Attentional and informational structure of Russian: interaction of reference, coherence, and word order

University of Pennsylvania
Undergraduate Honours Thesis in Linguistics

Sophia A. Malamud

My deepest gratitude is due to the people who have shaped my undergraduate encounter with linguistics: professors and teaching assistants in Linguistics and Cognitive Science. I want to thank especially my advisors: Ellen Prince and Mark Steedman, who have inspired and guided my growth in these disciplines. Their influence is evident in the content of this thesis. Without Ellen's prodding and generous advice this work would not have been possible. I am also grateful to Eleni Miltsakaki for lengthy and productive discussions of data and theory. Needless to say, all shortcomings and mistakes remain my own.

1 Introduction

The management of hearer's (reader's) attention is an integral part of cooperative communication in any language. The discourse is thus structured in a way that allows the hearer to focus his attention on various entities evoked in discourse, and to ensure that information about them is entered into his knowledge-store in a coherent way.

The universal properties of discourse structure are the subject of much ongoing research (Halliday 1967, 1985, Sgall *et al.* 1973, Givón 1984, Grosz and Sidner 1986, Vallduvi 1990, Grosz, Joshi, and Weinstein 1995, Walker and Prince 1996, Steedman 2000, Miltsakaki 2001, *inter alia*). At the same time, the language-specific tools that speakers utilise to manage

information in discourse have also been a subject of much exploration (see, e.g. Steedman 1996, Walker, Joshi and Prince 1998, Kruijiff-Korbayová *et al.* 2000). These tools may include particular syntactic constructions or intonation contours, as well as word order and referential devices.

In his 1990 dissertation, Enric Vallduvi writes: "It has long been noticed that there are syntactic operations which are not triggered by the need to satisfy any known 'purely' structural requirement..., and which are logico-semantically vacuous as well" (Vallduvi 1990, p.12). Consider the following Russian sentences:

(1) Rodion ubil staruhu.
Rodion(NOM.) killed old-woman(ACC.)

(2) Staruhu ubil Rodion.
Old-woman(ACC.) killed Rodion(NOM.)
'Rodion killed the old woman'

No matter whether syntactic derivations of (1) and (2) contain optional operations or not, the sentences are truth-conditionally equivalent and perfectly grammatical. However, a very strong intuition for a number of linguists and non-linguistist native speakers suggests that these two sentences, when pronounced with the same intonation, cannot be used felicitously in the same context. A theory of linguistic competence must be able both to model the syntactic structures available in a given language, and to explore the import of such truth-conditionally vacuous syntactic variation.

Following Prince 1988 and Vallduvi 1990, among others, I shall be referring to this non-truth-conditional component of sentence-meaning as "informational component" of a sentence. Packaging this meaning into syntactic structures will be termed **information pack-**

aging or **informational structure** of a sentence (*ibid.*). A part of this activity is concerned with focussing the hearer's attention on a single entity in each sentence (the topic in the sense of Givón 1983 and Miltsakaki 1999).

On the extra-sentential level, navigating the hearer's attention form one topic to the next determines attentional structure of discourse.

The goal of this thesis is to investigate the attentional and informational structure in Russian written narrative. To this end, the following questions are raised and the answers to them attempted:

1) What are the principles determining the change and maintainance of local topics in Russian discourse? What counts as a potential centre of attention?

2) How is information packaged in Russian sentences? In particular, to what extent and in what way does the word order in Russian serve information packaging purposes?

This paper will proceed as follows: in **Section 2**, I will give an overview of available theoretical tools for studying informational and attentional structure of discourse, and evaluate them for the purposes of the present study. **Section 3** will present a corpus study of Russian written narrative which attempts to answer the first question of the present investigation. A discussion of Russian zero subjects in **Section 4** will partially address the issue of potential centre of attention. **Section 5** will provide some background information on Russian scrambling, and existing claims about the relationship between discourse structure and word order in Russian. A descriptive information-packaging rule, based on the suggestions of previous authors and the findings of **Section 3** will be proposed in **Section 6**. The rule will then be modelled in the framework of Set-Combinatory Categorial Grammar. Finally,

a brief conclusion will follow in **Section 7**. **Appendix** will provide some worked examples and references.

2 Background: information structure and entity-tracking

A variety of theories of information and discourse structure have been developed in the last several decades. They have addressed a wide range of theoretical and empirical goals; a proliferation of inter-related (and at times ill-defined) notions and terminology resulted. The problem is not new: already in 1973, Sgall *et al.* write, "This way of specifying [the notions of theme, rheme, and transitional material]... is not convincing enough for many linguists wanting to use new methods of linguistic description. No overt criteria have been hitherto given for distinguishing the degree of CD..." (Sgall *et al.* 1973, p.24). A group of current researchers in the field complains: "Comparison and use is further complicated by the fact that similar terms are also often used (not always knowingly) to refer to quite different theoretical entities" (Kruijiff-Korbayová *et al.* 2000, p.9).

In half a century of attempts to define the basic elements of discourse structure, several dimensions of this structure have been distinguished, and denotations were proposed for the terms mentioned by Sgall and others. Enric Vallduvi provides a clear and comprehensive review of this literature in his dissertation (Vallduvi 1990, especially pp.35-56). For the purposes of this thesis, I shall restate some of Vallduvi's points, and outline several of the theoretical tools available to the researchers in the last decade. Usefulness and usability for an empirical investigation of the function-form relationship have been the main criteria in determining which of these tools to use in the present study.

2.1 Topic-comment and Halliday-theme

Practically every researcher has suggested at least a two-part division of informational labour within a sentence, in which "information is concentrated on a subpart of the sentence, while the remainder is licensed only as an anchoring vehicular frame for that informative part to guarantee an optimal entry into the hearer's knowledge store" (Vallduvi 1990, p.35). In the theme-rheme dichotomy in the sense of Halliday 1967, the theme is a single element within a sentence which forms the vehicular part, while the informative rheme is defined as the theme's compliment. In this, Halliday's approach is equivalent to the topic-comment dichotomy first formulated in Mathesius 1915 and later picked up by a plethora of researchers, including Hocket 1958, Strawson 1964, Gundel 1974, 1988, Kuno 1980, *inter alia*. The topic/theme is used as an address for the hearer under which to enter new information (comment) in his mental knowledge store.

Halliday defines this topic/theme to be sentence-initial. In this approach, we lose the utility of this term as designating a cognitive status of a sentence subpart, and cannot use it to study the correlation of cognitive status and syntactic realisation.

Other researchers, most notably Gundel 1974 and Reinhart 1982 say that the topic may be encoded by any constituent in the sentence. They set several tests to determine the cognitive status of sentence elements. These topichood tests all have at least three shortcomings: in many cases they fail to produce a single element of the sentence as the topic. More importantly, the tests are a matter of felicity judgements which are, at best, a shaky indicator of actual usage or processing. Finally, the topic-comment approach fail to distinguish a single most prominent element in the informative part of the sentence - however, the intuitive import of intonational focus seems to require such a distinguished element. The Praguean

topic-focus dichotomy, and several similar frameworks address this third problem.

2.2 Topic-focus, Firbas-theme, Steedman-theme, and open propositions

In his 1964 work, Firbas defines theme and rheme in terms of communicative dynamism. Under this view, the rheme is the most informative element in the sentence, i.e., one possessing the highest degree of communicative dynamism. Sgall *et al.* 1973 utilise the notion of contextual boundedness to define a similar division.

In a series of works in the seventies, Sgall, Hajičová and many others in Prague and elsewhere explore the topic-focus division as one of the factors determining the syntactic ('tectonic') structure of the sentence (see Kruijiff-Korbayová *et al.* for a review), or affecting its semantic import (von Stechow 1981). The Prague school theory of information structure is equivalent to the focus-presupposition dichotomy popular in American linguistics (Akmajian 1979, 1979, Chomsky 1971, Jackendoff 1972, Dahl 1974, Wilson and Sperber 1979, Ward 1985, Prince 1986, Lambrecht 1987, *inter alia*), and to the theme-rheme in the sense of Steedman 1991, 1996, 2000.

The topic/theme/presupposition/open proposition in these approaches is "the set of propositions evoked in an utterance which the speaker assumes the hearer already knows or believes or is ready to take for granted at the time of speech" (Lambrecht 1994). The focus/rheme is then defined negatively as the complement of the theme, the instantiation of the variable in the open proposition. Most researchers also note that a subpart of the focus is always marked prosodically by a pitch accent. This is an important operational criterion for distinguishing the focus/rheme, which can also be unambiguously determined for sentences that are answers to wh-questions.

