Studying Formal Properties of a Free Word Order Language*

Vladislav Kuboň and **Markéta Lopatková** and **Martin Plátek**
Charles University in Prague, Faculty of Mathematics and Physics
Czech Republic
{vk,lopatkova}@ufal.mff.cuni.cz, martin.platek@mff.cuni.cz

Abstract

The paper investigates a phenomenon of free word order through the analysis by reduction. It exploits its formal background and data types and studies the word order freedom by means of the minimal number of word order shifts (word order changes preserving syntactic correctness, individual word forms, their morphological characteristics and/or their surface dependency relations).

The investigation focuses upon an interplay of two phenomena related to word order: *(non-)projectivity* of a sentence and *number of word order shifts* within the analysis by reduction. This interplay is exemplified on a sample of Czech sentences with clitics.

1 Introduction

This paper studies the free word order phenomenon through a formalization of some important notions. The phenomenon itself plays a very important role in parsing, it constitutes a substantial challenge for all kinds of parsing algorithms. The languages with higher degree of word order freedom are usually more difficult to parse, they tend to achieve worse parsing results even when identical parsing methods are applied. Although modern stochastic or machine learning methods exploited in parsing in recent years achieved a substantial improvement and their results are widely accepted for a variety of applications, they do not answer the question whether the freedom of word order is really the crucial phenomenon which not only theoretically, but also practically constitutes the greatest parsing challenge.

In this paper we are not going to discuss any particular parsing algorithm or system; instead, we would like to clarify some basic features and notions which may play a role in the investigations of the word order freedom. For this purpose we are going to exploit the method of analysis by reduction and the formal data type derived from this method, so-called D-trees. A complete description of both the method and the data type can be found for example in (Plátek, Mráz, and Lopatková 2010). The thorough formalization of both the method and the data types provides a formal background for our investigation of formal properties of the free word order.

1.1 Word Order Variations

A particular natural language falls into the category of languages with a higher degree of the word order freedom if it allows to modify the word order in its sentences without affecting their syntactical correctness. The higher number of those word order variants is possible in a given natural language, the higher its degree of word order freedom. Let us now look at this intuitive definition more closely.

The word order variations which do not affect syntactical correctness of a particular sentence can be divided into two major groups – those, which affect the word forms (and their morphological or even syntactic categories) and those who don't. The first group may be illustrated for example by the differences between an active and passive sentence:

Peter invited Mary for a walk.
Mary was invited by Peter for a walk.

Because English is a language with very sparse number of word forms derived from a single lemma, these word forms don't change as it is the case with inflective or agglutinative languages. Instead, the changes of the word order are accompanied by insertions or deletions of functional words or prepositions. This mechanism is common also in languages with richer inflection and higher degree of word order freedom, like, e.g., in German, as well as in free word order languages, as e.g. in Czech. Let us look at the translations of this sentence to German and to Czech:

Peter hat Maria für einen Spaziergang eingeladet.
Maria war von Peter für einen Spaziergang eingeladet.
Petr pozval Marii na procházku.
Marie byla pozvána Petrem na procházku.

Further, the languages with rich inflection usually involve more changes of individual word forms in addition to insertions or deletions, as we can notice in the Czech example.

However, this first type of word order variations does not constitute a good basis for the investigation of word order freedom, because too many factors are involved.

The latter group is more interesting from our point of view: the constraint that the word forms and their morphological and syntactic properties should not be changed together with the change of a word order, helps to study the

*The paper reports on the research supported by the grant of GAČR No. P202/10/1333.

word order freedom separately from other phenomena. Let us look at the following set of examples for all three languages mentioned above.

English word order is fixed, a sentence with no adverbials or other words which might be able to occupy various places in the sentence actually does not allow any variants:

Peter watched the movie with Mary.

Although German generally demonstrates more freedom than English, it also has a number of constraints which make the changes of the word order in such simple sentence relatively difficult:

Peter sah den Film mit Maria.
Mit Maria sah Peter den Film.
Den Film sah Peter mit Maria.

