Below are excerpts from Marjorie McShane's *A Theory of Ellipsis*, Oxford University Press, 2005.

## CONTENTS

### Abbreviations

### Notes about Examples

1. Getting Started
2. Object Ellipsis: Preliminaries
3. Direct Object Ellipsis with a Like Antecedent
4. Direct Object Ellipsis with a Nominative Antecedent
5. Direct Object Ellipsis with an Oblique Antecedent
6. Elided Lexically Case-Marked Objects
7. Unexpressed Objects That Do Not or May Not Represent Syntactic Ellipsis
8. Head Noun Ellipsis . . . or Not?
9. Verbal Ellipsis with One Licensor
10. Verbal Ellipsis with a Combination of licensors
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Ellipsis of Minor Parts of Speech</td>
<td>178</td>
</tr>
<tr>
<td>12. Dependencies in Ellipsis: A Polish Case Study</td>
<td>190</td>
</tr>
<tr>
<td>13. More Elliptical Phenomena</td>
<td>198</td>
</tr>
<tr>
<td>Notes</td>
<td>235</td>
</tr>
<tr>
<td>References</td>
<td>243</td>
</tr>
<tr>
<td>Index</td>
<td>253</td>
</tr>
</tbody>
</table>
Getting Started

Syntactic ellipsis is the nonexpression of a word or phrase that is, nevertheless, expected to occupy a place in the syntactic structure of a sentence. For example, in Mary got an A on the math test and Louise θ a B, the verb 'got' in the second conjunct is elided. Although for many linguists syntactic ellipsis has come to represent the default interpretation of the term "ellipsis," ellipsis is actually a much broader phenomenon whose many aspects vie for immediate attention, particularly in the realm of natural language processing (NLP). For example, semantic ellipsis—the nonexpression of elements that, while crucial for a full semantic interpretation, are not signaled by a syntactic gap—occurs in I forgot my keys and He is reading Tolstoy, since the meanings are actually “I forgot to take/bring my keys” and “He is reading a book written by Tolstoy.” The fact that a component of semantics is, in fact, missing in sentences like these can be detected, among other ways, by cross-linguistic comparison: in Chinese, for example, one cannot use the elliptical read an author.

Ellipsis is a universal property of natural language, but its scope and means of realization differ substantially from language to language. Considering the ubiquity of this phenomenon, it may seem rather surprising that ellipsis studies are relatively undeveloped or, at least, lack breadth and depth of coverage. This state of affairs can be explained by the dizzying complexity of the phenomenon, which is not readily recognizable until one attempts to solve the ellipsis problem. Working on ellipsis—which requires reference to syntax, lexical semantics, discourse, prosody, semantics, and stylistics—is a prime example of doing linguistics across language modules.

The interaction between factors from different modules, while long of interest in linguistics, has not been a strong suit for linguistic theories, which—with the
exception of Optimality theory, recently expanded from phonology into morphol-
yogy and syntax—tend to wander little outside their respective domains (phonology, 
syntax, semantics, lexicology). The raw material of theoretical linguistics, language 
description, also lags behind on cross-modalist issues, with the lion’s share of time 
and energy devoted to the traditional domains.

This is not to say that ellipsis has not been productively studied; it most defi-
nitely has. For example, syntacticians have delineated licensing and recoverability 
conditions for instances of ellipsis subject to a syntax-oriented treatment; discourse-
theoreticians have incorporated ellipsis into models of theme-rheme structure and 
shared spaces of current concern between interlocutors, and descriptive and typo-
logical linguists—in addition to describing ellipsis potential in the world’s lan-
guages—have used ellipsis as one of the diagnostics for the presence or absence of 
so-called discourse orientation in language.

All current treatments have one thing in common: they address those subtypes 
or aspects of ellipsis that are most salient for a given established framework, be it 
thoretical or descriptive. I have a different interest that correspondingly requires a 
new framework: painting the whole canvas of ellipsis at a single go, considering all 
the factors from all the realms that determine when a speaker of some language can, 
should, or even virtually must elide categories. Moreover, although research in the 
trenches must be oriented toward some finite number of languages (here those lan-
guages are Russian, Polish, and English, with other languages treated in connection 
with specific phenomena), the results should be generalizable, with parametric varia-
tion across languages. With these goals in mind, in this work I develop an extensible 
theory of ellipsis that:

