position. This ambiguity is similar to that produced by generalized *you* in English or Russian. For example, in describing my experience yesterday I could say: *It's unbelievable how strict they are around here: you walk into a meeting all of four minutes late and everyone stares at you as if you were some sort of deviant!* There is, however, one slight semantic difference between a missing object in Russian and generalized *you* in Russian or English: when a Russian object is absent, the generalizing effect is tempered by the still-available specific interpretation. The examples below show such slight ambiguity in Russian when the object is ACC (34) and oblique (35–36). The English equivalent *you* closely approximates the Russian meaning.

- (34) «А если читаete про seбja в gazetax?» — «Когда and if read_{2,PL} about self in newspapers when xvalyat Ø, priyatno, а kogda branjat Ø, praise_{3,PL} Ø_{ACC} pleasant_{IMPERS} but when criticize_{3,PL} Ø_{ACC} to potom dva dnja čuvstvuješ' seбja ne v duxe.» PART then two days feel_{2,SG} self NEG in mood “And if you read about yourself in the newspaper?” “When they praise you, it’s nice, but when they criticize you, you’re out of sorts for a couple of days.” (Čexov 2: 413)

- (35) Oba tjaželo vzdoxnuli, potom ešče raz, a
 both_{NOM} deeply sighed_{PL} then another time and
 potom — tak kak vzdyxat' im nadoelo, kak
 then since sigh_{INFIN} them_{DAT} bored_{IMPERS} as
 nadoedaet Ø rano ili pozdno vse na
 bores Ø_{DAT} sooner or later everything_{NOM} on
 svete, — Pančo vzjal v ruku kamešek i kinul
 earth Pancho took into hand stone and threw
 ego v staruju avtomobil'nuju pokryšku [...].
 it at old car tire
 'They both sighed deeply, then again, then — since sighing got
 boring, just like everything in life gets boring sooner or later,
 Pancho picked up a stone and threw it at an old car tire ...'
 (Šinov: 15)

- (36) ... ideja poezdki v London utopična, poskol'ku
 idea_{NOM} trip_{GEN} to London utopian since
 amerikanskaja potogonka ne daet Ø perevesti
 American rat-race NEG permit Ø_{DAT} catch_{INFIN}
 dux ni na minutku
 breath not for minute
 '... A trip to London is a utopian idea since the American rat
 race never gives you a breather.' (Račko: 9)

- (37) «Ty že znaeš', čto ot kritiki ja otošel,
 you PART know that from criticism I left
 potomu čto vser'ez zanimat'sja kritikoj Ø
 because seriously engage-in_{INFIN} criticism_{INSTR} Ø_{DAT}
 ne dajut, a ne vser'ez eju zanimat'sja
 NEG permit_{PL} and NEG seriously it_{INSTR} engage-in_{INFIN}
 ne stoit.»
 NEG is-worth-it
 “But you know that I stopped writing reviews because they
 don't let you do it right, and doing it any other way isn't worth
 the effort.” (Vojnović 2: 10)

4.4. Non-Selection or Ellipsis?

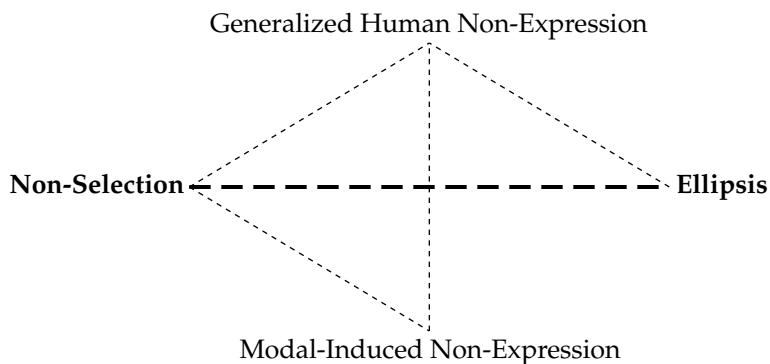


Diagram 5. Non-Selection *vs.* Ellipsis

The question of object non-selection or ellipsis arises when a referent for the missing object is contextually available, but the verb is not required to select any object at all. In such Russian contexts, one can lean more toward a process-oriented reading (and therefore non-selection), more toward a specific reading (and therefore ellipsis), or not lean in any direction at all: there is no need to resolve this ambiguity overtly in Russian. There is, however, a need to resolve it when translating into English because English, lacking object-ellipsis potential, provides no means of preserving the ambiguity.