Vallduvi 1990 proposes to unify the topic-comment and the topic-focus frameworks by defining a three-way informational partition of a sentence: focus, link, and tail. Vallduvi's focus corresponds to the variable-instantiating focus/rheme in the topic-focus and focus-presupposition approaches. The open proposition is called 'ground', which contains the special topic-like element and a 'tail'. The special element in the ground part of the sentence, the theme/topic that serves as a file-card/address in the mental knowledge store is aptly termed a 'link' (Vallduvi 1990, p.57). It is required to be sentence-initial in order to avoid the bog of ill-defined or non-operational topichood tests. While acknowledging the usefulness of this three-part division, the present thesis is an investigation of form-function relationship, and so in need of a cognitively-defined form-independent notion of a link. Such a notion is provided by the Centering Theory (Grosz, Joshi and Weinstein 1995, Brennan, Friedman, and Pollard 1987, Walker, Iida and Cote 1994).

2.3 Centering Theory

Joshi and Kuhn 1979 and Grosz, Joshi and Weinstein 1995 develop an entity-based model of local attention structure. By centering a single semantic entity in an utterance and predicating over it, the theory constrains the inferencing required for such tasks as anaphora resolution; computed transitions from one centre to the next model the local discourse coherence.

The Centering Theory has been proposed as such in Grosz, Joshi, and Weinstein 1983 (published as Grosz, Joshi and Weinstein 1995); the ideas were subsequently developed and expanded both by the original authors, and by others (Brennan, Friedman, and Pollard 1987; Walker, Iida, and Cote 1994; Walker, Joshi, and Prince 1998, *inter alia*). The theory

provides the most operational definition of a topic so far. It will therefore be utilised here as a formal framework for exploring the topic-structure of written Russian discourse. The basic notions of the theory are defined and discussed below.

2.3.1 The Centering Transitions

The Centering Algorithm allows us to compute the smoothness of transition between utterances based on salience ranking of entities in a discourse.

Definition 1: In each utterance, the set of discourse entities evoked in it is the set of *forward-looking centres* (Cf). Centres are semantic entities that are part of the discourse model (see Heim 1983), or items in the set of shared current concern for speaker and hearer (Yokoyama 1986).

Definition 2: There is a special member of this set called the *backward-looking centre* (Cb). This is the entity that is most central in the utterance (Walker and Prince 1996), the file card you're writing on (Reinhart 1982, Heim 1983), corresponding to "the utterance topic/theme" (Kuno 1980, Reinhart 1982). The Cb links the current utterance with the previous discourse.

The set of forward-looking centres is ranked according to discourse salience, or "activatedness". The factors that determine ranking are the crux of the Centering Algorithm. **By definition**, if any centres are evoked in the next utterance, the highest-ranked of them is the Cb of that utterance. In fact, the following **Pronoun rule** is established, reflecting the observation that the most cognitively prominent entities should need the least description for successful reference.

Pronoun rule (production prediction): if there is a pronoun in an utterance,

then the Cb of this utterance is also denoted by a pronoun.

Definition 3: The highest-ranked centre is the *preferred centre* (Cp). It predicts what the next utterance is going to be about.

The interaction between Cb and Cp determines smoothness of transition from one utterance to the next as shown in Table 1 below. When the most central entity in an utterance ($\mathrm{Cb}(\mathrm{U}_n)$) is the same as the most central entity in the previous one ($\mathrm{Cb}(\mathrm{U}_{n-1})$), and the same item is also predicted to be central in the next utterance ($\mathrm{Cp}(\mathrm{U}_n)$), the resulting discourse is very coherent, and the transition is *Continue*. On the other hand, when the Cb from a previous utterance is retained as such, but not predicted to be as salient in the next utterance, the transition type is *Retain*. The two Shifts result when the most central entity changes: the *Smooth-Shift* predicts that it should not change again in the next utterance, while the *Rough-Shift* does (Table 1).

Table 1. Transitions from $U_n - 1$ to U_n.

	$Cb(U_n) = Cb(U_{n-1})$	$Cb(Un) \neq Cb(U_{n-1})$
$Cb(U_n) = Cp(U_n)$	Continue	Smooth-Shift
$Cb(U_n)\|not = Cp(U_n)$	Retain	Rough-Shift

U_n - n^{th} utterance, $Cb(U_n)$ - the backward-looking centre of the n^{th} utterance, $Cp(U_n)$ - the preferred centre of the n^{th} utterance.

The transitions, smoothest to roughest are, thus: Continue, Retain, Smooth-shift, and Rough-shift (Walker and Prince 1996). Centering analyses have shown that smoother transi-

tions are preferred over rougher ones within a discourse segment (Di Eugenio 1998, Rambow 1993, inter alia).

2.3.2 The Ranking

The ranking of entities determines the Cp of the current utterance, and the Cb of the next one. The ranking principle arrived at by most Centering analyses (e.g. Di Eugenio 1998, Miltsakaki 1999), is based on the grammatical function of the entities, which are ranked as follows: EMPATHY → SUBJECT → OBJECT → OTHER. Here, 'empathy' denotes phrases grammatically marked as 'empathic' (e.g., in Japanese); the grammatical category was later transformed into a semantic one to include constructions in other languages clearly emphasising the experiencer (e.g., in the dative subject constructions (Yokoyama 1986)).

Studies of Italian (Di Eugenio 1998), Turkish (Hoffman 1998), and Greek (Miltsakaki 1999) have shown that this ranking indeed correctly predicts full noun phrase, pronoun, or zero-pronoun usage, and is independent of the utterance word order in these languages.

However, a study of German (Rambow 1993) showed that whereas topicalisation interacts with Centering in an ambivalent way, scrambling in the Mittelfeld directly affects the ranking: "the Cf (ordered set of forward-looking centres) of an utterance is the list of constituents of the Mittelfeld in that order." Thus, re-arranging constituents that follow the inflected verb (V2) will change the centre (topic) of the next utterance, and affect the local coherence of discourse.

2.3.3 The segment, the utterance, and other ranking assumptions

Centering is a model of local discourse structure: it operates within 'discourse segments.' Hence, it is important to know how to determine the segmentation of a discourse. However,

determination of segment boundaries is a separate question of much current investigation. For the purposes of this study therefore, I assume no *a priori* segmentation in written discourse outside of the writers' segmentation of large books into chapters.

Within each segment, the Centering algorithm calculates the Cf list for every 'utterance' - another notion in need of formal definition. Early Centering analyses seem to assume the utterance to be approximately the tensed clause (Kameyama 1998). In a later investigation (Miltsakaki 1999), this was revised, and 'utterance' was defined as a full sentence, i.e., "the main clause and its accompanying subordinate and adjunct clauses" (Miltsakaki and Kukich 2000). I follow here the revised definition. Miltsakaki 1999 argues that the ordering of subordinate and main clauses does not affect Centering. In this study, therefore, unless there were two or more coordinated subordinate clauses, nothing outside the main clause had significant effect on the Cf-list.

There has been much variation as to the correct ranking of entities within a complex noun phrase (e.g., possessive). In my corpus, the proportion of possessive noun phrases for which the various theories would predict different rankings is negligible, and so one may safely assume here the simple principle of left-to-right ranking.

3 Centering study of Russian

3.1 Pronoun rule

Recast in the terms of Centering theory, the question "What are the principles determining the change and maintainance of local topics in Russian discourse?" may be reformulated as follows:

What is the ranking principle for the list of forward-looking centres in Russian?

The two main possibilities for the ranking are the ones already stated for other languages: the ranking by grammatical function and left-to-right in the main clause. The two principles would rank the entities differently only in sentences with OVS, OSV, or VOS word order. A corpus study of written Russian was therefore conducted. The main source of data in this study was the online library of Russian literature (www.lib.ru). I have chosen a number of literary narrative segments containing scrambled sentences. A computerised search was used to select the segments containing scrambled sentences from electronic books. A total of 44 analysable segments of two or more sentences were found, each containing at least one scrambled sentence.

In an attempt to determine the ranking principle for Russian, the Centering Theory's Pronoun rule was utilised. The corpus was searched for sentences in which of the two or more possible Cb candidates only one was pronominalised. There were 16 such sentences in the data (cf. 127 total Cbs). For these $Cb(U_n)$, the word order of U_{n-1} and the grammatical role of the expression in U_{n-1} referring to the $Cb(U_n)$ were traced. The results are summarised in **Table 2** below.