1. takes a bold step toward the crucial but, as yet, not commonly sought 
goal of formally addressing ellipsis as a cross-disciplinary linguistic 
phenomenon;
2. presents all elliptical phenomena within a single descriptive frame-
work, rather than limiting the scope of inquiry to those phenomena or 
even aspects of phenomena that neatly fall under the purview of 
some theory or modular approach;
3. incorporates data and phenomena from many languages and exploits 
literature from many domains, with the goal of ultimately creating a 
comprehensive inventory of cross-linguistic elliptical patterns;
4. develops a parameter-and-value-oriented description methodology that 
can be utilized to account for ellipsis potential in any natural language;
5. incorporates syntactic, semantic, and, as yet to a lesser degree, 
morphological ellipsis and relates all of these to the larger task of 
reference resolution;
6. is independent of established schools and thus, presumably, able to 
feed into many applications over a longer period of time than 
theoretically restricted approaches;
7. limits the complexity of the metalanguage of description yet pro-
vides means of extending this metalanguage dynamically in order to 
handle unforeseen phenomena;
8. embraces useful partial solutions if complete ones are currently out of reach, thus rejecting the all-or-nothing constraints imposed by many frameworks;
9. maximally formalizes the descriptions while writing in plain English, avoiding obfuscating jargon;
10. orients the description toward the goal of NLP, since doing so both imposes rigor on the description and in no way excludes its application to other realms; and
11. presents sample processing algorithms that show how the descriptive generalizations can be turned into useful rules.

1. Theory across language modules

In common usage, the word ‘theory’ tends to carry connotations of abstraction or hypothesis—and, indeed, in the realm of hard science this is the case. But for linguistic inquiry, the first definition of ‘theory’ in Collins English Dictionary seems a better fit: “a system of rules, procedures, and assumptions used to produce a result.” For example, although certain schools of theoretical syntax derive from the hypothesis that there exists an independent, innate language mechanism in the human brain devoted exclusively to syntax, most of the work done in such schools revolves around accounting for language phenomena using rules supported by notational devices. There is a similar lack of hypothetical or abstract quality to, say, lexicology, which is the theory of dictionary making that provides principles and guidelines for the more practical task of lexicography. In short, many existing theories of linguistics represent a means toward a formal description. With theory conceived of this way, it is hardly radical to propose a new theory to answer a new set of research goals.

However, not every formal description represents a theory, a statement that begs the question, what is a theory, after all? Fortunately, I am saved the labor of independently answering this question by adopting the definition of theory from Nirenburg and Raskin’s “Prolegomena to the Philosophy of Linguistics” (2004, chapter 2). Such adoption, of course, can reflect neither the depth nor the well-argued rationale of the original work; however, even in bare form the tenets convey enough self-evident logic to act as a scaffolding for this theory of ellipsis.

According to Nirenburg and Raskin, a theory can be defined as a combination of its function and its components. Its function must be to promote (1) selection of the best description methodology and (2) subsequent evaluation of the quality of the description, leading to iterative improvements in both methodology and description. The components of the theory—purview, premises, body, and justification—as well as methodologies to support theory building are discussed below as applied to this nascent theory of ellipsis.

1.1. The purview of the theory

The purview of this theory is all instances of syntactic, semantic, and morphological ellipsis in natural language that can be accounted for by rules of syntax, lexical and
compositional semantics, pragmatics, and combinations thereof. By “rules” I mean statements subject to formalization in a way suited for machine processing in the foreseeable (non-science-fiction) future. This excludes, for example, unexpressed material in highly colloquial speech, in which grunts and pointing can count for a speech act or in which background information known only to the interlocutors themselves renders unexpressed material recoverable. Not every phenomenon to be discussed here fully falls within this purview: for example, an elliptical pattern that has clear-cut syntactic and lexico-semantic components might have an intangible pragmatic aspect that provides the only available explanation for differing grammaticality judgments in minimal pairs. However, since one goal of the theory is to describe factors from all realms that bear on ellipsis and clearly delineate between those that can currently be formalized and those that lie beyond reach such hybrid phenomena are quite in place.

Obviously, this purview is too vast to be fully treated in one book, especially when the goal of the theory is, ultimately, to ensure coverage of all elliptical phenomena in all natural languages. So, here I lay down the foundations for the theory (its conceptual structure and raison d’être), treat select phenomena in detail, then show how the same approaches can be applied to other relevant phenomena.