Each of the examples below is provided with two English translations. The first represents the non-selection interpretation and conveys a slightly stronger verbal focus: *no medication helped; I sympathize; the others were envious*. The second represents the object-ellipsis interpretation and conveys a slightly stronger object orientation: *no medication helped him; I sympathize with you; the others envied her*. Again, as in earlier sections, the semantic difference between these English variants is minor.

- (39) «Požalejte menja, xorošij, dobryj čelovek». — «Vy znaete, pity_{IMPER} me good kind person_{NOM} you know ja sočuvstvju Ø vsej dušoj.» I sympathize Ø_{DAT} whole soul_{INSTR} “Pity me, good, kind person.” “You know, I wholeheartedly sympathize {Ø / with you}.” (Čexov 1: 588)
- (40) Odnoj byvšej zakadyčnoj podruge Ljuda podarila one former busom friend_{DAT} Luda_{NOM} gave-as-gift pomadu dlja vek, drugie zavidovali Ø. creme for eyelids others_{NOM} envied Ø_{DAT} ‘Ljuda gave one of her former bosom friends eyeshadow and the others {were envious/envied her}.’ (I. Grekova)

4.5. Non-Selection or Modal-Induced Non-Expression?

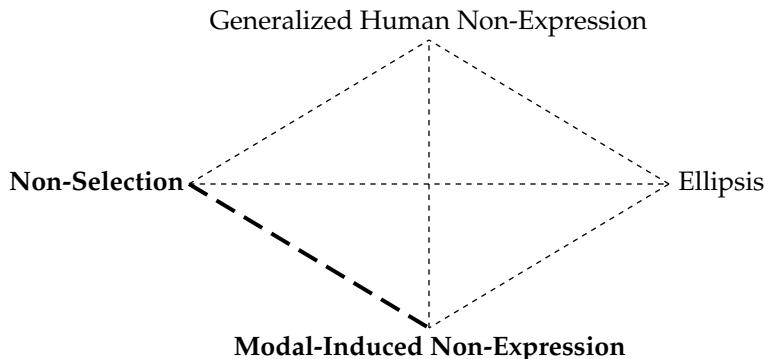


Diagram 6. Non-Selection vs. Modal-Induced Non-Expression

Object non-selection and modal-induced non-expression apply to different classes of verbs. Non-selection applies to optional-object verbs, whereas modal-induced non-expression applies to verbs that are lexically specified to require an object. For some verbs, however, object status cannot be unequivocally determined. This is true, for example, of *prosit'* 'to ask' and *trebovat'* 'to demand', whose objects are more necessary than the object of *čitat'* 'to read', but less necessary than the object of *podvergat'sja* 'to undergo'. *Prosit'* and *trebovat'* can be used without an object in non-modal clauses, but only in restricted kinds of contexts, like those expressing contrast: e.g., *Ja ne budu prosit', ja budu*

trebovat' 'I'm not going to request, I'm going to demand'. Thus, we must view object necessity/optionality as a continuum rather than as a binary option. In modal contexts containing verbs like *prosit'* and *trebovat'*, whose objects are of unclear status, it is practically impossible to settle upon a source for the missing object: non-selection or modal-induced non-expression.

- (41) On ne umeet prosit' Ø, on umeet
 he NEG knows-how ask_{INFIN} Ø_{ACC/GEN} he knows-how
 tol'ko trebovat' Ø.
 only demand_{INFIN} Ø_{ACC/GEN}
- 'He doesn't know how to ask, he only knows how to demand.'

4.6. Discussion

Perhaps the most important thread running throughout the analysis of hybrid contexts is that the semantic ambiguity produced by the Russian structures is minor. This is not surprising, since language systems generally do not foster gross ambiguity that could lead to confusion.²³

This "semantically similar" property of hybrid contexts can be exploited in the MT arena. That is, we have little to lose by simply selecting the Russian interpretation for which an English equivalent is most easily generated. Let us consider an example from this point of view, assuming (again, prematurely) that the Russian and English lexicons are expanded to include the following information for all verbs.