Czech allows more permutations, for example:

Petr se díval s Marií na film.
S Marií se Petr díval na film.
Na film se Petr díval s Marií.
Petr se díval na film s Marií.
Petr se s Marií díval na film.
Petr se na film díval s Marií.
S Marií se díval Petr na film.
S Marií se na film díval Petr.
Na film se díval Petr s Marií.
Na film se s Marií díval Petr.
...

These sentences have the same syntactic structure (apart from the word order) – the same morphological (case, number, gender, tense, ...) and syntactic categories (Subject, Predicate, Direct or Indirect Object, ...) are assigned to individual words.[1] Although the Czech example demonstrate a high number of acceptable permutations, the fact that the reflexive particle *se* occupies the second position in each sentence is not an accident. Actually, it is one of very few constraints on word order in Czech, which concerns not only reflexive particles but clitics in general.

Investigating the free word order by means of enumerating all possible permutations does not seem to be a good idea. We have to take into account the existence of a certain *gray zone* existing especially in languages with higher degree of word order freedom (and higher number of possible permutations) in which it is very difficult to judge individual permutations because they may be acceptable only in a very obscure reading. The second issue is related to the fact that the maximal number of permutations in a sentence with n words reaches $n!$ – a number too big for a manual enumeration of all variants.

The second clue obtained from our set of examples indicates that it might be better to concentrate on various constraints affecting the number of possible permutations than at the number itself. The stricter and more frequent the constraints are, the lower the number of acceptable word order permutations. The Czech examples indicate that the role of clitics might provide interesting material for a detailed investigation of the word order freedom.

This investigation will be backed up by a sound theoretical and formal background as well as by syntactically annotated data which will eliminate the uncertainty concerning the syntactic structure of sentences being investigated.

In this paper we are going to rely on the theoretical background of Functional Generative Description (Sgall, Hajičová, and Panevová 1986), here Sect. 2.1, which constitutes a theoretical basis for the data we are going to exploit, namely sample sentences from the Prague Dependency Treebank (PDT),[2] a large-scale treebank of Czech.

We are also adopting the methodology of an analysis by reduction described for example in (Lopatková, Plátek, and Sgall 2007; Plátek, Mráz, and Lopatková 2010); the authors propose and further enrich a formal model of a stratificational dependency approach to natural language description. The model is based on an elementary method of analysis by reduction (AR, see (Lopatková, Plátek, and Sgall 2007), here Sect. 2.2). The analysis by reduction has served as a motivation for a new family of automata, so called *restarting automata*, see (Otto 2003). The first step in the direction of more formal treatment of the word order freedom has been done in (Holan et al. 2000), where the authors discussed it without the exploitation of the analysis by reduction and without setting the constraints on unchanged morphological and syntactic properties of individual words. We will also use some examples from (Holan et al. 2000) and modify them according to the methods mentioned in (Plátek, Mráz, and Lopatková 2010) and according to Functional Generative Description, the linguistic theory we use as a background for the formalization.

2 The Background

2.1 Functional Generative Description

The theoretical linguistic basis for our research is provided by the Functional Generative Description (FGD in the sequel), see esp. (Sgall, Hajičová, and Panevová 1986). FGD is characterized by its stratificational and dependency-based approach to the language description.

The *stratificational approaches* split language description into layers, each layer providing complete description of a (disambiguated) sentence and having its own vocabulary and syntax. As we focus on surface word order phenomena in this project, we make use of three *surface layers* of FGD only: analytical layer (a-layer, layer of surface syntax), morphological layer (m-layer), and word layer (w-layer).[3]

FGD as a *dependency-based approach* describes surface syntactic information in a form of dependency trees (see Sect. 3.1). Individual words of a sentence are represented as nodes of the respective dependency tree, each node being a complex unit capturing the lexical, morphological and syntactic features; relations among words are represented

[1] These sentences differ in their communicative dynamism – what is an 'old information' referring to a previous context and what is a 'new information', i.e., the 'core' of the message. This difference is not significant for our purposes because our primary interests are in morphology and syntax.