Needless to say, there is no single correct way to divide up and treat elliptical phenomena and, as such, there is no self-evident logic in the organization of its analysis. As a means of orientation into my approach, I summarize here the top-level classification of elliptical phenomena and the representative sample of subtypes to be discussed in this book (chapters and subsections are noted in square brackets as chapter.subsection). Although all the material in the book is required to fully develop and motivate this classification, this snapshot should act as an initial lay of the land, helping the details of the exposition to fall into place.

**Syntactic ellipsis with coreference (i.e., with an accessible syntactic antecedent)**

[3, 12]

**Accusative object ellipsis with an Accusative object antecedent**

[4]

**Accusative object ellipsis with a Nominative antecedent**

[5]

**Accusative object ellipsis with an oblique antecedent**

[6]

**Head noun ellipsis**

[8]

**Gapping**

[9.1]

**Stripping**

[9.2]

**Sluicing**

[9.3]

**Verb Phrase Ellipsis**

[9.4]

**Multilicensor Verbal Ellipsis**

[10]

**The ellipsis of conjunctions and relative pronouns**

[11.1]

**The ellipsis of prepositions**

[11.2]

**The ellipsis of conditional particles**

[11.3, 12]

**The ellipsis of reciprocal and reflexive particles**

[11.4, 12]
Subject ellipsis [13.1.1]

Object ellipsis with an extralinguistic antecedent [13.1.2]

Nonfinite clauses [13.1.3]

Syntactic ellipsis without coreference (i.e., with no syntactically accessible antecedent)

The ellipsis of objects due to clause modality [7]
The ellipsis of objects with a generalized-human referent [7]
The ellipsis of objects in a series of actions [7]

Multilicenseor Verbal Ellipsis [10]
The ellipsis of conjunctions [11.1]
The ellipsis of relative pronouns [11.1]
The ellipsis of subjects with a generalized-human referent [13.2]

Semantic ellipsis (ellipsis of meaningful elements, but with no syntactic gap)

Unexpressed agents in passives [13.3.1]

Agentive impersonals [13.3.2]

Unexpressed experiencers and possessors [13.3.3]

Unexpressed arguments in derived nominals [13.3.4]

Unexpressed morphemes

Haplology [13.4]

Morpheme Ellipsis [13.4]

Morpheme loss during incorporation [13.4]

Language strategies

Dialogue strategies [13.5.1]

Sentence fragments [13.5.2]

Nominal sentences [13.5.3]

Unagentive impersonals [13.5.4]

One aspect of this inventory that might strike native speakers of English as unexpected is the many references to object ellipsis. Since standard English does not employ object ellipsis, the largely Anglo-oriented literature likewise makes little reference to it. However, I will show that object ellipsis, like many of the other phenomena listed here but not employed in English, offers particularly rich insights into the difficult problems of ellipsis as well as material to stimulate the search for practical and integrated approaches to their resolution.
Direct Object Ellipsis
with a Like Antecedent

The process of determining how different factors affect ellipsis involves holding certain factors stable while changing others, and seeking differences in ellipsis potential in minimal pairs. In this chapter, the control factors for studying object ellipsis in Russian are as follows:

- the potentially elided categories are direct objects with basic configurational (not quirky, lexical) case marking, which is ACC for Russian;
- all antecedents are, likewise, direct objects with configurational ACC case marking; and
- all antecedents are located in the clause that directly precedes the ellipsis clause, making them unquestionably syntactically accessible.

The parameters and their value sets relevant for this class of ellipsis are listed in table 3.1. They cover syntactic, lexicosemantic, and pragmatic factors that can affect ellipsis potential.

The status of direct objects ranges from "can't be elided" to "virtually must be elided," with many gradations between. For different applications, different approaches to object status might be appropriate. For example, an NLP system used to generate text for information analysts might be designed to produce overt objects in all instances, not seeking the stylistic refinement of variations on this theme. However, a system built to analyze text must be prepared for all instances of object ellipsis.

This chapter first closely analyzes Russian data, with the description organized at the top level according to syntactic structures. For each syntactic structure, the ellipsis effects of nonsyntactic factors are explored, like the nature of the antecedent
(R-expression or pronoun) and the choice of lexical items. After the major syntactic structures are covered, some additional lexico-semantic and discourse factors that apply to all of them are discussed. Next, cross-linguistic comparisons are drawn that provide initial validation for the inventory of parameters and values. Finally, sample processing algorithms are presented, showing how the description deriving from this theory might be applied in practice.

