

The function of word order in Russian compared with Danish and English

Ph.D. thesis by Johannes Kizach, Department of English, SLK, University of Aarhus

1. Russian word order	1
1.1 What is the problem?	1
1.2 The traditional approach	2
1.3 An alternative approach	3
2. The three questions	5
3. Processing efficiency and Russian word order	6
4. Information structure and Russian word order	10
5. Processing and syntax	12
6. Conclusion	13
<i>References</i>	14

1. Russian word order

1.1 What is the problem?

The problem is the freedom. In Russian, the word order is much less restricted than in e.g. Danish or English. Consider (1):

- | | |
|-----------------------------|-----|
| (1) a. Vladimir ubil sobaku | SVO |
| <i>Vladimir killed dog</i> | |
| "Vladimir killed the dog" | |
| b. Vladimir sobaku ubil | SOV |
| c. Sobaku Vladimir ubil | OSV |
| d. Sobaku ubil Vladimir | OVS |
| e. Ubil Vladimir sobaku | VSO |
| f. Ubil sobaku Vladimir | VOS |

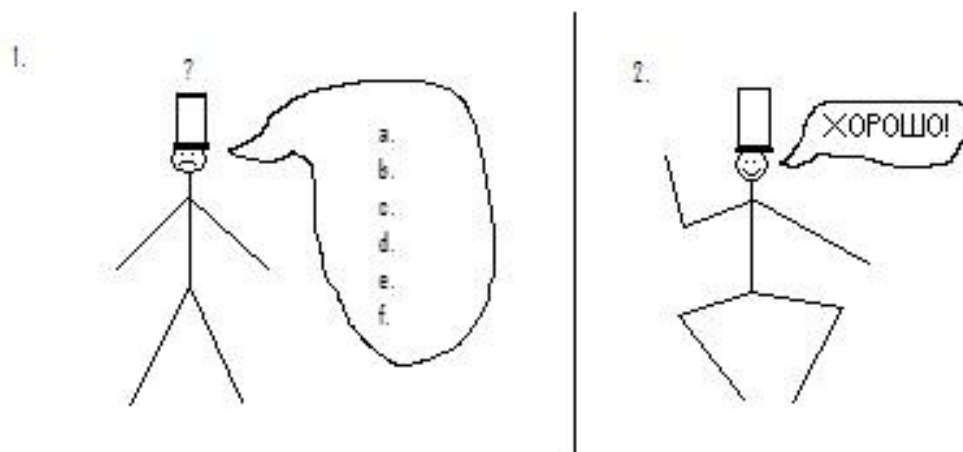
All the sentences in (1) are declarative and express the same basic predication, namely that Vladimir killed the dog, despite the differences in word order. This is obviously not so in e.g. Danish, where the alteration of the word order leads to meaning differences ((2)d), clause type differences ((2)e), both ((2)f), or gibberish ((2)b and c):

- | | |
|--------------------------------|-------|
| (2) a. Valdemar dræbte hunden | SVO |
| <i>Valdemar killed dog.the</i> | |
| "Valdemar killed the dog" | |
| *b. Valdemar hunden dræbte | (SOV) |
| *c. Hunden Valdemar dræbte | (OSV) |
| d. Hunden dræbte Valdemar | (OVS) |
| "The dog killed Valdemar" | |
| e. Dræbte Valdemar hunden? | VSO |
| "Did Valdemar kill the dog?" | |
| f. Dræbte hunden Valdemar? | (VOS) |
| "Did the dog kill Valdemar?" | |

The point is that when Russians speak, they are faced with more options than Danish and English speakers, and the question is why they choose one order instead of the other available orders? Do they choose randomly? Do they have idiosyncratic sympathies for specific orders? Are there subtle meaning differences between the orders?

In short, why do Russians behave like the man in (3)2 and not like the man in (3)1?

(3) Russian stick man:



1.2 The traditional approach

The usual explanation within Russian linguistics (see e.g. Kovtunova 1976, Sgall, Hajičová and Panevová 1986, Firbas 1992, King 1995, Slioussar 2007) is that the information structure of the

message determines the word order of the sentence. The basic idea is that given material is placed at the beginning of the sentence, and new material at the end.