In both lexicons:

- object status in basic (non-modal) clauses
- object status in modal clauses

In the Russian lexicon only:

- whether generalized-human non-expression of the object is possible

²³ However, calling this ambiguity minor applies strictly to semantic interpretation, not to structural analysis. It would seem that when a speaker produces a sentence with a missing category, he/she must have a concrete source in mind: either ellipsis, non-selection, or one of the types of non-expression.

In the English lexicon only:

- objectless workarounds for verbs that require an object but whose Russian equivalent does not

Consider again example (30), repeated here for convenience.

- (30) On znal za soboj: umel nravit'sja Ø. No —
 he knew about self knew-how please_{INFIN} Ø_{DAT} but
 ljudjam opredelennogo sklada.
 people_{DAT} certain_{GEN} type_{GEN}
- 'He was well aware that he knew how to {please/please people}. But only a certain kind of people.' (Ju. Trifonov)

The Russian verb *nravit'sja* will be lexically specified as:

- [+ obligatory 1st internal argument]
- [+ modal-induced non-expression]
- [+ generalized human non-expression]

The verb *please* in English will be specified as

- [+ obligatory 1st internal argument]
- [+ modal-induced non-expression]
- [use generalized human workaround, if necessary]

When the parser cannot fill the obligatory first internal-argument position of *nravit'sja*, it will carry both of the null-object possibilities to the feature structure. When the ambiguous feature structure is used to generate English text, one of the interpretations of object status will have to be selected and realized. The default will be set to the simplest solution: here—leaving the object position blank under the interpretation of modal-induced non-expression, since the task of choosing the best variant from the generalized-human inventory (*people*, *a person*, etc.) is not trivial.

In short, considering that MT systems are still in their infancy and that their level of sophistication is still rather primitive, it is entirely acceptable to opt for the simplest solution to the hybrid-contexts problem: select whatever interpretation of the Russian will be easiest to generate on the English side. However, for the realms of theoretical

syntax, semantics, lexicography and translation studies, hybrid contexts present a rich territory for further exploration.

5. Conclusions

The investigation arose from the task of computationally processing unrestricted Russian text, which includes extensive empty categories, among which are missing objects of various profiles. The nature of the task forced a departure from well-defined and highly constrained realms of linguistic investigation, and the simultaneous exploration of phenomena generally discussed in diverse realms.

This paper first classified missing-object phenomena into elliptical and non-elliptical, creating three subcategories of non-elliptical phenomena: object non-selection, non-expression under modality, and non-expression due to a generalized human referent. Then it was shown that even this four-pronged classification is not sufficient to cover real-world contexts because missing-object phenomena overlap in Russian, creating slight ambiguity that is natural within the language system but raises question when one works cross-linguistically.

The data collected for this study showed that objects with configurational ACC, configurational GEN, and lexical oblique case-marking can participate in all of the missing-object phenomena under investigation, supporting the notion that there exists a theoretically solid notion of **first internal argument** that transcends morphological case-marking.

The necessity of dealing with mismatches in missing-object potential in Russian and English led to the proposal that computational lexicons (and even those intended for human consumption) be significantly expanded to include fine-grained classification of object status for both languages, and objectless workarounds in English to be exploited when translating certain classes of verbs.

Finally, and perhaps most importantly, this study has presented missing-object phenomena in Russian as a multi-faceted phenomenon that should be studied not only in its isolated parts, but also as an interconnected whole—not only to fulfill practical necessity like MT, but also to further theoretical work on linguistic subdisciplines including syntax, semantics, and translation studies.