1. Coordinate structures

Coordinate structures consist of two or more conjuncts that occur in series and are or at least can be joined by a conjunction like and, or, or but. In the following English sentences, the conjuncts are bracketed for clarity:
(1) a. He [grabbed her pigtail] and [pulled it].
   b. He either [bought the pumpkin], [borrowed it], or [stole it].
   c. I [made a nice dinner] but [didn’t eat it].

In Russian coordinate structures that contain coreferential direct objects, the latter direct object(s) often can be and sometimes virtually must be elided. For example, whereas ellipsis is optional in (2), it is strongly preferred in (3). (The symbol \( \llbracket \) indicates that the variant is stylistically highly marked albeit not ungrammatical.)

(2) Он поднял мяч и бросил (ego).
   On podnjal mjač i brosil (ego).
   henOM picked-up ball\(_{ACC}\) and threw (it)\(_{ACC}\)
   'He picked up the ball and threw it.'

(3) Он поднял его и бросил \( \llbracket \) ego (\( \llbracket \) ego).
   On podnjal ego i brosil \( \llbracket \) it\(_{ACC}\)
   henOM picked-up it\(_{ACC}\) and threw \( \llbracket \) it\(_{ACC}\)
   'He picked it up and threw it.'

A number of factors contribute to determining DO ellipsis potential in coordinate structures. These factors can be thought of as parameters that have more ellipsis-promoting [+ ellipsis] and less ellipsis-promoting [− ellipsis] values, as shown in table 3.2. (I use [− ellipsis] rather than [− ellipsis] for the less ellipsis-promoting values in order not to imply that these values necessarily work against ellipsis; they simply do not promote it like the values marked [+ ellipsis].)

The information in this table can be recast as a series of logical generalizations. They are numbered according to the keys in the table.

1. The more coreferential direct objects there are in a sentence, the more desirable it is to elide one or more of them. For example, in English one would say I saw a firefly, chased it around the yard, caught it in my hands, studied it carefully, then let it go. Four overt its in a row is inevitable in English and thus sounds fine. However, in Russian a corresponding degree of pronominal repetition would be considered excessive. Therefore, in sentences composed of three or more con-

<table>
<thead>
<tr>
<th>Key</th>
<th>Parameter</th>
<th>[+ Ellipsis]</th>
<th>[− Ellipsis]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Number of conjuncts that contain</td>
<td>Three or more</td>
<td>Two</td>
</tr>
<tr>
<td></td>
<td>the same DO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Status of the conjunction</td>
<td>Overt</td>
<td>Absent</td>
</tr>
<tr>
<td>3</td>
<td>Status of the antecedent</td>
<td>Pronoun</td>
<td>Referential expression</td>
</tr>
<tr>
<td>4</td>
<td>Type of conjunct</td>
<td>VP</td>
<td>Clause (different-subject conjuncts)</td>
</tr>
</tbody>
</table>
juncts with coreferential direct objects, one or more of the objects should generally be elided. Which one(s) should be elided requires further analysis.

2. Overt conjunctions promote ellipsis by emphasizing the parallel nature of the conjuncts, since it is generally parallel elements that are joined by a coordinating conjunction (although see exceptions noted in chapter 1, section 3.6). All kinds of parallelism promote ellipsis in Russian.

3. There are at least two reasons that ellipsis is more strongly preferred when the antecedent is a pronoun than when it is an R-expression. The first reason is phonetic. Pronouncing two phonetically identical direct objects in close succession (it...it) sounds more repetitive than pronouncing two direct objects with different phonetic forms (ball...it). The second reason derives from properties of discourse: in order for you to refer to someone or something using a pronoun, that entity must either already be established in the language context or be visible/audible in the real-world context. Thus, a pronoun is merely a reminder of an already relevant and discourse-activated person or thing. When two pronominal reminders occur in close succession, the second tends to be elided in order to avoid undue repetition of the obvious.

4. When the subjects of the conjuncts are different, the second subject shifts the topic of discourse and decreases the expectation that the direct object will remain the same. Decreased expectation significantly impedes object-ellipsis potential in Russian.

1.1. Coordinated verb phrases

This section considers eight types of sentences that contain coordinated verb phrases (predicates). Each type is listed here showing the most common elliptical pattern for Russian. Summaries of the most common two-conjunct and three-conjunct patterns are found in tables 3.3 and 3.4. Further discussion and alternative ellipsis patterns are presented in the corresponding subsections to follow.