[2] http://ufal.mff.cuni.cz/pdt.html

[3] We disregard here the *tectogrammatical layer*, which captures deep syntax comprising language meaning – the core concepts being dependency, valency, and topic-focus articulation.

by oriented edges. The dependency nature of these representations is very important particularly for languages with relatively high freedom of word order, which allow for non-projectivity (long distance dependencies).

For more compact representation, we can naturally integrate all relevant information from the FGD surface layers into a single dependency tree – as there is the one-to-one correspondence between items of the three surface layers, we can assign all a-, m-, and w- information for an individual word form (or punctuation mark), i.e. the whole *lexical bundle*, to a single node of a dependency tree, as in Fig. 1.

2.2 Basic Principles of the Analysis by Reduction

Analysis by reduction (AR) is based on a stepwise simplification of an analyzed sentence. It defines possible sequences of reductions (deletions) in the sentence – each step of AR is represented by *deleting* at least one word of the input sentence[4]; in specific cases, deleting is accompanied by a *shift* of a word form to another word order position. Consequently, it is possible to derive formal dependency relations between individual sentence members based on the possible order(s) of reductions, as it is described in (Lopatková, Plátek, and Sgall 2007; Plátek, Mráz, and Lopatková 2010).

Using AR, we analyze an input sentence (w-layer) enriched with the metalanguage information from the m- and a-layers. A sentence is simplified until so called *core structure* is reached (typically its predicate). When simplifying an input sentence, it is necessary to apply certain elementary constraints assuring adequate analysis on the surface layers, the most important being the principle of *correctness*: a grammatically correct sentence must remain correct after its simplification.

The basic principles of AR can be illustrated on the following Czech sentence:

(1) *Marii se Petr tu knihu rozhodl nekoupit.*
to-Mary REFL Peter that/the book decided not-to-buy
'To Mary, Peter decided not to buy the book.'

Let us look more closely at several possible reduction steps. For example, it is clear that the demonstrative pronoun *tu* 'that/the' has to be deleted prior to the noun *knihu* 'book' – otherwise, the simplified sentence would not be correct, e.g. **Marii se Petr tu rozhodl nekoupit.* '*Peter decided not to buy the to Mary.' It implies that the pronoun depends on the noun according to the AR principles. The dependency relation is represented as the edge [*tu*, *knihu*] in the dependency tree.

Similarly, the noun *knihu* 'book' must be reduced prior to the verb *nekoupit* 'not-to-buy' (as **Marii se Petr tu knihu rozhodl.* '*To Mary, Peter decided the book.' is an incorrect simplification) and thus the noun depends on the verb.

On the other hand, *Marii* 'to-Mary' and *knihu* 'book' can be reduced in an arbitrary order, thus these words are mutually independent.

We can continue in the same manner until the sentence is reduced to the pair *Se rozhodl.* '(He) decided.' However, the simplified sentence is not a correct Czech sentence – the reflexive morpheme *se* is a clitic and thus it has to be located in the 'second position' in a correct sentence.[5] For this reason, the shift operation is applied which results in a correct simplified sentence *Rozhodl se.* '(He) decided.'

This pair represents a core structure as it cannot be further simplified; technical rules are applied for creating the edge (a verb being always a governor for its REFL clitic), see (Lopatková, Plátek, and Sgall 2007). Figure 1 shows the resulting structure describing the previous sentence.

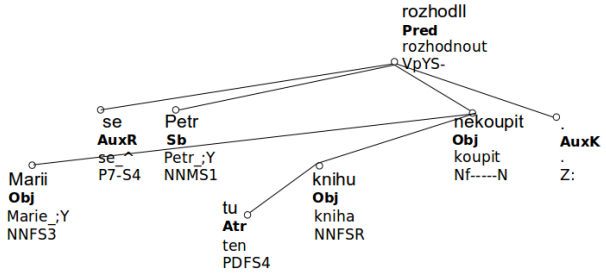

Figure 1: Dependency tree for sentence (1) (with integrated representation on a-, m- and w-layers according to FGD)

3 Formalization of Basic Notions

3.1 D-trees

One of the important factors of our formalization of word order freedom is a choice of an appropriate data type. In this paper we work with tree structures denoted as a (surface or analytical) D-trees (Delete or Dependency trees); D-tree is a rooted ordered tree with edges oriented from its leaves to its root. Nodes of each tree correspond to individual occurrences of word forms in a sentence. Moreover, we suppose a total ordering on the nodes that reflects word order in a sentence.