Table 2. The Pronoun Rule study.

#	Word order in U_{n-1}	Grammatical role of $Cp(U_{n-1})$	Ranking principle favoured
1	SVO	subject	both
2	SVO	subject	both
3	SVO	subject	both
4	OSV	subject	by grammatical function
5	OSV	subject	by grammatical function
6	DativeVCompl(SV)	EMPATHY (dative experiencer)	both if EMPATHY is a factor in Russian, by word order otherwise
7	SprepGenitiveV	EMPATHY (genitive experiencer)	by grammatical function if EMPATHY is a factor in Russian, by word order otherwise
8	OSViO	subject	by grammatical function
9	OSViO	subject	by grammatical function
10	OSViO	subject	by grammatical function
11	InstrumentalVS	subject	by grammatical function
12	SVCompl(VS)	subject of the complement clause	neither
13	pPrepositionalVS	subject	by grammatical function
14	iOSVO	indirect object	by word order
15	SVO	subject	both
16	SVO	subject	both

As is evident from the table above, the 4th, 5th, 8th, 9th, and 10th (maybe also 7th, 11th and 13th) tokens speak in favour of the grammatical-function-based ranking, while the 14th and possibly 6th and 7th work left-to-right. However, the number of tokens is too small for any significant or definite conclusion, especially because the OVS and VOS U_{n-1} word orders are not attested here, and the tokens 8, 9, and 10 form a parallel-construction sequence in a paragraph. Thus, I turn to a statistical study of the corpus, hoping that tracing the actual entities chosen for centering by the narrators will provide a clearer picture of their salience in preceding discourse.

3.2 Comparison with scrambling-less text

3.2.1 Control Data

To provide a measure of the true proportions of different transitions in Russian texts, a full short story "Pyat' minut vzajmy" was chosen and analysed. A total of about 70 transitions was calculated. Since of the 78 sentences containing 24 transitive clauses with overt arguments only 4 were scrambled, the ranking was performed by grammatical function only. Discounting the rough-shifts in the opening and closing paragraphs of the story as "necessities of artistic considerations," the analysis has shown that Rough-Shifts constituted 10% of all the transitions, with the remaining comprising 34 Continues (48.5%), 16 Retains (23%), and 13 Smooth-Shifts (18.5%).

3.2.2 The Rough-Shift measure

Twenty native Russian speakers, representing a wide variety of ages (17-74 years), occupations (students, sociologists, physicists, computer technicians, businessmen, beauty salon workers, translators, housewives, and a mover), and geographical backgrounds (Moscow, Kharkov, Kiev, Leningrad, various suburbs in Russia and the Ukraine, and several immigrants in Brooklyn, New York) were chosen as informants. They were made to read both the control story and a matched-sized portion of the scrambled data corpus (ten read the control first, and ten the scrambled portion first), and were at the end asked to rate the 'perceived coherence' of randomly chosen discourse segments. The reading times and the ratings were recorded. Although some individual variation in reading speed was detected, there was no significant difference in reading times between the control and scrambled data. The ratings indicated that the scrambled corpus was perceived as slightly less coherent, which was at-

tributed to its fragmentary nature: "You have to switch your brain from one excerpt to the next," as one of the informants has put it.

The Centering analysis of this data was done manually twice (see the **Appendix A**). The first analysis utilised ranking by grammatical function and produced 50 Continues, 46 Retains, 16 Smooth-Shifts, and 17 Rough-Shifts out of 129 total transitions. Then, using the left-to-right ranking hypothesis, the second analysis was performed, producing 49 Continues, 46 Retains, 21 Smooth-Shifts, and 13 Rough-Shifts.

In their 2000 Centering study, Miltsakaki and Kukich argue that "in general, Continues, Retains, and Smooth-Shifts do not yield incoherent discourses" (Miltsakaki and Kukich 2000). Therefore, only the presence of a Rough-Shift signals a significant incoherence. The presence of Rough-Shifts in a perceptually coherent discourse and the number of Rough-Shifts were therefore the first considered factors in this study.

Statistical tests were run on these numbers, with the transition percentages from the short story analysis serving as controls, i.e., the norm. Although the tests indicated that the second analysis was much closer to the normal data, the sample was not large enough to yield a degree of certainty above 75%. Meanwhile, a qualitative evaluation of data was performed. A closer examination of the transitions has indicated that of the 17 Rough-Shifts produced by the first analysis, 6 were found to be Smooth-Shifts in the second. One of these could have been a Continue changing to Retain in the second analysis, depending on the judgement of the main clause boundaries. The remaining 11 were Rough-Shifts in the second analysis as well.

At the same time, out of the 13 Rough-Shifts produced by the second analysis, 2 were Smooth-Shifts in the first. One of these could have been actually a Smooth-Shift in the

second analysis if a different ranking principle for the complex noun phrases were adopted.

Based on the Rough-Shift measure alone, we can conclude that word-order dependent ranking provides a more accurate measure of discourse coherence than the grammatical-function based one. However, both rankings perform poorly on scrambled data: while the informants' judgements and reading times indicated that this text was similar to control data, the distribution of transition percentages was very significantly different for the two. The chi-squared test resulted in less than 1% probability that the difference is due to chance, for both ranking hypotheses (see Table 3 in the next section).

3.2.3 New hypothesis: incorporating the verb

The two analyses have produced approximately the same number of Continue and Retain transitions. Moreover, both analyses have 'improved' and 'worsened' about the same number of these transitions. This suggests that neither hypothesis sufficiently accounts for the more coherent data.

In the original left-to-right ranking hypothesis, no consideration has been given to the verb. However, it has been noted for many languages, including Russian, that the pre-verbal and post-verbal positions in an utterance have different informational functions (Yokoyama 1986, Rambow 1993, Kiss 2000, inter alia). Therefore, the position of the verb was traced in the 34 scrambled transitive sentences with overt arguments for which the two analyses give different transitions. For 15 of them the word-order dependent ranking (second analysis) produced a smoother transition, whereas for the remaining 19, the other analysis did.

Crucially, 12 of the former sentences had OVS and the remaining 3 - VOS word order, whereas 16 of the latter had the order OSV. The remaining 3 sentences contradicting the

left-to-right hypothesis were OVS. However, two of them were a part of the 6-utterance parallel construction segment, and one more a part of a segment in which calculation of segment boundaries and, therefore, of the Cbs was very difficult. It becomes obvious, thus, that simply scrambling the object to the sentence-initial position in Russian doesn't affect its discourse salience, but serves some other purpose. When, however, the subject is demoted to the post-verbal position, the salience of both entities is affected. The corpus contained no instances of the VSO word order; thus it is unclear whether post-verbal status of the subject is sufficient to demote it in the ranking, or whether it must follow the object as well.

The new ranking hypothesis is formulated as follows: **the entities in Cf are ranked by grammatical function, unless the subject is in the post-verbal position.** This revised hypothesis was used for the third and final analysis of the data. The analysis produced only 11 Rough-Shifts, 20 Smooth-Shifts, 35 Retains, and 63 Continues. These are the smoothest resulting transitions yet. The chi-squared test was used to measure the probability that observed differences in all the variables (the number of occurrences of each transition type) are the result of chance variation. Low chi-square values indicate that the distribution of transitions is essentially the same as in the control data set. In this test, the significance of the Rough-Shift measure is somewhat downplayed, since each variable is given the same significance in the calculation of the chi-squared value. Again, the percentages from the short story analysis were used as controls. As is evident from Table 3, the new hypothesis results in a significantly more normal analysis of the scrambling data (60% probability that the difference from normal is chance).

Table 3. The chi-squared test.

Ranking hypothesis	Chi-squared value	Probability
By word order	12.69	Less than 1%
By gram. function	16	Even less! (about 0%)
The new hypothesis	2.07	60%

Thus indeed, the scrambling that places the subject in clause-final position reduces the salience (topicality) of the subject and predicts an earlier element to be centered in the subsequent utterance. Possible reasons for this peculiar ranking principle are considered below in **Section 6**.

4 Anaphoric reference in Russian and its implications for Centering

4.1 Phenomenon and Data:

In the process of data collection for the Centering study reported in the previous section, we have noticed that Solzhenicyn's "Gulag Archipelago" contains many sentences in which third-person-plural verbs without overt subjects are used to express a kind of impersonal meaning (example (3) and (4) below).