I. Он поднял мяч и бросил (его).
   Он поднял мяч и бросил (его).
   He picked up ball$_{ACC}$ and threw (it)$_{ACC}$
   ‘He picked up the ball and threw it.’

II. Он поднял его и бросил Ø.
    Он поднял его и бросил Ø.
    He picked up it$_{ACC}$ and threw Ø$_{ACC}$
    ‘He picked it up and threw it.’

III. Он поднял мяч, бросил его.
    Он поднял мяч, бросил его.

he_{nom} picked-up ball_{acc} threw it_{acc}
'He picked up the ball and threw it.'

IV. Он поднял er, бросил 0.
On podnal ego, brosil 0.
he_{nom} picked-up it_{acc} threw 0_{acc}
'He picked it up and threw it.'

V. Он нашёл мяч, поднял (его) и бросил 0.
On našel mjač, podnal (ego) i brosil 0.
he found ball_{acc} picked-up (it)_{acc} and threw 0_{acc}
'He found the ball, picked it up, and threw it.'

VI. Он нашёл er, поднял 0 и бросил 0.
On našel ego, podnal 0 i brosil 0.
h_{nom} found it_{acc} picked-up 0_{acc} and threw 0_{acc}
'He found it, picked it up, and threw it.'

VII. Он нашёл мяч, поднял er, бросил 0.
On našel mjač, podnal ego, brosil 0.
h_{nom} found ball_{acc} picked-up it_{acc} threw 0_{acc}
'He found the ball, picked it up, and threw it.'

VIII. Он нашёл er, поднял 0, бросил 0.
On našel ego, podnal 0, brosil 0.
h_{nom} found it_{acc} picked-up 0_{acc} threw 0_{acc}
'He found it, picked it up, and threw it.'

Type I 2 conjuncts ♦ R-expr. ant. ♦ Overt conj. ♦ Optional DO
On podnal мяч и бросил (его).
On podnal mjač i brosil (ego).
h_{nom} picked-up ball_{acc} and threw (it)_{acc}

Sentences of Type I are extremely common in Russian, and ellipsis of the second direct object is always optional:

<p>| Table 3.3 Two-conjunct patterns for DO ellipsis with a DO antecedent |
|---------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>No.</th>
<th>Antecedent</th>
<th>Conjunction?</th>
<th>Coreferential NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Referential expression</td>
<td>Conjunction</td>
<td>(Pronoun)</td>
</tr>
<tr>
<td>II</td>
<td>Pronoun</td>
<td>Conjunction</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td>Referential expression</td>
<td>Conjunction</td>
<td>Pronoun</td>
</tr>
<tr>
<td>IV</td>
<td>Pronoun</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
TABLE 3.4 Three-conjunct patterns for DO ellipsis with a DO antecedent

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>Referential expression</td>
<td>(Pronoun)</td>
<td>Conjunction</td>
<td>Ø</td>
</tr>
<tr>
<td>VI</td>
<td>Pronoun</td>
<td>Ø</td>
<td>Conjunction</td>
<td>Ø</td>
</tr>
<tr>
<td>VII</td>
<td>Referential expression</td>
<td>Pronoun</td>
<td></td>
<td>Ø</td>
</tr>
<tr>
<td>VIII</td>
<td>Pronoun</td>
<td>Ø</td>
<td></td>
<td>Ø</td>
</tr>
</tbody>
</table>

(4) С большим трудом преступник одалел следователя, вырвал у него приказ на собственный арест и спрятал Ø к себе в карман (Брагинский и Рязанов: 155).

S bol’šim trudom prestupnik odošel sledovatela, vyrval u nego priказ na sobstvennyj arest i sprijtal Ø k sebe v karman (Braginskij i Rjazanov: 155)

with great effort criminal ном overpowered inspector, tore from him order for

self’s arrest acc and hid Ø acc to self in pocket

‘With great effort the criminal overpowered the inspector, grabbed the order for his own arrest and hid it in his pocket.’

Ellipsis is possible in sentences of Type I even if a parenthetical phrase separates the antecedent conjunct and the elliptical conjunct, as shown in (5):

(5) Она подняла ладонь, словно чтобы защитить волосы от ветра, но сразу же опустила Ø — её стрижка лишала это движение всякого смысла (Пелевин: 105).

