# Corpus evidence for word order freezing in Russian and German

**Aleksandrs Berdicevskis**
Språkbanken
University of Gothenburg
Gothenburg, Sweden
`aleksandrs.berdicevskis@gu.se`

**Alexander Piperski**
School of Philological Studies
HSE University
Moscow, Russia
`apiperski@hse.ru`

## Abstract

We use Universal Dependencies treebanks to test whether a well-known typological trade-off between word order freedom and richness of morphological marking of core arguments holds within individual languages. Using Russian and German treebank data, we show that the following phenomenon (sometimes dubbed *word order freezing*) does occur: those sentences where core arguments cannot be distinguished by morphological means (due to case syncretism or other kinds of ambiguity) have more rigid order of subject, verb and object than those where unambiguous morphological marking is present. In ambiguous clauses, word order is more often equal to the one which is default or dominant (most frequent) in the language. While Russian and German differ with respect to how exactly they mark core arguments, the effect of morphological ambiguity is significant in both languages. It is, however, small, suggesting that languages do adapt to the evolutionary pressure on communicative efficiency and avoidance of redundancy, but that the pressure is weak in this particular respect.

## 1 Introduction

Languages are optimized for communicative efficiency, at least in some aspects of their structure (Gibson et al., 2019). One manifestation of evolutionary pressures for optimization is the trade-off in core argument marking: languages tend to have either rigid order of subject (S), verb (V) and object (O) or rich morphological marking that shows which argument is the subject and which is the object, rarely both, and never neither (Sinnemäki, 2014). This tendency, long recognized in qualitative linguistic literature (Sapir, 1921; Kiparsky, 1997), has recently received strong quantitative support, both from grammar-based (Sinnemäki, 2014), corpus-based (Futrell et al., 2015; Koplenig et al., 2017; Levshina, 2019) and experimental (MacWhinney et al., 1984; Fedzechkina et al., 2012) studies.

The explanation of the trade-off is quite straightforward. For successful communication, it is convenient to have the semantic distinction between S and O overtly coded. However, one means of coding (either syntactic or morphological) is enough, two are redundant, and languages optimize also by eliminating redundancy (Berdicevskis and Eckhoff, 2016; Fedzechkina et al., 2017).

To consider an example: English is a language with a very rigid word order (SVO) and almost no morphological marking, while Russian is exactly the opposite: there is rich morphological marking (through nominal cases and verbal agreement), and all six word order permutations are grammatically possible. It should be highlighted that both word order freedom and presence of morphological marking are gradual phenomena, not binary. English does have morphological marking for pronominal arguments (both cases and verbal agreement), while Russian word order is not entirely free, as it usually conveys pragmatic and/or stylistic information. A corpus-based approach allows us to adequately quantify these phenomena.

All quantitative typological studies cited above focused on the correlation *across* languages. We test whether an equivalent tendency can be observed *within* a language. In other words, in the previous studies, datapoints were languages, while in ours, datapoints are clauses.

More specifically, we test the following hypothesis: is it true that if a language has relatively free word order, then those clauses where S and O cannot be distinguished by morphological means (due to case syncretism or other kinds of ambiguity) will have more rigid order of S, V and O than those clauses where unambiguous morphological marking is available? This phenomenon has sometimes been dubbed *word order freezing* in the literature (see section 2).

We also make a stronger prediction: in ambiguous clauses, word order will be more often equal to the one which is basic or dominant (most frequent) in the language (for instance, SVO for Russian). Note that the second prediction entails the first, but not vice versa, at least in principle.

## 2   Word order freezing

One of the earliest explicit wordings of the word order freezing hypothesis belongs to Jakobson (1971, p. 585): "If in a language like Russian the nominal subject and object are not distinguished by morphological means, the relative order SO is compulsory". Since then, word order freezing has been claimed to exist in Russian (Mahowald, 2011), Korean, Bulgarian, Papuan languages (Lee, 2001), Japanese (Flack, 2004), German (Vogel, 2004), Hindi and Urdu (Mohanan, 1994).