This is typically illustrated with very simple three word sentences like the following, taken from Slioussar (2007:1-2, (1.1)):

- (4) a. Programmist kupil kofevarku. SVO
programmer.NOM bought [coffee machine.ACC]
“The programmer bought a coffee machine”
- b. Kofevarku kupil programmist OVS
[coffee machine.ACC] bought programmer.NOM
“A programmer bought the coffee machine”

These sentences mean the same thing, but whereas (4)a is used in a context where *the programmer* is given, (4)b is used in contexts where *the coffee machine* is given (Slioussar 2007:1-2).

It is interesting that nothing is said on why the SOV-order is not as good as the suggested SVO-order in (4)a, nor on why the OSV-order is not as good as the suggested OVS-order in (4)b. In fact, this is usually not addressed in the literature on information structure and Russian word order (see references above). The example involves a single word subject and object which is typical in the literature on information structure in Russian, and which is hardly a coincidence as we will see later.

1.3 An alternative approach

The idea that processing efficiency influence word order choice is suggested in Hawkins (1994, 2004), where a correlation between efficiency and frequency is demonstrated in data from several languages.

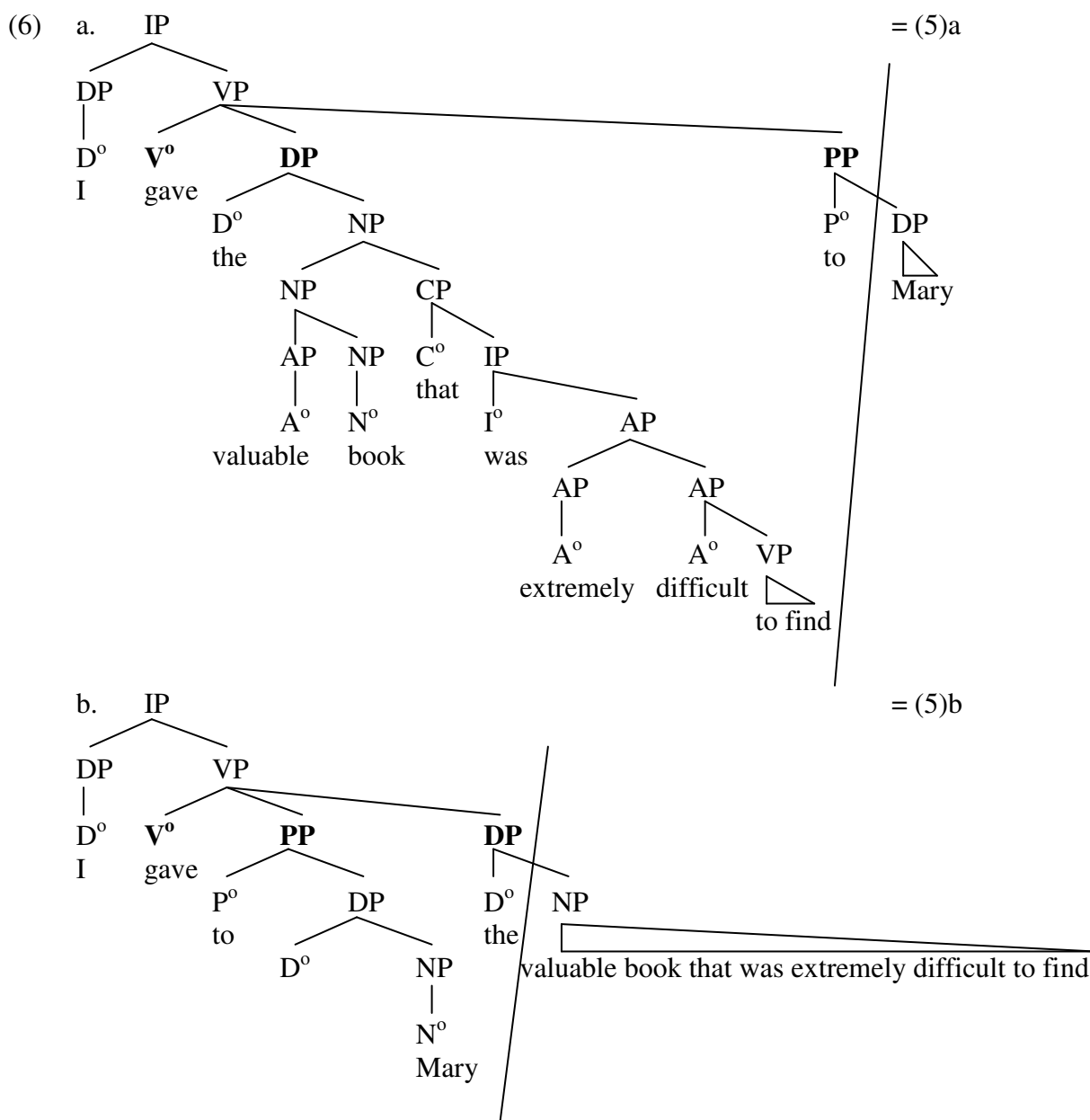
Consider (5):

- (5) a. I gave [the valuable book that was extremely difficult to find] [to Mary]
b. I gave [to Mary] [the valuable book that was extremely difficult to find]

(Hawkins 1994:57, (3.1a) and (3.1b))

Let us imagine that when we encounter the verb *give*, then we start looking for two objects (due to the subcategorization frame of *give*). This means that we will be on the lookout until we have found both objects. Clearly (5)a is much clumsier than (5)b, and this is possibly due to the fact that in (5)a we have to process the entire nominal object, before we reach the prepositional object, whereas in (5)b, we find both our objects within the first five words of the sentence.

The point is that we have to process lots of material/structure before we have the basic double object structure in place in (5)a, whereas we get the basic structure very rapidly in (5)b.



(Adapted from Hawkins 1994:59, 60 (3.1'a), (3.1'b))

The difference between these two examples could be expressed as a difference in the ratio of constituents compared to phrasal nodes. So in (5)a we have two constituents in the clause, the subject (DP) and the predicate (VP), and inside the predicate we have three constituents, the verb and the two objects. We know all this when we reach the word *to*. The amount of phrasal nodes that we have to process is 15. The efficiency ratio is thus $5/15 = 33\%$.

In (5)b we still have 5 constituents, but now we only have to process 7 phrasal nodes to establish this. The efficiency ratio is thus $5/7 = 71\%$.

Now we have a method to calculate the relative efficiency of two (or more) sentences which contain the same words, but in different orders.¹

Notice that this is not an absolute measure – it can only be used to decide between different orderings of the same elements.

2. The three questions

Now we have established that Russian poses a puzzle with regards to how speakers choose between the alternative word orders, we have seen the traditional approach, and we have seen that an alternative approach exists.

This brings us to the three main topics of the thesis:

- I. Is there a correlation between frequency and efficiency in Russian (and Danish) performance data?
- II. What exactly is claimed about word order by information structure theories, and is there any evidence that they are right?
- III. Can processing facts be of any benefit to syntax, especially as a method to choose between alternative analyses?

In the following we will look at the first question in detail, and then briefly comment on the second and third question.

¹ This method is simply meant to illustrate the idea of a complexity metric (as developed in Hawkins 1994), but the method used in the thesis is somewhat more complex and crucially relies on the concept *parsing domain*, which is not explained here. I refer to the explanation on pages 50-55 in the thesis.

3. Processing efficiency and Russian word order

The aim is to test whether Russian speakers prefer the most efficient orders in cases where they have a choice. To test this we need to find structures where there is in fact a choice, then find examples in a corpus, analyze all examples, so that for each we know which order is most efficient, and then finally compare the actual order with the most efficient order.

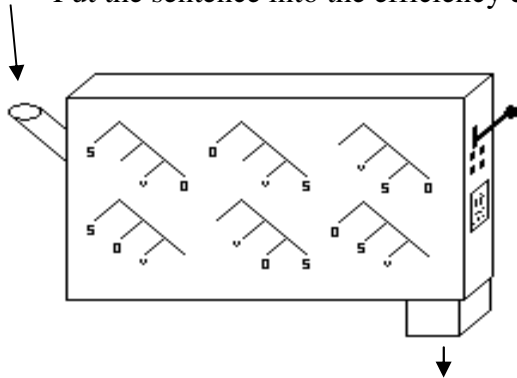
Prediction: the most efficient orders = the most frequent orders

1. Gather examples of transitive sentences from a corpus

2. Take the first of the transitive sentences

The sentence might have the order SOV

Put the sentence into the efficiency calculator:



The output is the efficiency of the sentence with the actual order: SOV
and with the alternative orders: SVO, OSV, OVS,
VSO and VOS

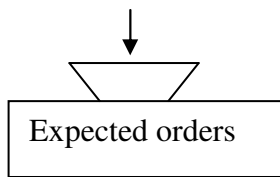
The efficiency calculator tells us which of the orders is most efficient, which is the second most efficient etc.

3. Compare the actual order with the most efficient order

If the actual order that the sentence has in the data and the most efficient order is the same, put the sentence in the box labelled "expected orders", if the actual order and the most efficient order is not the same, then put the sentence in the box labelled "unexpected orders":

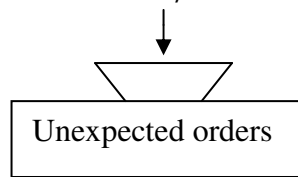
Most efficient order = Actual order

SOV = SOV



Most efficient order ≠ Actual order

SVO ≠ SOV



This procedure is repeated for all the transitive sentences in the data until all sentences have been placed in one of the two boxes – we then get a percentage demonstrating whether there is a correlation between efficiency and frequency or not.

I have examined the following structures in Russian, Danish and English:

Transitive sentences - Russian

- (7) a. Vladimir ubil sobaku
Vladimir killed dog
"Vladimir killed the dog"
- b. Vladimir sobaku ubil
c. Sobaku Vladimir ubil
d. Sobaku ubil Vladimir
e. Ubil Vladimir sobaku
f. Ubil sobaku Vladimir

Adversity impersonals - Russian

- (8) a. Ščuku vskolyxnulo vodovorotom
pike.ACC stirred up.3.P.NEUTER whirlpool.INST
"The pike was stirred up by the whirlpool"
- b. Ščuku vodovorotom vskolyxnulo
c. Vodovorotom ščuku vskolyxnulo
d. Vodovorotom vskolyxnulo ščuku
e. Vskolyxnulo ščuku vodovorotom
f. Vskolyxnulo vodovorotom ščuku

Double objects construction - Russian

- (9) a. Ivan dal [Masju] [sobaku]
Ivan gave Masja.DAT dog.ACC
"Ivan gave Masja a dog"
- b. Ivan dal [sobaku] [Masju]

Postverbal prepositional phrases - Russian

- (10) a. Ja dolgo privykala smotret' [čerez linzy] [na okružajuščij mir]
I long got.used.to look through lenses at surrounding world
"For a long time I got used to looking through lenses at the world around me"

- b. Ja dolgo privykala smotret' [na okružajuščij mir] [čerez linzy]
I long got.used.to look at surrounding world through lenses
 "For a long time I got used to looking at the world around me through lenses"

Postverbal prepositional phrases – Danish

- (11) a. Der er ikke langt [fra det brede fortov] [til rendestenen]
there is not far from the broad sidewalk to gutter.the
 “There is not far from the broad sidewalk to the gutter”
- b. Der er ikke langt [til rendestenen] [fra det brede fortov]
there is not far to gutter.the from the broad sidewalk
 “There is not far to the gutter from the broad sidewalk”

Particle construction – English

- (12) a. John took [the garbage] [out]
 b. John took [out] [the garbage]

In all cases there is a clear correlation between efficiency and frequency (i.e. the most frequent orders are the most efficient orders).²

Table 1: The results

		successes	relevant cases	success ratio
Russian	Transitive sentences - 1 order predicted	122	155	79%
	Transitive sentences - 2 orders predicted	150	162	93%
	Adversity impersonals - 1 order predicted	85	127	67%
	Adversity impersonals - 2 orders predicted	35	50	70%
	Double object construction	129	147	88%
	Postverbal PPs	193	218	89%
Danish	Postverbal PPs	78	111	70%
English	Particle construction	237	247	96%

² The correlation is statistically significant in all the cases (exact binomial test used, calculated using R software, R Development Core Team 2009).

These tests clearly demonstrate that efficiency influences word order choice in Russian, and this gives further reason to reconsider the importance of information structure with regards to Russian word order.

4. Information structure and Russian word order

As mentioned above, the traditional view is that word order variation in Russian can be explained by referring to the information structure of the sentence, but as it turns out, information structure theories only consider simple declarative sentences, and say nothing about embedded clauses, interrogatives or imperatives:

(13) Clauses where information structure influence ordering:

Simple declarative sentences ~~embedded clauses~~ ~~interrogatives~~ ~~imperatives~~

Furthermore it turns out that within the simple declarative sentences, the theories only concern referring expressions³, so verbs, most adverbials and clausal arguments are not necessarily ordered according to information structure:

(14) Elements inside declarative sentences ordered according to information structure status:

Referring expressions ~~verbs~~ ~~adverbials~~ ~~clausal arguments~~ ~~negation~~

The information structure theories divide declarative sentences into two types: Emotive and non-emotive sentences. In so called emotive sentences, the referring expressions are ordered in the opposite way of how we would expect. So in these sentences new precedes given elements.

³ The concepts givenness, salience, contextually boundedness and topichood all rely on previous mention of the referent of the phrase, and clearly this is irrelevant for non-referring expressions.

(15) Referring expressions ordered according to their information structure status:

Referring expressions in non-emotive simple declarative sentences ~~referring expressions in emotive simple declarative sentences~~

The idea that speakers order their phrases according to information structure relies on the fact that e.g. all six logically possible orders in transitive sentences are available. And we should expect that speakers take advantage of this option irrespective of how heavy the subject and object are. But this is in fact not the case at all. On the contrary, Russian speakers only use the many word order options with transitive sentences when the subject and object are very light – which is expected from a processing point of view, since it is in precisely these cases that processing cost of the less efficient orders is smallest (a very short sentence is easy to process).

Table 2: Distribution of SVO and non-SVO orders

	All	One DP 3+ words	Both DPs 2+ words	Both pronominal
SVO	205	66	10	21
non-SVO	112	7	1	51
Ratio of non-SVO	35%	10%	9%	71%
Totals	317	73	11	72

So to recapitulate: Information structure theories might have something to say about the ordering of referential expressions in non-emotive simple declarative clauses with pronominal arguments.

This is a little less impressive than how the traditional view is normally presented, which is that all ordering facts follow from information structure.

A few other problems with the traditional approach: Given elements/contextually bound elements/salient elements are possibly very rare, making the theory inapplicable, because how can we order elements according to their status as contextually bound or not if only a very small percentage of the elements are contextually bound? In a text count Firbas (1992:30) finds that out of 1.167 words only 198 are context dependent, so a full 83% of the text is context independent.

Some claim (see e.g. Siewierska & Uhlířová 1998) that the order given-new-given is the most common one, others claim that the sentence is always organized so that the further right we go, the less given the elements are (see e.g. Slioussar 2007).

Tests suggest that information structure influence word order in English (Gries 2003, Arnold et al. 2000) and sometimes suggest that information structure does not (Hawkins 1994). The one test that I have found where Russian is examined show that information structure only marginally influences word order (Holden & Krupp 1987).

A pilot study presented in the thesis demonstrates that in specific context more than one order is available, so information structure may at best narrow down the options, but can hardly be said to determine the word order choice.