Ona podnjalaladon’, slovno’tochby zaščitit’ volosy ot vetra, no sužaže opustila Ø — eë striz’ka лишala eto dvizhenie vsjakogo smysla (Pelevin: 105).

she raised palm acc as if in order to shield hair from wind but immediately particle lowered Ø acc her hair cut ном deprived that movement acc all gen sense gen

‘She held up her hand, as if to shield her hair from the wind, but immediately lowered it—he short haircut made that motion entirely unnecessary.’

Type II 2 conjuncts  ♦ Pron. ani.  ♦ Overt conj.  ♦ Ellipsis preferred

Он поднял его и бросил Ø.
On podnal’ego i brosil Ø.
he nom picked-up it acc and threw Ø acc

Sentences of Type II are also very common in Russian. Since pronoun . . . pronoun tends to sound overly repetitive, ellipsis of the second direct object is consistently preferred. In fact, all of the collected examples of this type show ellipsis.

(6) «Запомни: только об одном мы и думаем, только одна у нас цель и есть — освободить тебя и увезти Ø с собой» (Шварц 1: 81).

«Zapomni: tol’ko ob odnom my i dumaem, tol’ko odna u nas sel’ i est’ — osvobodit’ tebja i uvezti Ø s soboj» (Švarc 1: 81).
Unexpressed Objects That Do Not or May Not Represent Syntactic Ellipsis

Syntactic ellipsis—that is, the non-expression of a syntactically obligatory category with a contextually determined referent—is only one of several sources of unexpressed objects. Others include:

1. The Nonselection of Optional Objects. Some verbs, like sing, are optionally transitive, meaning that they may or may not select an object. Nonselection occurs in sentences like The girl on the balcony is singing.
2. Object Nonexpression Triggered by Modality. In some cases, verbs that would generally require an object can be used without one if the clause has certain types of modality. For example, love in English is obligatorily transitive, but one can say I know that you know how to love. It is the modality of know how to that shifts the focus to the verbal process and licenses the nonexpression of the object (cf. *You love/*You are loving/*You always love).
3. The Nonexpression of Generalized-Human Referents. This type of object nonrealization is not found in English but is found in languages such as Russian and Polish. 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 use of you in English).
4. The Nonexpression of Objects in Series. When numerous verbs are presented in an action-focused series, their objects can often be
unexpressed, even if they are lexically specified as being mandatory. That is, the syntactic/pragmatic nature of the utterance can override the subcategorization requirements of the verbs in question. For example: People are really very much alike: they love, hate, marry, retire, have regrets and maybe some fun, then die. Although the English verbs love and hate have obligatory direct objects (and marry may or may not), in a series like this the focus is on actions, not their arguments, so object nonexpression is permitted.

Depending on what one considers a full syntactic structure, these phenomena may or may not represent syntactic ellipsis. But for semantically rich NLP, structures like the preceding ones certainly represent semantic ellipsis, since some type of object is always implied. Prerequisites for analyzing and generating such elliptical structures include the following:

1. The computational lexicon must carefully and consistently specify which objects are optional.
2. The system must include lexicosyntactic rules to the effect that objects become optional if used with a modal, and the computational lexicon must note the relevant modals as well as any verbs for which such object nonexpression is semantically impossible: for example, undergo.
3. The computational lexicon or ontology must specify selectional restrictions for verbs, indicating which ones can or must take a human object, and the computational grammar for a language like Russian must contain a rule that permits nonspecification of human objects.
4. The system must include lexicosyntactic rules to the effect that objects become optional if verbs are used in series, with system trials used to determine the cutoff point for what represents a series (perhaps three actions, perhaps four . . . ) and what other constraints might obtain.

For text analysis, each time a potential object gap is encountered in the syntax, several processes for resolving it must be launched. In the best case, all but one analysis will be excluded based on textual clues; however, more often multiple competing analyses will be valid, requiring some procedure for ranking outcomes. For text generation, a knowledge of the rules that govern the use of these types of missing objects and an understanding of their expressive power broadens the inventory of expressive means available to a language generator.

Missing-object phenomena can be described as five distinct linguistic processes: syntactic ellipsis plus the four strategies mentioned earlier. However, there is actually much gray area between them, meaning that the source of a missing object in a given context can defy clear analysis. For example, in a given context one might wonder: "Was the object selected, then elided, or not selected at all?" "Was it not selected, or was it selected but not expressed because it refers to generalized humans?"—and so on.
We might think of missing-object phenomena as a broad plane with five 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 in Figure 7.1. Each line represents hybrids—missing objects that cannot be unambiguously attributed to either of the sources at its end points. Although I have found no examples that are ambiguous between modal nonexpression and ellipsis, I do not exclude the possibility that such examples exist. In addition, examples whose elliptical sources are three-ways ambiguous could exist as well.