(3) Kogda ne hvatalo boksov, delali eshche i tak.
When not it-was-enough of-cells, did also and so.
'When there wasn't enough cells, they would act in this way as well.'

(4) Elenu Strutinskuyu v novocherkasskom NKVD
Elena Strutinskaya(ACC) in Novocherkassy police-station
posadili na shest' sutok v koridore na taburetku...
forced-to-sit-down(3rdP.PL) for six days in hallway on stool...

'In the police station in Novocherkassy, Elena Strutinskaya was made to sit for six days in a hallway on a stool...'

The perceived agent in such a sentence is an indefinite subset of 'perpetrators of Gulag'.

Russian is not considered to be a pro-drop language (Franks 1995). Several researchers have noted, however, that it does sometimes allow zero subjects (see e.g., Avrutin and Rohrbacher 1995). Examples in (3) and (4) raise a number of questions on the referential status of zero subjects in Russian and their possible implications for Centering. In particular, one might wonder whether zero subjects in Russian ever refer anaphorically, or whether they appear only in impersonal constructions. Both possibilities have important implications for entity-based model of attentional structure. In the former case, the presence of zero anaphora will call for modification of Centering Theory's Pronoun Rule as in Turkish and Greek (Hoffman 1998, Miltsakaki 1999):

Modified Pronoun Rule: If an utterance contains a zero pronoun, then the Cb of that utterance is also realised by a zero pronoun. If an utterance does not contain zero anaphora, but does contain an overt pronoun, then the Cb of that utterance is also realised by an overt pronoun.

If, on the other hand, zero subjects do not refer anaphorically, a Centering study of segments containing zero-subject utterances may shed light on the status of inferrable entities in attention management. In particular, the entities inferable from the case-frame of the verb (e.g., agents) may or may not become potential centres for subsequent discourse.

In order to investigate these questions, a corpus study of written data was performed. For the purpose of tractability, the study was limited to zero subjects of only plural verbs. The corpus was "Master and Margarita," a novel by Mikhail Bulgakov (1984, online edition

at www.lib.ru), approximately 126 000 words. The online edition was searched for all plural verbs using the grep programme. Lines containing plural verbs whose subjects appeared in the same sentence were deleted. For verbs that had no overt subjects, the context at least up to the full sentence was retrieved from the text.

An examination of data showed that there are three main kinds of zero subjects in the corpus:

a. the ones that refer to old entities, just like pronouns do:

(5) Udivitel'noe u nih nastroenie. Vezut pokojnika, a
Amazing at them mood. (They)-are-driving dead-man, and
dumayut tol'ko o tom, kuda devalas' ego golova!
(they)-are-thinking only about that, where got-lost his head!
'They are in an amazing mood. They are driving a dead man [in a funeral procession], but thinking only about where his head went!'

b. the ones that evoke a new entity (indefinite zeroes):

(6) Poshel ja kupat'sya na Moskva-reku, nu i popyatili moyu odezhu, a etu
Went I to-swim to Moscow-river, well and (they)-stole my clothing, and this
dryan' ostavili!
shit (they)-left-behind!
'I went to swim in the river Moscow, so then someone stole my clothing, and left this shit behind!'

These indefinites were for the most part restricted by a locative phrase:

(7) V zale perestali dyshat'.
In auditorium (they)-stopped to-breathe.
'The people in the auditorium have held their breath.'

c. the ones that do not refer to any entities at all (impersonal constructions):

(8) Govorit Azazello, – skazali v trubke.
Speaking Azazello, – (they)-said in receiver.
' "Azazello speaking," was said in the receiver.'

The great majority of the data tokens were of the 'hard-to-tell' kind, ambiguous between the indefinite and non-referential readings. A psycholinguistic test of the pronominalisability of the referent in the subsequent utterance may show that these are not really ambiguous, but belong to one of the categories (b) or (c). No well-defined procedure allows me to disambiguate these using text-internal evidence.

(9) – A eto nas arestovyvat' idut, – otvetil Azazello.
– And this us to-arrest (they)-are-coming, – replied Azazello.
'[This means that] they are coming to arrest us, – replied Azazello.' or '[This means that someone is coming to arrest us, – replied Azazello.'

Besides, there were generic sentences with zeroes, which will not be treated in the present study.

(10) "Vot kak, okazyvaetsya, shodyat s uma!" – podumal on.
"Here how, it-turns-out, (they)-come-down from reason!" - thought he.
' "So this is how, it turns out, one loses one's mind!" - thought he.'

Application of the Centering algorithm to the (a)-kind zero subjects and grammaticality judgements of ten native informants revealed that these zeroes are Cbs of their own utterances, which can be realised as pronouns as well as zeroes. They can be pronominalised Cbs of the next utterance (in fact, they are preferred as Cbs of the next utterance). Unfortunately, there is not enough data to confirm or disconfirm any of the ranking hypotheses suggested in Section 4. A worked example of Centering a fragment containing zero subjects is included in the Appendix.

At the same time, the indefinite zeroes clearly introduce a new entity, and thus a potential centre into the discourse model. These entities may be grammatically referred to by a pronoun in the next utterance, as native speakers' judgements indicate. The indefinites

restricted by a locative phrase are even better pronominalisable in the next utterance. Consider, for example, the following sequence modified from Bulgakov (in the second utterance, the subject pronoun was inserted to co-refer with the zero subject of the first utterance).

(11) Tela Dismasa i Gestasa s vyklevannymi hishchnymi
Bodies(ACC) Dismas(GEN) and Gestas(GEN) with plucked-out by-wild
pticami glazami podnyali i totchas zhe brosilis' na poiski
birds eyes they-picked-up and immediately FOC threw-themselves on search
tret'ego tela.
of-third body.
'[They] picked up the bodies of Dismas and Gestas with eyes plucked out by wild birds, and immediately rushed to look for the third body.'

Oni ego obnaruzhili v ochen' skorom vremeni.
They it(ACC) discovered in very quick time.
'They discovered it very soon.'

The grammaticality and interpretability (under the intended reading) of the above example indicates that some of the zero subjects in Russian function like indefinite reference.

A further investigation of the factors which determine the interpretation and referential status of zero subjects in Russian is imperative, and is indeed under way now. To-date results indicate that situational verb aspect and immediately preceding discourse may be the main determinants of the zero subject interpretation.

5 Background: word order and information structure of a Russian sentence

We now turn to the second question of this thesis. In particular, the next two sections examine the claim that imformational structure of the sentence is encoded in word order in Russian.

5.1 Russian word order and motivation of scrambling

Russian is a 'free' word order language. This means that the word order in Russian does not encode 'who did what to whom.' Instead, the word order and suprasegmental phonology are used to encode different pragmatic and grammatical factors in an utterance.

In a Russian simple transitive "John killed Mary" theoretically all six permutations of the words yield grammatical sentences. Indeed, the following examples (3)-(8) are extracted from literary Russian prose and judged perfectly grammatical by a multitude of native speakers:

(12) Vozhataya rvanula elektricheskij tormoz.
Conductor(FEM)(NOM) yanked electric brake.
'The [female] conductor yanked the/an electric brake.' (SVO, writer: Bulgakov)

(13) Nemnogochislennyh svoih chitatelej zhurnal nemalo podivil.
Non-numerous(ACC) its-own(ACC) readers(ACC) magazine very-much surprised.
'The magazine surprised its few readers.' (OSV, writer: Solzhenicyn)

(14) YA ved' slovo svoe sderzhu.
I indeed word mine shall-keep.
'Indeed, I shall keep my word.' (SOV, writer: Bulgakov)

(15) Otkryla dver' devica, na kotoroj nichego ne bylo...
Opened door lass(NOM), on whom nothing not it-was...
'The door was opened by a lass who was not wearing anything...' (VOS, writer: Bulgakov)

(16) YEyo oskorblyal maleyshiy namyok na yeyo bednost'.
Her insulted smallest hint at her poverty.
'She was insulted by the smallest hint at her poverty.' (OVS, writer: Chekhov)

(17) Vzyal ya na kuhne svechechku.
Took I in kitchen candle(DIM)(ACC).
'I took a (small) candle in the kitchen.' (VSO, writer: Bulgakov)

However, a random association of a permutation with a discourse context typically produces infelicity. Intonation contours also constrain the use of different word orders. In fact, researchers of every background are turning to discourse for motivation of at least some cases of Russian scrambling.