There is, however, a dearth of quantitative corpus-based studies. Bouma and Hendriks (2012) demonstrate word order freezing for Dutch, while Øvrelid (2004) does the same for Norwegian. Note, however, that both Dutch and Norwegian have relatively rigid word order (and no case marking on nouns), and virtually the only deviation from the default word order (SVO) that is possible is OVS (which results from object fronting). It is thus unclear whether the results of these studies can be generalized to typologically different languages. In addition, Øvrelid (2004) focuses on a somewhat different facet of word order freezing hypothesis ("word order freezes when properties of the arguments are maximally marked in some sense") and does not directly test the role of morphological ambiguity.

We are interested in directly testing the "morphological ambiguity freezes the word order" hypothesis on two languages, Russian and German, using corpus data from Universal Dependencies (UD) 2.6 (Zeman et al., 2020). The choice of languages is driven by the following considerations: a) languages must have relatively free word order; b) there must be a substantial proportion of nominative-accusative syncretism in the nominal paradigm; c) large treebanks must be available (a pilot study on Russian data suggests that the effect is not visible on smaller datasets). Given the current state of the UD collection, that basically leaves Germanic and Slavic languages, of which we choose one language per group.

We will call a clause containing S, V and O "unambiguous" if S and O can be distinguished from morphological marking and "ambiguous" if they cannot. We will ignore semantic and pragmatic information. In the vast majority of clauses, S and O can be distinguished using context, background knowledge or common sense. For instance, in a German sentence like *Die Zahlung wickelt die Deutsche Bank ab* 'Deutsche Bank is processing the payment', it is clear that *Die Zahlung* 'payment' is the object and *die Deutsche Bank* 'Deutsche Bank' is the subject, because banks normally process payments, and not vice versa. It is, however, extremely difficult to formalize and quantify these factors (besides, they are orthogonal to our hypothesis about formal marking).

Likewise, we ignore other factors that might influence word order, including, for instance, information structure, dependency locality (Futrell et al., 2020), predictability (Ferrer Cancho, 2017), the length (weight) of arguments (Wasow, 1997) etc. Unlike Bouma and Hendriks (2012) and Øvrelid (2004), we will not use animacy and definiteness as predictors. First, the availability of this information varies: German treebanks are not annotated for animacy, while Russian treebanks are not annotated for definiteness. Second, animacy and definiteness strongly overlap with availability of overt morphological marking of syntactic role (in Russian, for instance, animate arguments will almost always be marked), which makes it difficult to reliably estimate their contribution.

## 3   Materials and methods

For Russian, we concatenate all available UD treebanks, which yields 73K sentences with 1.3M tokens. For German, we concatenate HDT and GSD (excluding the smaller treebanks, PUD and LIT, that lack

the necessary annotation[1]), which yields 206K sentences with 3.7M tokens.

We extract all clauses that contain a verb that has a nominal subject (related via NSUBJ) and a nominal object (OBJ). We include both proper and common nouns, but exclude arguments that belong to any other part of speech (pronouns etc.) to avoid adding potential confounds. If there are several coordinated arguments, we include only those that satisfy these criteria. See below for other, language-specific filters.

To establish whether a clause is ambiguous we check whether at least one of the arguments and/or the verb mark which argument is the subject and which is the object. For every argument, all available conjuncts (that pass the filter) are checked, and if at least one of them is marked, then the argument is considered marked. We treat morphological ambiguity/markedness as a binary feature, not trying to quantify degree of markedness. To achieve high accuracy, we do it using a set of language-specific handcrafted rules (see below). We manually check the accuracy of the algorithms on subsample of 100 clauses, randomly drawn from all suitable clauses.

### 3.1 Russian

We accept only nominative subjects and accusative objects. This filter excludes a small amount of non-canonical arguments (arguments of a negated verb in genitive; experiencer subjects in dative; objects of certain verbs in dative and instrumental; arguments that have a numeral modifier and get their case from the numeral rather from the verb; idiomatic constructions etc.) and misannotations.