Figure 7.1 is an approximation of the real interaction of phenomena, which would require a far more complex graphic. The reason that Series Nonexpression does not fit neatly into the schematic is because a series of verbs can contain verbs with different lexical properties: some may have optional objects and others obligatory objects; some select humans and some do not; and so on. The placement of this phenomenon in the upper left corner is arbitrary.

The following subsections present examples and discussion of nine of the ten missing-object phenomena—all clear-cut and hybrid phenomena except syntactic ellipsis, which was covered in chapters 3 through 6. Since the missing objects in the following examples do not or may not represent syntactic ellipsis, they are not indicated by Ø. Instead, the verbs whose objects are in question are in boldface for emphasis.

1. Clear-cut nonselection

Clear-cut nonselection of an object occurs when an optionally transitive verb is used without an object either to convey a process (1a) or to convey a general state of affairs (1b):

(1) a. Таня читает на крыльце.
   Tanja čitaet na kryli'ce.
   Таня,νοм reads, masc. on porch
   "Tanya is reading on the porch."

b. Таня читает книгу.
   Tanja čitaet knigu.
   Таня,νom reads, fem. book
   "Tanya is reading a book."

Copyright © 2005. Oxford University Press. All rights reserved.
Verbal Ellipsis with a Combination of Licensors

All of the well-studied types of verbal ellipsis discussed in chapter 9 are licensed by one of two basic strategies: interclause parallelism (Gapping and Stripping) or a lexical licensor (Sluicing and VP Ellipsis). There is, however, another strategy for licensing verbal ellipsis that has not made it to the agenda of mainstream theories or practical systems: licensing by a combination of lexical categories, as illustrated by Russian examples (1) and (2).¹

   «Pochtaj, korol’». — «Kuda ty Ø?» — «Pojdu k sosednomu korolju» (Shvarc 1: 69).
   Good-bye King where DIRECTIONAL you NOM Ø will-go to neighboring king DAT
   “Good-bye, King.” “Where are you going?” “I’m going to see the neighboring king.”

2. [Discussing what seems like an unfair decision]
   «Ведь он не из мест Ø, вы понимаете, что не из мест?» (Вампилов 1: 26).
   «Ved’ on ne iz mest Ø, vy ponimaete, chto ne iz mest?» (Vampilov 1: 26).
   after-all he NOM NEG from revenge GEN Ø you NOM understand that NEG from revenge GEN
   “After all, he’s not doing this for revenge—you do understand that he’s not out for revenge?”

In (1) the combination of directional where and NOM-case you implies motion-related action, while in (2) the combination of NOM-case he and the adverbal not for revenge
implies behaving in some manner that must be understood from the context. So it is the combined semantics of the overt categories that both licenses the ellipsis and, sometimes with the help of the context, ensures recoverability of verbal meaning. I will refer to this as Multilicensor Verbal Ellipsis, or Multi-VE for short.

One salient aspect of Multi-VE is that often, but not always, a whole semantic class of verbs is implied rather than one specific meaning associated with one specific verb. The notion "semantic class" requires further specification. For example, when a verb of motion is elided, the motion might be on foot or in a vehicle, fast or slow; when a verb of speaking is elided, the speech might be storytelling, asking, lecturing, or blathering on; and when a verb of hitting is elided, the hitting might be punching, smacking, or walloping with a frying pan.

Multi-VE, which is highly productive in Russian and is also used, although more limitedly, in Polish and Czech, is almost exclusively found in colloquial speech. Because of this register restriction and because this is not among the simpler types of ellipsis to process, Multi-VE is probably not a priority for most current NLP systems. However, it is important to the current study for two reasons. First, it must be included in a full description of ellipsis in those languages that use it, whether or not one chooses to include rules for it in a given NLP system. Second, this is a realm where Ontological Semantic text processing can flex its muscles, so it serves as a good example of the goal toward which semantics-rich systems are moving.

1. Orienting the description toward processing

Unlike most of the types of ellipsis described so far, for which one could posit a relatively neutral description that could be applied to many fields, this subclass is best described with a class of applications in mind (like NLP), since the form of the description could vary significantly, being more oriented toward semantics, syntax, patterns with variables, and so on.2 Each of the Subsections 1.1–1.4 discusses a type of preparatory descriptive work that would be required for taking the approach to processing these sentences that is put forward in section 2.