In a position paper at the SLING2000 conference, John Bailyn writes, "Minimalism forces [the researchers in the generative syntax paradigm], with its emphasis on the *motivation* of movement, to finally address the hard questions about word order and it is here that functional and generative approaches to syntax will find their eventual reconciliation... Functional descriptions... are characterising the *motivation* of using such a structure, and generative grammar can... characterise the *means* of achieving the result needed for discourse purposes" (Bailyn 2000, pp.20-21, original emphasis).

A number of works investigating the phenomenon of scrambling have appeared in the past decade in the generative tradition (Mahajan 1990, King 1993, Bošković and Takahashi 1998, Bailyn 1995, Miyagawa 1997, Sekerina 1997, and many others). Some of these researchers have worked in a framework that allows optional (i.e., not motivated by syntax or semantics) movement (e.g., Mahajan 1990); later works claim that scrambling involves movement which, at least in part, is motivated by discourse factors (e.g., King 1993, Bailyn 1995, Miyagawa 1997, Sekerina 1997). Bošković and Takahashi 1998 argue that scrambled categories are base-generated, with an obligatory movement at LF. Many researchers have, however, suc-

cessfully argued that base-generation is ruled out for scrambling, since the latter is subject to island constraints (see Bailyn 1995, Sekerina 1997 for a detailed treatment). Bošković and Takahashi 1998 cite the following examle to show that scrambling, unlike wh-movement doesn't obey these contraints in Russian. However, at least three native linguists and a score of non-linguist native speakers find both (18) and (19) *e*xtremely ungrammatical.

(18) * Kto ty videl kogda t podjezžal?
* Who you saw when t was-coming?

(19) * Doktor ty videl kogda t podjezžal?
* Doctor you saw when t was-coming?
Intended reading: 'Did you see when the/a doctor was-coming?'

Thus, the researchers looking for the motivation of this movement turn to information structure (Functional Projection, or Focus Form) at least in some cases of scrambling (see Bailyn (in press) for a discussion of various cases).

At the same time, "in the Slavonic linguistic tradition, information structure has always been considered the main factor determining the linear ordering within a sentence" (Kruijff-Korbayová *et al.* 2000 vis. Mathesius 1939 and the Prague School). In her monumental 1986 study of Russian word order, Olga Yokoyama argues that the mechanisms of encoding the pragmatic and grammatical information in an utterance depend most heavily on the "speaker's subjective evaluation of the discourse situation" (Yokoyama 1986, p.331). That is, the word order and suprasegmental phonology that are used in Russian to encode this information depend on the speaker's assessment of the 'knowledge' and 'current concern' sets for those involved in the discourse.

Yokoyama's study was concerned with the spoken discourse, and her most definite conclusions dealt with Type 1 ('neutral') intonation contour for the utterances. Her work has

partially formalised the Prague school's 'theme-rheme condition,' which suggests that "an NP may scramble over less rhematic NP" (cited in Rambow 1993).

The multiplicity of factors affecting Russian word order have been explored in a very principled way by Yokoyama. It seems, nevertheless, to be an exceptionally difficult task to formalise the process of "subjective evaluation of the discourse situation." In this work, I will therefore turn to the notions of backward-looking centre, salience marking, and theme-rheme structure of a sentence in the sense of Steedman 2000 to explore the informational import of various word orders in written Russian narrative.

6 Attaching INFORMATION feature to Categorial Grammar for a fragment of Russian.

This section examines the information structure of simple transitive sentences in Russian from the perspective of Combinatory Categorial Grammar proposed for a fragment of Russian by Natalia Nygren (1999). I assume the syntactic analysis of Nygren 1999, and attempt to expand it to include the feature INFORMATION. To reiterate, the following discussion is limited for tractability purposes to simple transitive sentences with overt arguments.

In her 1999 work, Natalia Nygren proposes a Set-CCG syntax for transitive Russian sentences. Her analysis accounts for the coordination facts in Russian. Her work, however, is not concerned with exploring the informational import of the various possible word orders. The study reported in **Section 3** provides evidence that word order affects the attentional information in Russian, a claim supported by the work of various researchers discussed in the previous section.

In talking about informational structure of a sentence, I shall use a three-part partition

like the one proposed in Vallduvi 1990. Vallduvi's 'link' is replaced with the Centering Theory's backward-looking centre, and the ground-focus distinction is termed theme-rheme in the sense of Steedman 2000.

The *theme* is the part of the utterance which presupposes the existence of one or more alternative ways to update the discourse model, while the *rheme* is the part which actually updates the model in one of these ways. The sentence topic, i.e., the backward-looking centre is always a part of the theme.

6.1 The empirical ordering rule

In the study of Russian word-order and discourse, Olga Yokoyama suggests two ordering rules for Russian utterances. The ordering of the discourse-initial statements with 'Type 1' (neutral) intonation is

$$C_a \bigcap C_b \rightarrow Verb \rightarrow C_a \bigcap (B - C_b) \rightarrow A \bigcap (C_a - B),$$

where A and B are the knowledge sets of the speaker and the addressee, respectively, and C_a and C_b are the corresponding subsets of their current concern (p. 234).

For the non-discourse-initial utterances, "the primary dichotomy in word order is determined by the preceding utterance" (p. 326). Knowledge items found in $C_a \bigcap C_b$ after preceding utterance must be placed in the beginning in utterances with neutral intonation; the rest of the knowledge, which is found in $A \bigcap (C_a - B)$ (i.e., unknown to the addressee) must appear at the end of the utterance. Within these sections the discourse-initial rule

applies (Yokoyama 1986).

In terms of Steedman 2000, this rule would approximately translate to "theme is sentence-initial, rheme is sentence-final," with thematic verbs following, and rhematic verbs preceding other elements (if any) in the corresponding segments.

Comparing the rule with the results of the Centering study of Russian scrambled data, we see, however, that only 54% of all Cbs were sentence-initial. At the same time, if we stipulate that **rhematic status of an expression corresponds to the lower salience of its denotation** (i.e., in non-discourse-initial utterances, a rheme of U_n does not contain a $Cp(U_n)$), the rheme-final rule is confirmed by the lower salience of sentence-final elements in our data.

Thus, we modify Yokoyama's ordering rule as follows:

{ *Thematic arguments, thematic verb* } → { *rhematic verb* } → { *rhematic arguments* }

No ordering within the theme is assumed.

This rule can be illustrated by the semi-contrived examples in the Appendix.

Yokoyama's rule was formulated for the Type I intonation contour. Natalia Svetozarova (1998) provides a precise description of this intonation in the corresponding chapter of Intonation Systems survey: "the kind of sentence stress which performs this organising function can be called *neutral*... Neutral sentence stress at the end of a final declarative sentence is characterised by a simple falling tone and increased length of the stressed vowel with relatively low intensity." (p.266).

A number of researchers have argued that the element bearing the sentential stress in Type I contour is thereby marked: "A falling nuclear accent (HL*) corresponds to the natural focus. The exponent of natural focus in Russian is constituted by the last lexical accent, i.e.

at the right periphery of a sentence" (Zybatow and Mehlhorn 1999, cf. Bryzgunova 1971, Krylova and Khavronina 1988).

Here, I propose an analysis for the existing data, in which the neutral sentential stress, the centering status of an expression, and the lexical category of the transitive verb in such sentences constrain the information-structure properties of the sentence.

6.2 Syntactic Analysis.

Following Nygren 1999, I propose the following categories for the transitive Russian verb:

1. $S_i \,/\, \{NP_{inom},\, NP_{iacc}\}$

where i is a variable ranging over three values of the feature INFORMATION: θ, ρ, and η, indicating the theme, rheme, and unmarked material respectively. To iterate, the theme is defined as presupposing a *rheme alternative set* of propositions; the rheme restricts this set to a single proposition (Steedman 2000). The above categories for the verb force two different categories for its arguments: while the verb-initial word-orders are easily derived by assigning a cased NP category to the noun phrases, the other word-orders cannot. Instead, the pre-verbal arguments have to be type-raised to a rightward-looking category S/ {S/NP}.

I assume that the default NP_{nom} and NP_{acc} categories bear a neutral (η) value of the INFORMATION feature. The following categories are therefore proposed for non-rhematic noun phrases in transitive sentences (suppressing case feature marking):

2. NP_η (post-verbal categories),

or

3. $S_\theta \,/\, \{S_\theta \,/\, \{NP_\theta\}\}$

and

4. $S_\eta \,/\, \{S_\eta \,/\, \{NP_\eta\}\}$ (pre-verbal categories; empirical evidence indicates that arguments type-raised to be able to occur pre-verbally may not be rhematic).