An argument is considered non-marked if it is:

1. masculine...
   (a) and inanimate (nominative and accusative are syncretic);
   (b) and animate, and the lemma ends in a vowel other than *-a* or *-ja* (indeclinable with very few exceptions);
2. or feminine...
   (a) and not animate and plural, and the lemma ends in ' (soft sign) (nominative and accusative are syncretic);
   (b) and ends in a vowel other than *-a* or *-ja* (indeclinable);
   (c) and inanimate and plural (nominative and accusative are syncretic);
3. or neuter and not animate and plural (nominative and accusative are syncretic)...
4. or a plurale tantum (nominative and accusative are syncretic).

Otherwise the argument is considered marked, since accusative is different from nominative. If an argument is non-marked according to rules 1b, 2a (singular only) or 2b, but has an adjectival or a pronominal modifier (via AMOD or DET) and the modifier is not an indeclinable possessive pronoun (*ego* 'his', *ee* 'her', *ih* 'their'), then the argument is considered marked, since nominative and accusative are not syncretic for the modifier. This rule may generate a small amount of false positives due to (rarely occurring) indeclinable adjectives.

Russian verbs agree with the subject in number and, in past tense singular, also in gender. Thus, a verb is considered marked iff:

1. it is in past tense singular and subject and object have different genders;
2. subject and object have different numbers. If an argument is in singular, but has at least one conjunct, it is treated as plural, since plural agreement is most likely in such cases.

The manual spotcheck of 100 randomly selected clauses yielded an accuracy of 0.99 at the clause-annotation level. The only error (ambiguous clause labelled as non-ambiguous) is due to a misannotation in the treebank. In addition, two errors in labelling of individual arguments were discovered (but the clause as a whole was still labelled correctly), the rule that generated one of them was corrected in the script, no measures were taken against the other one.

---

[1]LIT lacks morphological features; PUD, unlike the other treebanks, does not have XPOS tags VAFIN and VMFIN, which are convenient for identifying auxiliary verbs.

## 3.2 German

We aspire to apply the same filter as for Russian, but the situation is complicated by the fact that most nouns in the German treebanks are not annotated for case (German has much less overt case markers on nouns than Russian). For this reason, we adapt the filter, excluding arguments that have any dependent in dative or a dependent preposition in accusative.

In addition, we restrict our analysis to main clauses in German, because in subordinate clauses SOV order is almost obligatory.

An argument is considered non-marked by default, since noun modifiers (articles, pronouns, and adjectives) do not distinguish nominative and accusative in feminine and neuter singular as well as in plural.

An argument is considered marked if it is masculine and singular and has a modifier (via AMOD or DET or their variants, such as via DET:POSS). The rationale is that most nouns except for a small group called "weak declension" (*Löwe* 'lion', *Name* 'name') and substantives derived from adjectives (*Abgeordneter* 'member of parliament') are unmarked for case, and case marking is expressed on dependent words.

German verbs agree with the subject in number. A verb is considered marked if its subject and object have different numbers. If an argument is in singular, but has at least one conjunct, it is treated as plural, since plural agreement is most likely in such cases.

If an argument has a dependent numeral, it is considered plural, even if it is annotated as singular (e.g., *zwei Prozent* 'two percent'). Many proper nouns (30%) lack annotation for number in the treebanks. In such cases, we consider them singular, since proper nouns are most often singular (this is true for 97% of the proper nouns annotated for number).

The manual spotcheck of 100 randomly selected clauses yielded an accuracy of 0.95. The errors are mostly due to incorrect annotation of compounds in the corpus.

## 4 Results

Our final dataset contains 8,575 clauses for Russian and 53,373 for German (remember that for German we have main clauses only, see section 3.2). Table 1 provides information about the proportion of ambiguous and non-ambiguous clauses and about how exactly the ambiguity is resolved (by marking subject, object, verb or several words at once).