1.1. Compile a list of syntactic structures

Analysis of extensive examples of Multi-VE shows that, in most clauses that employ it, the overt syntactic categories fall into a number of patterns. The most common of these patterns—whose elements can occur in any order in Russian—are shown in table 10.1 (when other patterns are used, they are limited lexically and should be covered by special rules). When turned into syntactic rules, this list will constitute an inventory of "legal" sentence structures that, when encountered, will trigger the search for the elided verb based on the semantics of the overt categories.

The licensors in table 10.1 are intentionally not called minimal licensors, which is a matter of more theoretical than practical concern. For example, whereas (1) and (2) each have two licensors, in examples like (3) it is more difficult to draw the line between licensors and facultative categories:
TABLE 10.1 Syntactic structures in which Multi-VE is commonly used

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Agent or experience</th>
<th>Tensed auxiliary or impersonal predicate word</th>
<th>PP or Adv</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agent&lt;sub&gt;SOM&lt;/sub&gt;</td>
<td>Tensed aux.</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>Agent&lt;sub&gt;SOM&lt;/sub&gt;</td>
<td>—</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>Tensed aux.</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>—</td>
<td>PP or Adv</td>
<td>Object</td>
</tr>
<tr>
<td>5</td>
<td>Agent&lt;sub&gt;SOM&lt;/sub&gt;</td>
<td>—</td>
<td>—</td>
<td>Object</td>
</tr>
<tr>
<td>6</td>
<td>Experimenter&lt;sub&gt;DAT&lt;/sub&gt;</td>
<td>Impers. pred. word</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>Experimenter&lt;sub&gt;DAT&lt;/sub&gt;</td>
<td>—</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>—</td>
<td>Impers. pred. word</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>—</td>
<td>Impers. pred. word</td>
<td>—</td>
<td>Object</td>
</tr>
<tr>
<td>10</td>
<td>—</td>
<td>Davaj(ie) ‘Let’s’</td>
<td>PP or Adv</td>
<td>—</td>
</tr>
<tr>
<td>11</td>
<td>—</td>
<td>Davaj(ie) ‘Let’s’</td>
<td>—</td>
<td>Object</td>
</tr>
</tbody>
</table>

(3) a. «Вели за билетами, если ещё хочешь ∅ со мной в кино» (Вампилов 2: 124).

«Вели за билетами, если ещё хочешь ∅ со мной в кино» (Вампилов 2: 124).
run<sub>PER</sub> for tickets if still want<sub>2SG</sub> ∅ with me<sub>INSTR</sub> to movies<sub>ACC</sub>

"Run and get tickets if you still want to go to the movies with me."

b. «Вели за билетами, если ты ещё хочешь ∅ со мной в кино».

«Вели за билетами, если ты ещё хочешь ∅ со мной в кино».
run<sub>PER</sub> for tickets if you<sub>NOM</sub> still want<sub>2SG</sub> ∅ with me<sub>INSTR</sub> to movies<sub>ACC</sub>

"Run and get tickets if you still want to go to the movies with me."

In (3a), the ellipsis is licensed and recovered based on the combination of want<sub>2SG</sub> and to the movies; with me is uninvolved in licensing. In (3b) there is one extra category that may or may not play a role in licensing the ellipsis: the subject, you<sub>NOM</sub>. Obviously, you cannot be considered a minimal licensor in (3b) because (3a), which lacks it, is grammatical. However, it is possible that you somehow facilitates the ellipsis without being minimally required to license it. Thus, categories within elliptical sentences can have a number of statuses: they can clearly be minimal licensors, perhaps be minimal licensors, not be minimal licensors but potentially facilitate the ellipsis, or not be involved in licensing/facilitating the ellipsis at all. For practical purposes, any category that potentially facilitates the ellipsis and is typically a part of the given elliptical pattern will be listed among the licensors. (There is no practical benefit to reducing the number of patterns.)

Since Russian permits ellipsis in general so widely, it is common for a given category to imply the existence of another, which can thereby be elided. This was discussed in chapter 1 with respect to direct and reverse valency. Some examples:

- Pattern 3 lacks but implies a subject, since the inflectional ending on the tensed auxiliary licenses its ellipsis and provides clues to the nature (person, number, and sometimes gender) of its referent.