References

- Babby, Leonard H. (1980) "The syntax of surface case marking". W. Harbert and Julia Herschensohn, eds. *Cornell working papers in linguistics* 1: 1–32.
- . (1984) "Case conflicts and their resolution: A contribution to EST case theory". *Cornell working papers in linguistics* 6: 1–21.
- . (1991) "Noncanonical configurational case assignment strategies". Almeida J. Toribio and Wayne E. Harbert, eds. *Cornell working papers in linguistics* 9: 1–55.
- . (1994a) "Case theory". Carlos P. Otero, ed. *Noam Chomsky: Critical Assessments*. London and NY: Routledge, 630–52.
- . (1994b) "A theta-theoretical analysis of adversity impersonal sentences in Russian". Sergey Avrutin et al., eds. *Formal approaches to Slavic linguistics: The MIT meeting*. Ann Arbor: Michigan Slavic Publications, 25–67.
- Bailyn, John F. (1995) *A configurational approach to Russian "free" word order*. Ph.D. dissertation, Cornell University.
- Chvany, Catherine V. (1986) "Jakobson's fourth and fifth dimensions: On reconciling the cube model of case meanings with the two-dimensional matrices for case forms". Richard D. Brecht and James S. Levine, eds. *Case in Slavic*. Columbus, OH: Slavica Publishers, 107–29.
- Chvany, Catherine V. (1996a) "Hierarchies in the Russian case system: for N-A-G-L-D-I, against N-G-D-A-I-L". Olga T. Yokoyama and Emily Kleinin, eds. *Selected essays of Catherine V. Chvany*. Columbus, OH: Slavica Publishers, 175–87. First published in 1982, *Russian language journal* 36/125: 133–47.
- . (1996b) "A continuum of lexical transitivity: Slightly-transitive verbs". Olga T. Yokoyama and Emily Kleinin, eds. *Selected Essays of Catherine V. Chvany*. Columbus, Ohio: Slavica Publishers, Inc., 161–68. First published in 1990 in Martina Björklund et al., eds. *Carina Amicorum, Carina Davidsson septuagenariae* 28.3.1990. Åbo : Pargas : Åbo akademis förlag, 51–63.
- Cole, Peter. (1987) "Null objects in universal grammar". *Linguistic inquiry* 18(4): 597–612.
- Farrell, Patrick. (1990) "Null objects in Brazilian Portuguese". *Natural language and linguistic theory* 8(3): 325–46.
- Firbas, Jan. (1992) *Functional sentence perspective in written and spoken communication*. Cambridge, England: Cambridge University Press.

- Fowler, George (1996) "Oblique passivization in Russian". *Slavic and East European journal* 40(3): 519–45.
- Franks, Steven. (1995) *Parameters of Slavic morphosyntax*. New York: Oxford University Press.
- Günther, H. (1978) "Valence in categorial syntax". Werner Abraham, ed. *Valence, semantic case, and grammatical relations*. International Congress of Linguists, 12th, University of Vienna, 1977. Amsterdam: John Benjamins, 127–56.
- Huang, C.-T. James. (1984) "On the distribution and reference of empty pronouns". *Linguistic inquiry* 15(4): 531–74.
- Jakobson, Roman. (1958/1984) "Morphological observations on Slavic declension". Linda R. Waugh and Morris Halle, eds. *Russian and Slavic Grammar: Studies 1931–1981*. Berlin: Mouton de Gruyter, 105–33.
- Komar, Eric. (1999) "Dative subjects in Russian revisited: Are all datives created equal?" *Formal approaches to Slavic linguistics: The Seattle Meeting*. Ann Arbor: Michigan Slavic Publications, 245–64.
- Law, Paul. (1993) "On null subjects and null arguments". *Canadian journal of linguistics* 38(1): 1–41.
- Lillo-Martin, Diane C. (1991) *Universal grammar and American Sign Language: Setting the null argument parameter*. Dordrecht: Kluwer Academic Publishers.
- Massam, Diane and Yves Roberge. (1989) "Recipe context null objects in English". *Linguistic inquiry* 20(1): 134–39.
- McShane, Marjorie J. (1998) *Ellipsis in Slavic: The syntax-discourse interface*. Ph.D. dissertation, Princeton University.
- . (1999a) "Predictive rules of direct object ellipsis in Russian". Katarzyna Dziwirek et al., eds. *Formal approaches to Slavic linguistics: The Seattle Meeting*. Ann Arbor: Michigan Slavic Publications, 329–48.
- . (1999b). "The ellipsis of accusative direct objects in Russian, Polish and Czech". *Journal of Slavic linguistics* 7(1): 45–88.
- . (2002) *What's missing in Russian?* Ms., University of Maryland, Baltimore County.
- Nillson, Barbro. (1982) *Personal pronouns in Russian and Polish: A study of their communicative function and placement in the sentence*. Stockholm, Sweden: Almqvist & Wiksell International. Translated from the Swedish by Charles Rougle.
- Padučeva, E.V. and T.K. Ljaščenko. (1973) "Ellipsis kak nulevoj anaforičeskij znak". *Naučno-texničeskaja informacija*, serija 2, #5. Moskva, 20–25.