Until proven otherwise, I shall assume that **only** belonging to the rheme licenses sentence-final position, i.e., that occurrence sentence-finally marks an item as rhematic. This presupposes that a theme-only utterance will not be able to utilise a neutral intonation contour - however, since a theme-only utterance, it seems, shall be breaking the Gricean maxims, it should naturally require a marked intonation.

Following the studies of English (Steedman 2000), and Italian (Cinque 1993), and noting the 'natural focus at the right periphery' observations of Zybatow and Mehlhorn, I hypothesise that the lexical category and value of the feature INFORMATION for the sentence-final item are dependent not on its position as such, but rather on the sentence-final neutral sentential stress in the sentences in question.

Value assignment for information:

In particular, I hypothesise that like the H* pitch accent in English (Steedman 2000), the neutral sentential stress in Russian forces the item that bears it to type-raise and assigns the ρ value to the feature INFORMATION, creating the following category:

5. $S_\rho \backslash \{S_\rho / NP_\rho\}$.

This constrains the item to being sentence-final and rhematic.

At the same time, in non-discourse-initial utterances, the backward-looking centre will always be marked as part of the theme, assuming either

6. NP_θ , or the type-raised thematic category listed in 3.

The rule for assigning a value to the INFORMATION feature on the verb is left to future research. It is clear that this may be done prosodically or through a wh-question in the

immediately preceding utterance; otherwise, I shall assume a neutral value for the verb.

The lexical items may combine with others bearing either the same, or neutral (η) value of the feature INFORMATION. At any point in the derivation the theme and rheme alternative sets may be identified and an identity function may apply mapping the informationally marked categories onto unmarked ones to allow further combination.

A few sample derivations are shown below:

(20) Ja znaju, Manya lyubit Vanyu. A kogo lyubit Anya?
I know, Mary-NOM loves John-ACC. And whom Anna-NOM loves?
'I know that Mary loves John. But whom does Anna love?'

Anya	lyubit	Petyu	
$S_\theta/\{S_\theta/NP_\theta\}$	$S_\theta/\{NP_\theta, NP_\theta\}$	$S_\rho\backslash\{S_\rho/NP_\rho\}$	
	forw.comp.		
	$S_\theta/$ NP$_\theta$	$S_\rho\backslash\{S_\rho/NP_\rho\}$	*identity function applies*
	S/NP	$S\backslash\{S/NP\}$	
		S	

(21) Chto proishodit s Anej?
What is-happening with Anna-INSTR?
'What is going on with Anna?'

Anya	lyubit	Petyu	
$S_\theta/\{S_\theta/NP_\theta\}$	$S_\rho/\{NP_\rho, NP_\rho\}$	$S_\rho\backslash\{S_\rho/NP_\rho\}$	
		S$_\rho/$ NP$_\rho$	
$S_\theta/\{S_\theta/NP_\theta\}$		S$_\rho/$ NP$_\rho$	*identity function applies*
S/S/NP		S/NP	
		S	

Note that this last sentence will still compute if the verb category

is $S_\eta/\{NP_\eta , NP_\eta\}$.

The incompatibility of the feature θ with the sentence-final position blocks some derivations in infelicitous contexts. For example, in the following answer to the question in (11), the subject Anya is marked as a theme by the preceding wh-question, but at the same time is occupying the rheme-only sentence-final position.

(22) Whom does Anya love?

#	Lyubit	Petyu	Anya
	$S_\theta/\{NP_\theta, NP_\theta\}$	NP_θ	?!?

When more than one item is left unmarked as thematic by the context, feature mismatch cannot rule out all of the infelicitous answers, as in the following answer to the question in (12). In such cases, the neutral sentential stress falls on the rhematic verb, and can, in principle, mark the entire sentence as a single 'informational phrase.' However, the derivation of the sentence fails, since the type-raised category for rhematic noun phrases disallows them to appear pre-verbally.

(23) What is going on with Anna?

#	Anya	Petyu	lyubit.
	$S_\theta/\{S_\theta/NP_\theta\}$	$S_\rho\backslash\{S_\rho/NP_\rho\}$	$S_\rho/\{NP_\rho, NP_\rho\}$
	?		

6.3 Summary of this section

Here, we have proposed on the basis of previous literature and the empirical evidence from the Centering study of Russian narrative, a simple discourse-status motivated ordering rule for Russian. This descriptive rule was then formally modelled by assuming a Set-CCG syntax for Russian transitive sentences as proposed in Nygren 1999, and postulating a feature INFORMATION, which can mark the transitive verb and post-verbal argument categories in Russian as either neutral, rhematic or thematic, and the pre-verbal argument category as neutral or thematic. The mechanism of assigning the value for the INFORMATION feature is twofold: firstly, the neutral sentential stress marks its bearer as rhematic (while also type-raising the arguments that bear it to become the rightmost ones). Secondly, the Centering Algorithm marks the backward-looking centre of an utterance as belonging to the theme. The postulated underlying verb-initial word-order in Russian forces preverbal arguments to have type-raised rightward-looking categories, which are then easily restricted to be non-rhematic. Under this analysis. the peculiar Centering properties of subject-final word orders are derived naturally from their assumed intonational properties.

7 Conclusions

The Centering study of a written corpus suggests that word order and attentional structure of discourse are interdependent phenomena in Russian. The entity-based approach to local discourse coherence shows a special informational status of the subject-final word orders in Russian. The study shows that in such word orders the subjects are dispreferred as potential topics for subsequent discourse.

As previous studies suggest, this artefact may be due to the informational marking induced by the sentence-final sentential stress in neutral Russian intonation. A Set-Combinatory Categorial Grammar for a fragment of Russian is then utilised to model the informational structure of simple transitive sentences. In the model, the intonational stress is assumed to mark its bearer as rhematic (in the sense of Steedman 2000), in addition to the contextual marking of sentential centres (topics) as thematic. The informational structure of Russian transitive sentences is then straightforwardly derived by the model.

In addition, an investigation of the referential properties of zero subjects in Russian suggests that these might pose a challenge to the entity-tracking algorithm, unless the factors that determine which zeroes refer to entities are made clear.

References

Akmajian, A. (1979). *Aspects of the grammar of Focus in English.* New York: Garland.

Avrutin,S. and Rohrbacher, B. (1995). Null Subjects in Russian Inverted Constructions. Browne, W., Dornisch, E., Kondrashova, N., and Zec, D. (eds.). *The Fourth Annual Workshop on Formal Approaches to Slavic Languages. The Cornell Meeting.* Ann Arbor: Michigan Slavic Publications.

Bailyn, J.F. (1995). *A configurational approach to Russian 'free' word order.* Cornell University Ph.D. Dissertation.

Bailyn, J.F. (2000). Generative Syntax (GB/Minimalism). (in press). In Fowler, G. ed. *Slavic Linguistics in the 21st Century.* Bloomington, IN: Slavica. (conference position paper available at http://www.indiana.edu/ slavconf/SLING2K/pospapers/bailyn.pdf).

Bailyn, J.F. (in press). Does Russian Scrambling Exist? In S. Kirimi, ed., Word Order and Scrambling, Blackwell. (conference presentation and handout).

Bošković, Ž. and D.Takahashi (1998). Scrambling and Last Resort. Linguistic Inquiry 29, p.347.

Brennan, S., Friedman, M. and Pollard, C. (1987). A Centering Approach to Pronouns. *Proc. 25th Annual Meeting of the ACL*, Stanford, CA.

Bryzgunova, E. A. (1971). O smyslorazlichitel'nyx vozzmozhnostjax russkoj intonacii [On meaning-distinguishing possibilities of Russian intonation]. *Voprocy jazykoznanija* [Issues in linguistics]. 4, p.42.

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. In Steinberg, D. and Jacobovits, L. (eds.) *Semantics.* Cambridge: Cambridge University Press.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Cinque, G. (1993). A Null Theory of Phrase and Compound Stress. Linguistic Inquiry 24, p.239.

Dahl, O. (1974). Topic-Comment structure revisited. In Dahl, O. ed. *Topic and comment, contextual boundedness and focus.* Hamburg: Helmut Buske.

Di Eugenio, B. (1998). Centering in Italian. In M. Walker, A. Joshi, and E. Prince (1998).