Thus D-tree is a triple $T = (V, H, \Gamma)$, where V, H, and Γ are sets of nodes, edges, and finite vocabulary, respectively, with the following properties:

- Each node $u \in V$ is a pair $[i, a]$, where
  - i denotes a horizontal position, which preserves left-to-right ordering of the input sentence (i.e., word order position in a sentence, also called a horizontal index);
  - a is an input word, $a \in \Gamma$; this item is referred to as a *lexical part of a node*.
- Each edge is a pair of the form $([i, a_i], [j, a_j])$, $i \neq j$.

In fact, this version of D-trees actually constitutes a special case of a DR-tree introduced in (Plátek, Mráz, and Lopatková 2010) which does not consider rewriting.

The concept of D-tree reflects the analysis by reduction (AR) (without rewriting) – its structure reflects a way how individual words of a sentence are deleted during reduction steps of the corresponding analysis by reduction. Each edge

[4]Here we leave aside possible rewriting steps necessary for an adequate analysis on the tectogrammatical layer.

[5]In Czech, clitics have specific constraints on their surface word order position, see Sect. 4.

of a D-tree connects a word form $[i, a_i]$ to some other word form $[j, a_j]$, which cannot be deleted earlier then $[i, a_i]$ during (any branch of) analysis by reduction of the same sentence.

The root of such a D-tree is one of the nodes corresponding to the word forms which remain in the sentence after the last reduction step of AR.

3.2 Measures of Non-projectivity

When considering word order freedom, we have to take into account one phenomenon which is common in languages with higher degree of word order freedom, namely non-projective constructions. In order to classify this phenomenon it is useful to define certain notions allowing for an easy definition of projectivity/non-projectivity and also for the introduction of measures of non-projectivity (these notions are formally defined in (Holan et al. 2000)).

Let us start with only an informal description of the first notion, a notion of a *coverage of a node of a* D*-tree*. The coverage of a node u identifies nodes from which there is a path to u in the D-tree (including empty path). It is expressed as a set of horizontal indices of nodes directly or indirectly dependent upon a particular node. For example, the coverage of the node of the verb *nekoupit* in Fig. 1 consists of the horizontal indices of nodes representing the words *Marii, tu, knihu, nekoupit*.

The notion of a coverage leads directly to a notion of *a hole in a subtree*. Such a hole exists if the set of indices in the coverage is not a continuous sequence. In Fig. 1 there is only a single subtree with at least one (actually two) hole in its coverage, the subtree rooted in the verb *nekoupit*.

We say that D-tree T is *projective* if none of its subtrees contains a hole; otherwise, T is *non-projective*.

In order to be able to describe necessary word order shifts in the course of the AR, we need to define a notion of equivalence for D-trees. Such equivalence (denoted as DP-equivalence) is defined as follows: DP*-equivalent trees* are those D-trees which have (i) the 'same' sets of nodes, i.e., the nodes have identical lexical parts and may differ only in their horizontal indices, and (ii) their edges always connect 'identical' pairs of nodes (nodes with identical lexical part). It actually means that a particular set of DP-equivalent trees contains the D-trees representing sentences created by a permutation of the words of the original sentence but having the same dependency relations.

For the investigation of the word order freedom it is also necessary to limit our scope and to exclude sentences which would bring into the play different phenomena than the word order. Let us therefore limit our considerations to *correct sentences* of a natural language and their *correct syntactic and morphological analysis* based on the principles of FGD.

As already mentioned, we can naturally integrate all relevant information from the FGD surface layers into a single D-tree: we assign a lexical bundle for an individual word form (or punctuation mark) – collecting a-, m-, and w- information – to a single node of a D-tree; such a D-tree is referred to as a *(correct) surface* D*-tree* (see Fig. 1 for a correct surface D-tree for sentence (1)).