The strategies of morphological marking are clearly different for the two languages. For Russian, the distribution is more uniform. The most frequent strategy is marking subject and verb, but not the object. It is followed by the following strategies: marking subject only; marking verb only; not marking anything; marking everything. The remaining strategies are less common, marking object only is the least frequent one. For German, the two most frequent strategies (marking verb only; not marking anything) account for 71% of all cases, followed by marking object only, the rest are less common, marking everything is virtually non-existent (it happens when both arguments are masculine singular and have a modifier, and in addition one and only one of them is coordinated with other nouns and thus behaves as if in plural). Most importantly for the current study, the proportion of ambiguous sentences is 13% for Russian and 35% for German.

The proportions of different word orders in ambiguous and unambiguous clauses are provided in Table 2. For both languages, the proportion of the basic word order (SVO) is slightly higher for ambiguous sentences, the other word orders either experience small drops in frequency or remain unchanged. Note that the distribution is different across languages: in Russian, SVO accounts for 85% of clauses (across all clauses, both ambiguous and unambiguous), followed by OVS (9%), the other four orders are infrequent, VSO extremely infrequent. In German, SVO accounts only for 56%, followed by VSO (35%), OVS (8%), the other three orders are extremely infrequent.

To quantify word order freedom, we calculate Shannon entropy (in bits) over the six possible word orders, see Table 2. In both languages the entropy is lower for the ambiguous clauses, which is in line with our main prediction. To estimate whether the difference is significant we perform a bootstrap test. We take the set of all clauses, randomly split it into two subsets A (same size as the set of unambiguous clauses) and B (same size as the set of ambiguous clauses), and calculate the absolute difference between word-order entropies for A and B. We repeat the procedure 10,000 times and estimate $p$-value as the

| Subject | Object | Verb | % in Russian | % in German |
|---------|--------|------|--------------|-------------|
| yes | yes | yes | 12 | 0.2 |
| yes | yes | no | 8 | 3 |
| yes | no | yes | 23 | 6 |
| yes | no | no | 16 | 6 |
| no | yes | yes | 8 | 4 |
| no | yes | no | 6 | 10 |
| no | no | yes | 15 | 36 |
| no | no | no | 13 | 35 |

Table 1: Proportions of various morphological marking strategies (out of 8,575 clauses for Russian and 53,373 for German); "yes" and "no" denotes whether the respective word disambiguates the clause.

|  | **Russian** | | **German** | |
|--------|--------|------|--------|------|
|  | unamb. | amb. | unamb. | amb. |
| SVO | 84.18 | 87.15 | 55.38 | 58.55 |
| SOV | 1.27 | 1.07 | 0.94 | 0.90 |
| VSO | 0.62 | 0.36 | 35.38 | 33.51 |
| VOS | 2.66 | 2.77 | 0.25 | 0.33 |
| OSV | 2.28 | 1.07 | 0.06 | 0.07 |
| OVS | 8.99 | 7.58 | 7.98 | 6.65 |
| Entropy | 0.91 | 0.76 | 1.39 | 1.34 |

Table 2: Proportions of word orders, in %; entropy of word order.

proportion of splits that yielded the absolute difference larger than or equal to the absolute difference between ambiguous and unambiguous clauses. The resulting $p$-values are 0.004 for Russian and 0 for German.

To test whether the observed differences for the proportion of the basic word order are significant (i.e. whether our second prediction holds), we apply the same kind of test: $p$-values are 0.011 for Russian and 0 for German. The bootstrap tests are visualized on Figure 1.

It can be argued that potential word-order preferences of individual verbs should be controlled for in order to ensure more rigorous testing. To do that, we fit a mixed-effects logistic regression model with word order as the dependent variable. To make the model simpler and more interpretable, we treat word order as a binary variable: SO (reference level) or OS. The independent variable is whether the clause is unambiguous (reference level), with by-verb random intercept and random slope. In *lme4* (Bates et al., 2015) notation, the model looks as follows:

$$wo \sim ambiguity + (1 + ambiguity | verb)$$

We performed the calculations in *R* (R Core Team, 2020), using the `lmerTest` package to calculate $p$-values (Kuznetsova et al., 2017). The summary of the model is presented in Table 3.