Firbas, J. (1964). On defining the theme in functional sentence perspective. *Travaux linguistiques de Prague* 1, p.267.

Franks, S. (1995). *Parameters of Slavic Morphosyntax.* New York: Oxford University Press.

Givón, T. (1983). Introduction. In T. Givón (ed.). *Topic continuity in discourse.* Amsterdam: John Benjamins.

Grosz, B., Joshi, A. and Weinstein, S. (1995). Centering: a framework for modelling local coherence of discourse. http://www.cis.upenn.edu/ ircs.

Grosz, B. and Sidner, C. (1986). Attentions, Intentions and the Structure of Discourse. *Computational Linguistics* 12, p.203.

Gundel, J. (1974). *The role of topic and comment in linguistic theory.* University of Texas Ph.D. dissertation.

Gundel, J. (1988). Universals of topic-comment structure. In Hammond, M., Moravcsik, E. and Wirth, J. (eds.). *Studies in syntactic typology.* Amsterdam: John Benjamins.

Halliday, M.A.K. (1967). Notes on transitivity and theme in English. Part II. *Journal of Linguistics* 3, p.199.

Halliday, M.A.K. (1985). *Introduction to Functional Grammar.* London: Edward Arnold.

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In R.

Bauerle, C. Schwarze, and A. von Stechow (eds.). *Meaning, use and the interpretation of language.* Berlin: Walter de Gruyter.

Hocket, C.F. (1958). *A Course in Modern Linguistics.* New York: Macmillan.

Hoffman, B. (1998). Word Order, Information Structure, and Centering in Turkish. In M. Walker, A. Joshi, and E. Prince (1998).

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar* Cambridge, MA: MIT Press.

Joshi, A. and Kuhn, S. (1979). Centered Logic: The Role of Entity Centered Sentence ReEpresentation in Natural Language Inferencing. *6th International Joint Conference on Artificial Intelligence*, p.435.

Kameyama, M. (1998). Intrasentential Centering: A case study. In M. Walker, A. Joshi, and E. Prince (1998).

King, T. (1993). Configuring Topic and Focus in Russian. Ph.D. Dissertation, Stanford University.

Kiss, K. (2000). Movement to the left periphery. Ms. Linguistic Institute of the Hungarian Academy, *International Conference on Word Order and Scrambling*, Tuscon, AZ.

Kondrashova, N. (1997). Generativnaja grammatika i problema svobodnogo poryadka slov [Generative Grammar and the Problem of Free Word Order]. In A. Kibrik, I. Kobozeva, and Sekerina I. (eds.). *Fundamental'nye napravleniya v sovremennoj amerikanskoj lingvistike. [Fundamental Trends of Modern American Linguistics.]* Moscow: MSU Press.

Kruijff-Korbayová, I., Kruijff, G.J.M. and Bateman, J. (2000). Generation of Contextually Appropriate Word Order. To appear in *Information sharing.* Kees van Deemter and Rodger Kibble (eds.). CSLI Publishers. (AGILE project) (currently available at

http://www.cogsci.ed.ac.uk/ ivana/Publications/wo-booksubm.ps).

Krylova, O. and Khavronina, S. (1988). *Word order in Russian.* Moscow: Russkij Yazyk.

Kuno, S. (1980). The scope of the question and negation in some verb-final languages. In Kreiman, J., and Ojeda, A. E. (eds.). *Papers from the Sixteenth Regional Meeting, Chicago Linguistic Society*, p. 155.

Lambrecht, K. (1987). Sentence focus, information structure, and the thetic-categorial distinction. In *Proceedings of the 13th Annual Meeting of the Berkley Linguistics Society*, p.366.

Lambrecht, K. (1994). *Information Structure and Sentence Form.* Cambridge: Cambridge University Press.

Mahajan, A. (1990). *The A/A-bar Distinction and Movement Theory.* MIT Ph.D. dissertation.

Malamud, S.A. (2000). Centering and Scrambling: Towards a Discourse-Status Motivation of Russian Word Order. In Franks, S., King, T.H., and Yadroff, M. (eds.). *The Ninth Annual Workshop on Formal Approaches to Slavic Languages. The Indiana Meeting.* Ann Arbor: Michigan Slavic Publications.

Mathesius, V. (1915). O passivu v moderní angličtině. Sborník filologický 5, p.198.

Mathesius, V. (1939). On the information-bearing structure of the sentence. **Kuno, S.(ed.)**.*Harvard Studies in Syntax and Semantics.* Harvard University Press.

Miltsakaki, E. (1999). Dissociating Discourse Salience from Information Structure: Evidence from a Centering study in Modern Greek and Japanese. *Computational Linguistics in the Netherlands, (CLIN '99).*

Miltsakaki, E. (2001). *Attention Structure in Discourse: A Cross-Linguistic Investigation.* University of Pennsylvania Ph.D. dissertation proposal.

Miltsakaki, E. and Kukich, K. (2000). Automated Evaluation of Coherence in Student Essays. To appear in *Proceedings of ACL 2000.* http://www.ling.upenn.edu/ elenimi/grad.html

Miyagawa, S. (1997). Against optional scrambling. *Linguistic Inquiry* 28, p.1.

Moshkow, M. On-line Russian Library. http://www.lib.ru.

Nygren, N. (1999). Coordination and Word Order in Russian. Ms. University of Edinburgh, Edinburgh, UK.

Prince, E.F. (1981). Toward a Taxonomy of Given/New Information. In P. Cole (ed.). *Radical Pragmatics.* New York: Academic Press.

Prince, E.F. (1984). Topicalization and Left-dislocation: A functional analysis. White, S.J. and Teller, V. (eds.). *Discourses in reading and linguistics. Annals of the New York Academy of Sciences.* vol.433, p.213.

Prince, E.F. (1986). On the syntactic marking of presupposed open propositions. *Parasession Papers from the Twenty Second Annual Meeting of Chicago Linguistics Society.*, p.208.

Prince, E.F. (1988). Discourse analysis: A part of the study of linguistic competence. Newmeyer, F. (ed.) *Linguistics: The Cambridge Survey* vol.2. Cambridge: Cambridge University Press.

Radford, A. (1997). *Syntactic Theory and The Structure of English - A Minimalist Approach.* New York: Cambridge University Press.

Rambow, O. (1993). Pragmatic Aspects of Scrambling and Topicalization in German: A Centering Approach. Unpublished Manuscript, University of Pennsylvania.

Reinhart, T. (1982). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27, p.53.

Sekerina, I.A. (1997). *The Syntax and Processing of Scrambling Constructions in Russian.* City University of New York Ph.D. dissertation.

Sgall, P., Hajičová, E., and Benešová, E. (1973). *Topic, Focus and Generative Semantics.* Kronberg: Taunus.

Stechow, A. von. (1981). Topic, Focus and local relevance. Klein, W. and Levelt, W. (eds). *Crossing the boundaries in linguistics.* Dordrecht: Reidel.

Steedman, M. (1991). Structure and Intonation. *Language* 67, p.262.

Steedman, M. (1996). *Surface Structure and Interpretation.* Cambridge, MA: MIT Press.

Steedman, M. (2000). *The Syntactic Process.* Cambridge, MA: MIT Press.

Strawson, P.F. (1964). Identifying reference and truth-values. Steinberg, D. and Jacobovits, L. (eds.) *Semantics.* Cambridge: Cambridge University Press.

Svetozarova, N. (1998). Intonation in Russian. Hirst, D. and Di Cristo, A. (eds.). *Intonation Systems. A Survey of Twenty Languages.* Cambridge: Cambridge University Press.

Vallduvi, E. (1990). *The Information Component.* University of Pennsylvania Ph.D. dissertaion.

Walker,M. and Prince, E. (1996). A Bilateral Approach to Givenness: a Hearer-Status Algorithm And a Centering Algorithm. In T. Fretheim, and J. Gundel (eds.). *Reference and referent accessibility.* Philadelphia: John Benjamins.

Walker, M., Iida, M. and Cote, S. (1994). Japanese Discourse And the Process of Centering. *Computational Linguistics 21.*

Walker, M., Joshi, A. and Prince, E. eds. (1998). *Centering Theory in Discourse.*

Oxford: Oxford University Press.

Ward, G. (1985). *The semantics and pragmatics of preposing.* University of Pennsylvania Ph.D. dissertation.

Wilson, D. and Sperber, D. (1979). Ordered entailments: An alternative to presuppositional theories. Oh, C.K. and Dinneen, D. (eds.). *Presupposition (Syntax and Semantics, 11).* New York: Academic Press.

Webber, B. (1991). Structure And Ostension in the Interpretation of Discourse Deixis. http://www.ling.upenn.edu/~ellen/bonnie.ps .

Yokoyama, O. T. (1986). *Discourse and Word Order.* Philadelphia: John Benjamins.

Zybatow, G. and Mehlhorn., G. (1999). Experimental Evidence for Focus Structure in Russian. King, T.H. and Sekerina, I.A. (eds.). *The Eighth Annual Workshop on Formal Approaches to Slavic Languages. The Philadelphia Meeting.* Ann Arbor: Michigan Slavic Publications.

Appendix

A. Comparing left-to-right and grammatical-function rankings

Consider the following segment from Bulgakov, in which the second sentence is scrambled (The Cb of the previous utterance is K.):

(24) K. svistnul.
K. let-out-a-whistle.
'K. let out a whistle.'

Cf = {K.}

Cp = K.

Cb = K.

Etogo svista Margarita ne uslyhala, no ona ego uvidela v to vremya, kak
Of-this whistle Margarita not heard, but she it saw at that time, as
ee vmeste s goryachim konem brosilo sazhenej na desyat' v storonu.
her together with hot horse it-threw sazhens for ten to side.

'Margarita didn't hear this whistle, but she saw it at the same time when she, together with her hot-tempered horse, was thrown several meters to the side.'

Analysis 1: ranking by grammatical function.

Cf={Margarita, whistle, horse}

Cp=Margarita

Cb = whistle

Transition = Rough-Shift

Analysis 2: ranking by word order.

Cf = {whistle, Margarita, horse}

Cp = whistle

Cb = whistle

Transition = Smooth-Shift

B. Centering of a fragment containing several zero subjects

(25) Dumali, chto on dejstvitel'no vernetsya cherez minutku.
Thought(3rdP.)(PL.), that he really will-come-back after minute.
'[They] thought that he will really come back in a minute.'

Cb= Fagot("he") (determined by Pronoun rule)

Cf= Fagot, at least some of the employees denoted by the zero

Transition=Continue

(26) No proshlo i desyat' minut, a ego netu.
But passed and ten minutes, and of-him there-was-not.
'But ten minutes passed, and he wasn't there.'

Cb=Fagot(Default, no competitor)

Transition=Smooth-Shift or Continue

(27) Radost' ohvatila filial'cev – sbezhal.
Joy overcame employees – ran-away(3rdP.)(S.).
'Joy overcame the employees - he ran away.'

Cb=Fagot(Pronoun-rule)

Cf=employees, joy, Fagot

Transition=Retain

(28) I vdrug kak-to sami soboj zapeli
And suddenly some-how themselves by-themselves started-singing(3rdP.)(PL.)
vtoroj kuplet,
second stanza,
'And suddenly somehow on their own they started singing the second stanza (of the song),'

Cb=employees

Cf=employees, second stanza, the song

Transition=Smooth-Shift

(29) vseh povel za soboj Kosarchuk, u kotorogo, mozhet byt', i ne
everybody(ACC) led after himself K., at whom, may be, and not
bylo absolyutnogo sluha, no byl dovol'no priyatnyj vysokij tenor.
there-was absolute hearing, but there-was quite pleasant high tenor.
'everybody followed K., who perhaps did not have a perfect pitch, but who had a quite pleasant high tenor.'

Cb=employees

Cf=employees, K., his tenor

Transition=Continue

(30) Speli.
Finished-singing(3rdP.)(PL.).
'They finished singing [the song].'

Cb=employees

Cf=employees

Transition=Continue

(31) Regenta netu!
Of-regent there-isn't!
'No regent!' [regent here = choirmaster]

Cb=none? Or Fagot - pop to U12 at least

Cf=employees

Transition=Rough-Shift or Continue

(32) Dvinulis' po svoim mestam, no ne uspeli
Started-moving(3rdP.)(PL.) to their places, but not had-time(3rdP.)(PL.)
sest', kak, protiv svoego zhelaniya, zapeli.
to-seat, when, against their-own will, started-singing(3rdP.)(PL.).

'They started going to their places, but before they could seat they started singing, against their will.'

Cb=employees(Default, pop to U15)

Cf=employees, places

Transition=Continue

(33) Ostanovit', – no ne tut-to bylo.
To-stop, – but not here-FOC it-was.
'Stop this! But they couldn't' ("ne tut-to bylo" - "not here it was" is an idiom, meaning that the action just mentioned is beyond the abilities of those trying)

Cb=none or inferrable employees

Cf=singing, inferrable employees

Transition=Rough-Shift or Retain

(34) Pomolchat minuty tri i opyat'
Will-keep-silent(3rdP.)(PL.) for-minutes for-three and again
gryanut.
will-thunder(3rdP.)(PL.).
'They would keep silent for about three minutes, and then would start to thunder again.'

Cb=employees

Cf=employees

Transition=Smooth-Shift or Continue

(35) Pomolchat – gryanut!
Will-keep-silent(3rdP.)(PL.) – will-thunder(3rdP.)(PL.)!
'[They'd keep] silent for a while - and start to thunder!'

Cb=employees

Cf=employees

Transition=Continue

(36) Tut soobrazili, chto beda.
Here realised(3rdP.)(PL.), that wrongness.
'Here they realised that something's wrong.'

Cb=employees

Cf=employees Transition=Continue

(37) Zaveduyushchij zapersya u sebya v kabinete ot sramu.
Boss locked-himself at himself in office from shame.
'The boss locked himself in his office out of shame.'

Cf=boss, his office, possibly inferrable employees?

Cb=none, or employees

Transition=Rough-Shift or Retain

C. Illustration of the empirical Ordering Rule (Section 6)

A wh-question was used to unambiguously indicate the felicitous set of alternatives for the answer. A neutral intonation for the answers was assumed.

The felicity judgements below are my own and those of two native informants - therefore, this is certainly no more than an example illuminating the above rule.

In reading the answers to the question in the example, it must be born in mind that '-a' ending marks nominative case, '-u' ending marks accusative, and 'lyubit' is a third-person singular present form of the verb 'to love.' The felicity judgements are marked to the left of each answer. The item bearing the neutral sentential stress is sentence-final.

(38) Ja znaju, Manya lyubit Vanyu. A kogo lyubit Anya?
I know, Mary-NOM loves John-ACC. And whom Anna-NOM loves?
'I know that Mary loves John. But whom does Anna love?'

ok a. Anya lyubit Petyu.

b. Petyu lyubit Anya.

#? c. Anya Petyu lyubit.

#(?) d. Petyu Anya lyubit.

ok? e. Lyubit Anya Petyu.

#? f. Lyubit Petyu Anya.

(39) Ja znaju, Manya lyubit Vanyu. A kto lyubit Petyu?
I know, Mary-NOM loves John-ACC. And who loves Peter-ACC?
'I know that Mary loves John. But who loves Peter?'

a. Anya lyubit Petyu.

ok b. Petyu lyubit Anya.

c. Anya Petyu lyubit.

d. Petyu Anya lyubit.

?#? e. Lyubit Anya Petyu.

ok f. Lyubit Petyu Anya.

(40) Ja slyshala, Anya i Petya nenavidyat drug druga. Eto pravda?
I heard, Anna-NOM and Peter-NOM hate each other. This truth?
'I heard that Anna and Peter hate each other. Is this true?'

a. Nyet, Anya lyubit Petyu.

b. Nyet, Petyu lyubit Anya.

ok c. Nyet, Anya Petyu lyubit.

ok? d. Nyet, Petyu Anya lyubit.

e. Nyet, lyubit Anya Petyu.

f. Nyet, lyubit Petyu Anya.

The judgements seem to indicate that in this intonation, the most felicitous word orders will indeed have the rheme as the sentence-final element, and the theme as the sentence-initial. However, no definite prohibition on post-verbal placement of thematic items could be detected, as in e.g., the (e) answer to the first question. A pre-verbal rheme, however, seems infelicitous, even when the verb is rhematic as in the answers to the following question:

(41) Chto proishodit s Anej?
What is-happening with Anna-INSTR?
'What is going on with Anna?'

ok a. Anya lyubit Petyu.

#? b. Petyu lyubit Anya.

c. Anya Petyu lyubit.

d. Petyu Anya lyubit.

?ok? e. Lyubit Anya Petyu.

f. Lyubit Petyu Anya.