A set of such surface trees is denoted as CT.

We refer to a string $w = a_1, \ldots, a_n$ corresponding to a correct surface D-tree as to a *(correct) characteristic sentence*. Thus, a (complex) symbol $a_i, i \in \{1, \ldots, n\}$, reflects a word form enriched with the relevant information from each of a-, m-, and w-layers – we call such a complex symbol a *lexical bundle*. For example, the lexical bundle for the word form *rozhodl* 'decided' consists of the word form itself (w-leyer), from its analytical function `Pred` (a-leayer), and its lemma *rozhodnout* 'to decide' and morphological tag `VpYS-` (m-leayer), see Fig. 1.

Let T be a D-tree; the set of D-trees which are DP-equivalent to T will be denoted DPE(T). In other words, DPE(T) is a set of D-trees which differ only in the word order of their characteristic sentence.

The previous concepts allow us to introduce a new feature, a *number of reduction steps enforcing a shift* during a single branch of AR. Shifts make it possible to change word order and thus 'recover' from incorrect word order that may be incurred by an AR deleting step. The *shift operation* is such a change in a D-tree when (i) the ordering of all nodes except for one is preserved, and (ii) the edges are preserved (connecting 'identical' pairs of nodes with respect to their lexical parts). It means that both the original D-tree T and the modified one belong to the same set DPE(T).

More precisely, let T be a D-tree obtained as a result of several reduction steps which is not a correct surface tree, i.e. $T \notin$ CT. Our goal is to find – if possible – a modified D-tree T' such that T' is a correct surface tree (i.e., $T' \in$ CT) and T' is DP-equivalent to T (i.e., $T' \in$ DPE(T)) by applying as small number of shift operations as possible.

4 Pilot Study on Czech Sentences

4.1 Description of the Data

In our experiment we have analyzed a sample set of data from the Prague Dependency Treebank from the point of view of word order freedom. We have focused on two phenomena related to word order: (non-)projectivity of a sentence and a necessary number of shifts in the analysis by reduction.

According to (Hajičová et al. 2004), almost one quarter of sentences from PDT 1.0 contains non-projective constructions. More precisely, among the 73 088 sentences of training data in PDT 1.0, there are 23.2 % non-projective ones, i.e., 16 920 sentences.

From the linguistic point of view, the most interesting are those examples where the non-projectivity is given by a modal verb (or a verb with similar properties) with an infinite complementation – this type of non-projectivity appears 5 696 times in 4 708 trees.[6] We have concentrated our efforts on these sentences.

The sentences with verbal non-projectivities represent an interesting material for our investigations. Despite the fact that non-projective constructions constitute a challenge to

[6] We would like to express our special thanks to Daniel Zeman who has provided the data, see also http://ufal.mff.cuni.cz/žeman/projekty/neproj/index.html.

every parsing algorithm, from our point of view they are not so problematic. It is usually possible to reduce the number of non-projective constructions to zero while preserving the correctness of a sentence simply by reordering the words in the sentence. The most regular exception from this rule are sentences containing clitics.

Clitics constitute a certain fixed point in a typical Czech sentence. They are usually located on the sentence second (Wackernagel's) position and thus they are both a frequent source of non-projective constructions and an obstacle which requires special treatment when we attempt to reduce the number of non-projectivities (see (Avgustinova and Oliva 1997)). The situation is even more complicated because the sentence second position may contain a larger number of clitics whose mutual order is not arbitrary in some cases. Let us consider the following example (taken from (Hana 2007)):

(2) *Opravit jsem se mu to včera snažil marně.*
to-repair aux-1-sg REFL him it yesterday tried fruitlessly
'I tried to repair it for him yesterday without success.'