For both languages, the coefficient for ambiguity is negative, meaning that in ambiguous clauses the probability of OS word order is lower (and the probability of the basic word order, SO, is higher). For Russian, the significance of the coefficient is slightly higher than the 0.05 threshold, for German, it is much lower.

## 5   Discussion

Both in Russian and German, both our predictions are confirmed: clauses where subject and object are not marked morphologically have lower word-order entropy and higher proportion of basic word order. According to the bootstrap test, the differences are significant, but according to a mixed-effect logistic
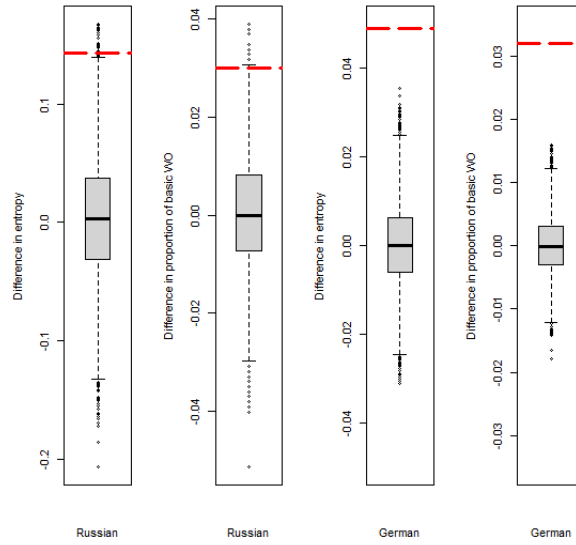
Figure 1: Boxplots showing the distribution of bootstrap test results for difference between word-order entropies and proportions of basic word order. The dashed line denotes the observed difference between ambiguous and unambiguous clauses.

| Predictor | Estimate | SE | *z* value | *p* value |
|-----------|----------|-----|-----------|-----------|
| **Russian** | | | | |
| (Intercept) | -2.11 | 0.06 | -34 | <0.001* |
| ambiguous=Yes | -0.58 | 0.30 | -2 | 0.057 |
| **German** | | | | |
| (Intercept) | -2.85 | 0.07 | -44 | <0.001* |
| ambiguous=Yes | -0.25 | 0.07 | -4 | <0.001* |

Table 3: Summary of the logistic-regression model: word order (whether it is OS) as predicted by clause ambiguity with by-verb random effects. Asterisk denotes significance at the 0.05 level.

regression model with by-verb random slope and intercept the effect of morphological ambiguity is significant only for German, but not for Russian.

Interestingly, Russian and German exhibit considerable differences with respect to how exactly the ambiguity is resolved (by marking subject, object, verb or several words at once) and to the distribution of word orders. Nonetheless, the effect is clearly present in both languages, and in both languages it is small. This implies that the pressure for word-order freezing is relatively weak (and probably mitigated by other factors).

*Pace* Jakobson (1971), the trade-off is not absolute: morphologically ambiguous clauses where word order is different from the basic one do occur in both languages. We inspected such counterexamples in order to see whether any factors that could explain why these clauses do not have any formal marking of core arguments, but did not discover any patterns. Absence of formal marking, however, is not surprising, given that in most cases the clauses can be disambiguated using semantic and pragmatic information, as well as context and background knowledge.

Further work in this direction might include psycholinguistic experiments (do human beings rely more on word order when interpreting morphologically ambiguous sentences?), diachronic studies (do morphological changes in Russian, German, or other relevant languages that increase or decrease the degree of syncretism affect word order) and machine-learning experiments designed to determine to what extent state-of-the-art parsers rely on morphology and word order (using e.g. ablation techniques à la Berdicevskis and Eckhoff (2016))

The scripts for morphological analysis and statistical tests, as well as detailed results are openly available[2].

## Acknowledgements

## References

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, 67(1):1–48.

Aleksandrs Berdicevskis and Hanne Eckhoff. 2016. Redundant features are less likely to survive: Empirical evidence from the Slavic languages. In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, and T. Verhoef, editors, *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANGX11)*. Online at `http://evolang.org/neworleans/papers/85.html`.

Gerlof Bouma and Petra Hendriks. 2012. Partial word order freezing in Dutch. *Journal of Logic, Language and Information*, 21(1):53–73.

Maryia Fedzechkina, T Florian Jaeger, and Elissa L Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.

Maryia Fedzechkina, Elissa L. Newport, and T. Florian Jaeger. 2017. Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science*, 41(2):416–446.

Ramon Ferrer Cancho. 2017. The placement of the head that maximizes predictability: An information theoretic approach. *Glottometrics*, 39:38–71.

Kathryn Flack. 2004. Ambiguity avoidance as contrast preservation: Case and word order freezing in Japanese. In Leah Bateman, Michael O'Keefe, Ellen Reilly, and Adam Werle, editors, *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*. Booksurge Publishing.

---

[2]https://github.com/AleksandrsBerdicevskis/word-order-freezing

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Quantifying word order freedom in dependency corpora. In *Proceedings of the third international conference on dependency linguistics (Depling 2015)*, pages 91–100.

Richard Futrell, Roger P Levy, and Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language*, 96(2):371–412.

Edward Gibson, Richard Futrell, Steven P Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How efficiency shapes human language. *Trends in cognitive sciences*, 23(5):389–407.

Roman Jakobson. 1971. *Selected writings [of] Roman Jakobson: Word and language*. Mouton.

Paul Kiparsky. 1997. The rise of positional licensing. In Ans van Kemenade and Nigel Vincent, editors, *Parameters of morphosyntactic change*, pages 460–494. Cambridge: Cambridge University Press.

Alexander Koplenig, Peter Meyer, Sascha Wolfer, and Carolin Müller-Spitzer. 2017. The statistical trade-off between word order and word structure – large-scale evidence for the principle of least effort. *PLOS ONE*, 12(3):1–25, 03.

Alexandra Kuznetsova, Per B Brockhoff, Rune HB Christensen, et al. 2017. lmertest package: tests in linear mixed effects models. *Journal of statistical software*, 82(13):1–26.

Hanjung Lee. 2001. Markedness and word order freezing. In Peter Sells, editor, *Formal and empirical issues in optimality theoretic syntax, volume 5 of Studies in constraint-based lexicalism*. CSLI Stanford, CA.

Natalia Levshina. 2019. Token-based typology and word order entropy: A study based on universal dependencies. *Linguistic Typology*, 23(3):533–572.

Brian MacWhinney, Elizabeth Bates, and Reinhold Kliegl. 1984. Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23(2):127–150.

Kyle Mahowald. 2011. An LFG approach to word order freezing. In Miriam Butt and Tracy King, editors, *Proceedings of LFG11*, pages 381–398. CSLI.

Tara Mohanan. 1994. *Argument structure in Hindi*. Center for the Study of Language (CSLI).

Lilja Øvrelid. 2004. Disambiguation of syntactic functions in Norwegian: modeling variation in word order interpretations conditioned by animacy and definiteness. In *Proceedings of the 20th Scandinavian Conference of Linguistics*, pages 1–17. Helsinki: University of Helsinki.

R Core Team, 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Edward Sapir. 1921. *Language: An introduction to the study of speech*. Harcourt, Brace and company.

Kaius Sinnemäki. 2014. Complexity trade-offs: A case study. In Frederick Newmeyer and Laurel Preston, editors, *Measuring grammatical complexity*, pages 179–201. Oxford University Press.

Ralf Vogel. 2004. Correspondence in ot syntax and minimal link effects. In Arthur Stepanov, Gisbert Fanselow, and Ralf Vogel, editors, *Minimality effects in syntax*, pages 401–441. Mouton de Gruyter.

Thomas Wasow. 1997. Remarks on grammatical weight. *Language variation and change*, 9(1):81–105.

Daniel Zeman, Joakim Nivre, et al. 2020. Universal dependencies 2.6. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.