5. Processing and syntax

Different syntactic analyses of a construction can change the efficiency calculation, so two analyses will typically lead to slightly different results – one analysis might lead to the result that the most efficient orders are chosen in 50% of the cases, and an alternative analysis may lead to the result that the most efficient orders constitute 85% of the orders. There could also be more subtle differences in the predictions depending on which analysis the calculation is based on.

If it is a fact about language that the most efficient orders are preferred in cases where there is a word order choice, then this fact could potentially be used to choose between the alternative analyses: The one that is compatible with a correlation between frequency and efficiency is preferred over the one that is not compatible with a correlation between frequency and efficiency.

The test of the transitive sentence data showed a strong correlation between frequency and efficiency. In the test we used Bailyn's (2004) analysis of the six different word orders, but now let us use the analysis in King (1995) as the basis for the efficiency calculations and compare the results.

Table 3: Results – transitive sentences, comparison of Bailyn’s and King’s results

		successes	relevant cases	success ratio
Calculated using Bailyn's structures	1 order predicted	122	155	79%
	2 orders predicted	150	162	93%
Calculated using King's structures	1 order predicted	7	317	2%

When the structures suggested in King (1995) are used as the basis for the efficiency calculation, PTOC makes a single prediction in all 317 cases. Of the 317 cases, only 7 (2%) have the order predicted by PTOC.

PTOC is unsuccessful when the calculations are made on the basis of King (1995).

When the calculations were based on the structures suggested in Bailyn (2004), we saw a strong correlation between the most efficient orders and the most frequent orders, with 79% successes for the 155 cases where there was a single prediction, and 93% successes in the 162 cases where two orders were tied for the first place.

6. Conclusion

Let us look at the three questions again:

- I. Is there a correlation between frequency and efficiency in Russian (and Danish) performance data? **Yes there is – a strong one.**
- II. What exactly is claimed about word order by information structure theories, and is there any evidence that they are right? **The claim is that the order of referential expressions in simple declarative non-emotive sentences may be influenced by information structure, and there is little evidence that this is so.**
- III. Can processing facts be of any benefit to syntax, especially as a method to choose between alternative analyses? **Potentially, yes.**

References

- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom (2000) "Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering", *Language*, 76, 1, pp. 28-55.
- Bailyn, John (2004) "Generalized Inversion", *Natural Language and Linguistic Theory* 22, pp. 1-49.
- Firbas, Jan (1992) *Functional sentence perspective in written and spoken communication*, Cambridge University Press, Cambridge.
- Gries, Stefan T. (2003) *Multifactorial Analysis in Corpus Linguistics – a study of particle placement*, Continuum, London – New York.
- Hawkins, John A. (1994) *A Performance Theory of Order and Constituency*, Cambridge University Press, Cambridge.
- Hawkins, John A. (2004) *Efficiency and Complexity in Grammars*, Oxford University Press, Oxford.
- Holden, Kyril T. & Maria Krupp (1987): "Word Order in Russian Transitive Sentences", *Folia Slavica*, vol 8, pp. 254-271.
- King, Tracy Holloway (1995) *Configuring Topic and Focus in Russian*, CLSI Publications, Stanford, California.
- Kovtunova, Irina I. (1976) *Sovremennyj russkij jazyk – porjadok slov i aktual'noe členenie predloženiya*, Prosveščenie, Moscow.
- R Development Core Team (2009) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rozental', D.Ě (1979) *Sovremennyj Russkij Jazyk – Čast 2 - Sintaksis*, Vysšaja Škola, Moskva.
- Sgall, Petr; Eva Hajičová & Jarmila Panevová (1986) *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*, D.Reidel Publishing Company, Prague.
- Siewierska, Anna & Ludmila Uhlířová (1998) "An Overview of Word Order in Slavic Languages", *Constituent Order in the Languages of Europe*, Anna Siewierska (ed.), Moyton de Grouter, Berlin – New York.
- Slioussar, Natalia (2007) *Grammar and Information Structure*, Ph.D. diss., University of Utrecht.
- Svedstedt, Dag (1976) *Position of Objective Personal Pronouns – a study of word order in Modern Russian*, Almquist & Wiksell International, Stockholm, Sweden.