In this sentence we may notice that the clitics are the main reason why the sentence is non-projective. While *jsem* 'aux-1-sg' and *se* REFL depend on the verb *snažil* 'tried', the pair of clitics *mu* 'him' and *to* 'it' depend on the infinite verb *opravit* 'to repair'. In this special case it is possible to make the sentence projective while preserving its correctness and all dependencies and morphological properties of all words by means of either swapping the two verbs and moving the adverb slightly forward: *Snažil jsem se mu to marně včera opravit.*; or by swapping the pairs of clitics: *Opravit mu to jsem se včera snažil marně.*

These examples actually show that the projectivization of non-projective sentences might provide a clue for calculating the degree of word order freedom for a particular sentence. Such a degree might be defined as the number of shifts or swaps performed in the course of the analysis by reduction with the purpose of preserving all important factors (grammatical correctness, morphological and syntactic information, dependency relations) in every step of the analysis.

4.2 Description of the Analysis

In order to obtain a deeper insight into the problem of mutual relationship between clitics, holes (non-projective constructions) and shifts necessary for the projectivization of sentences, we have chosen 200 non-projective sentences from the PDT with non-projectivity given by a modal verb (see above) and we have manually analyzed them using the method of analysis by reduction (AR, as described in Sect. 2.2).

As we are concentrating primarily on the clitic / (non-)projectivity interplay here (and we want to eliminate other language phenomena) we have simplified the input sentences using AR in such a way that only words related to these phenomena are preserved. In other words, we focus on those branches of AR where the words which do not contribute to the examined structures are already deleted (if it is possible without shifting). Let us exemplify this on sentence (3) (shortened sentence from PDT) and its initial simplification:

(3) *Naše firma by se možná mohla tvářit, že se jí premiérova slova netýkají (nebot' ...).*
'Perhaps our firm might pretend that the prime minister's words do not apply to it (as ...).'
⇒ *Firma by se mohla tvářit.*
'The firm might pretend.'

Typically, there are several possibilities how to analyze a simplified sentence. In our example, we can start with reducing the noun *firma* 'firm'. This results in the string starting with clitics *by* and *se* – thus a shift in word order positions must by applied to ensure the correctness of the simplified sentence. We have two possibilities of shifting: (a) We can shift the verb *tvářit* 'to pretend' to the first position, which results in the correct sentence *Tvářit by se mohla*. However, the only possible subsequent reduction step means deleting the pair *tvářit se* 'to pretend + REFL', which requires another shifting *By mohla.* → *Mohla by.* Or, (b) we can shift the verb *mohla* 'may' to the first position *Mohla by se tvářit.* The subsequent reduction of the pair *se tvářit* 'REFL + to pretend' does not require another shifting.

This example shows that if we aim at the minimal necessary number of shifts then we must apply the second type of shifting whenever possible, i.e., we shift the finite verb to the clause first position if a change of word order is enforced by analysis by reduction. Let us note that we have not found a sentence in the PDT data so far where more than one shift would be necessary when applying this strategy.

4.3 Evaluation

The results of our analysis are summarized in Tab. 1. The table shows that although clitics are usually a primary reason why a sentence contains non-projective constructions, there surprisingly seems to be no correlation between the number of holes (number of individual non-projectivities), the number of clitics and the number of shifts. It is also quite interesting that the maximal number of necessary shifts does not exceed one regardless of the number of clitics or the number of holes.

This result actually agrees with the intuition – as a language with a high degree of word order freedom Czech does not contain many constraints on the word order, therefore the number of necessary shifts is very low. This is quite an encouraging result for all parsing algorithms exploiting shifts of the word order - if such an algorithm would be able to find out what to move where, it could rely on the fact that only one shift is necessary, that once it was made, it is not necessary to search for it further.

The observations described in the Tab. 1 inspired a second experiment – if the number of holes is irrelevant, how does the relationship between the number of clitics in projective sentences and the necessary number of shifts look like? The results of this experiment are presented in Tab. 2. In this case we have taken 50 randomly chosen projective sentences with clitics from the PDT and we have performed the same evaluation as in the previous experiment. The results are also quite interesting.

# clitics	# holes	# shifts	# sentences
0	0	0	2*
	1	0	83
	2	0	1
	2	1	1
1	1	0	54
	1	1	25
	2	0	1
	2	1	3
2	1	0	11
	1	1	14
	2	0	1
	2	1	4
3	1	0	1
	1	1	2

Table 1: Sample non-projective sentences from PDT 1.0 (*annotation errors)

First, the number of clitics in projective sentences is generally lower. This supports the claim that clitics constitute one of the primary sources of non-projectivities in Czech. If the sentence contains more than two clitics, it is highly probable that it contains non-projective constructions as well.