- Raposo, Eduardo. (1986) "On the null object in European Portuguese". Osvaldo Jaeggli and Corvalan Silva, eds. *Studies in Romance linguistics*. Dordrecht: Foris Publications, 373–90.
- Rizzi, Luigi. (1986) "Null objects in Italian and the theory of *pro*". *Linguistic inquiry* 17(3): 501–57.
- Roberge, Yves. (1990) *The syntactic recoverability of null arguments*. Kingston: McGill-Queen's University Press.
- Rögnvaldsson, Eirikur. (1990) "Null objects in Icelandic". Joan Maling and Annie Zaenen, eds. *Modern Icelandic syntax*. San Diego: Academic Press, Inc., 367–79. (*Syntax and semantics*, 24.)
- Rothstein, Robert A. (1980) "Gender and reference in Polish and Russian". Catherine V. Chvany and Richard D. Brecht, eds. *Morphosyntax in Slavic*. Columbus, OH: Slavica Publishers, 79–97.
- Saloni, Zygmunt. (1974) "O możliwościach elipsy określenia czasownika w polskim tekście wielozdaniowym". M.R. Mayenowa, ed. *Tekst i język: problemy semantyczne*. Wrocław: Wydawnictwo polskiej akademii nauk, 59–81.
- Somers, Harold L. (1984) "On the validity of the complement-adjunct distinction in valency grammars". *Linguistics* 22: 507–30.
- Yokoyama, Olga T. (1986) *Discourse and word order*. Amsterdam: John Benjamins Publishing Company.
- Zemskaja, E.A. (1987) *Russkaja razgovornaja reč': Lingvističeskij analiz i problemy obučenija*. Moskva: Russkij jazyk.

Sources for Examples

- Vampilov, Aleksandr. (1984) "Prošlym letom v Čulimske". *Izbrannoe*. Moskva: Iskusstvo, 301–77.
- Vojnovič, Vladimir. (1963) *My zdes' živem*. Moskva: Sovetskij pisatel'. [Vojnovič 1]
- Vojnovič, Vladimir. (1988) *Šapka*. London: Overseas Publications Interchange Ltd. [Vojnovič 2]
- Račko, M. (1990) *Čerez ne mogu*. Tenafly, NJ: Hermitage.
- Tokareva, V. (1994) *Den' bez vran'ja*. Moskva: SP 'Kvadrat'.
- Tolstoj, Lev N. (1993). *Anna Karenina*. Moskva: Terra-Terra. [Tolstoj 1]
- Tolstoj, Lev N. (1973) "Detstvo". *Detstvo, otročestvo, junost'*. Moskva: Detskaja literatura, 5–104. [Tolstoj 2]
- Čexov, Anton P. (1967) "Višnevyyj sad". *Izbrannye proizvedenija v treх tomakh, tom III*. Moskva: Xudožestvennaja literatura, 555–606. [Čexov 1]

- Čexov, Anton P. (1967) "Čajka". *Izbrannye proizvedenija v trex tomach, tom III.* Moskva: Xudožestvennaja literatura, 391–442. [Čexov 2]
- Švarc, Evgenij. (1998) "Golyj korol'". *Drakon, Klad, Ten', Dva klena, Obyknovennoe čudo i drugie proizvedenija.* Moskva: Gud'jal press, 55–130. [Švarc 1]
- Švarc, Evgenij. (1998) "Drakon". *Drakon, Klad, Ten', Dva klena, Obyknovennoe čudo i drugie proizvedenija.* Moskva: Gud'jal press, 269–339. [Švarc 2]
- Švarc, Evgenij. (1998) "Zoluška". *Drakon, Klad, Ten', Dva klena, Obyknovennoe čudo i drugie proizvedenija.* Moskva: Gud'jal press, 533–71. [Švarc 3]
- Švarc, Evgenij. (1998) "Ten'". *Drakon, Klad, Ten', Dva klena, Obyknovennoe čudo i drugie proizvedenija.* Moskva: Gud'jal press, 195–267. [Švarc 4]
- Šinov, Č. (1987) *Poxiščenie avtobusa.* Translated by Maja Kačaunova. Sofia: Svat.

Dept. of Computer Science and Electrical Engineering

University of Maryland, Baltimore County

1000 Hilltop Circle

Baltimore, MD 21250

marge@umbc.edu

Date received: 16 December 2000

Date Revised: 19 March 2001