Second, even in projective sentences it is sometimes necessary to shift the words during the analysis by reduction, otherwise some of the general constraints would be violated (usually the correctness preserving constraint). This actually supports the claim that neither the number of holes nor the number of clitics in a sentence correlates with the necessary number of shifts.

On the other hand, the fact that in both experiments it took maximally 1 shift to make the sentence projective in the course of the analysis by reduction indicates that a simple counting of shifts is not subtle enough and more detailed analysis is necessary in the future.

# clitics	# holes	# shifts	# sentences
1	0	0	11
	0	1	34
2	0	0	5

Table 2: Sample projective sentences from PDT 1.0

The method of calculating the minimal number of necessary shifts seems to have one substantial drawback: the initial simplification of the sentence. In complex sentences it is not guaranteed that the parts removed from the sentence do not require similar shifts as the main clause or that some connecting expressions (conjunction, relative pronoun etc.) do not require shifts as well. The reason why we have decided to allow the reduction despite these drawbacks is very simple – in this paper we want, as a first step, to study the phenomenon of the free word order itself, not its interaction with other linguistic phenomena in a complex sentence. The number of shifts in a sentence can somehow express the degree of word order freedom (or the number of strict word-order constraints applied) only if it is studied on simple sentences. The total number of shifts does not have any meaning for a complex sentence, we can easily construct a complex sentence with an arbitrary number of shifts simply by coordinating a desired number of clauses requiring one shift each.

Conclusion

We have focused on the phenomenon of the free word order studied within the boundaries of the formal means defined for the analysis by reduction, a method of stepwise simplification of sentences in the course of syntactic analysis. The results presented in the paper show that the proposed characteristic, a number of shifts preserving syntactic correctness and other parameters in the course of the analysis, is an important factor which provides different than traditional view of the word order freedom of individual sentences.

In the future we would like to increase the number of manually evaluated sentences in order to gain more precise results and to investigate different phenomena than clitics and their role in the word order. Apart from that, the future research will also investigate other languages in order to compare the properties of languages with lower and higher degree of word order freedom and higher number of word-order constraints than the language of this pilot study, Czech.

References

Avgustinova, T., and Oliva, K. 1997. On the Nature of the Wackernagel Position in Czech. In Junghanns, U., and Zybatow, G., eds., *Formale Slavistik*. Frankfurt am Main: Vervuert Verlag. 25–47.

Hajičová, E.; Havelka, J.; Sgall, P.; Veselá, K.; and Zeman, D. 2004. Issues of Projectivity in the Prague Dependency Treebank. *The Prague Bulletin of Mathematical Linguistics* 81:5–22.

Hana, J. 2007. *Czech Clitics in Higher Order Grammar*. Ph.D. Dissertation, The Ohio State University.

Holan, T.; Kuboň, V.; Oliva, K.; and Plátek, M. 2000. On Complexity of Word Order. *Les grammaires de dépendance – Traitement automatique des langues (TAL)* 41(1):273–300.

Lopatková, M.; Plátek, M.; and Sgall, P. 2007. Towards a Formal Model for Functional Generative Description: Analysis by Reduction and Restarting Automata. *The Prague Bulletin of Mathematical Linguistics* 87:7–26.

Otto, F. 2003. Restarting Automata and Their Relation to the Chomsky Hierarchy. In Ésik, Z., and Fülöp, Z., eds., *Proceedings of DLT 2003*, volume 2710 of *LNCS*, 55–74. Berlin: Springer.

Plátek, M.; Mráz, F.; and Lopatková, M. 2010. (In)Dependencies in Functional Generative Description by Restarting Automata. In Bordihn, H. et al., ed., *Proceedings of NCMA 2010*, volume 263 of *books@ocg.at*, 155–170. Wien, Austria: Österreichische Computer Gesellschaft.

Sgall, P.; Hajičová, E.; and Panevová, J. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel.