

# Natural Syntax of Slovenian: The Complex Sentence

Varja Cvetko-Orešnik and Janez Orešnik

*Abstract:* This paper applies the framework of Natural Syntax to complex sentences in Slovenian, with the twin goals of introducing the framework to Slavicists and showing how it deals with Slavic language data. The framework of Natural Syntax as initiated by Janez Orešnik, in the tradition of (morphological) naturalness as established by Wolfgang Dressler and Willi Mayrthaler, is a developing deductive theory. The naturalness judgments are couched in naturalness scales, which follow from the basic parameters (or “axioms”) listed at the beginning of the paper. The predictions of the theory are calculated in what are known as deductions, the chief components of each being a pair of naturalness scales and the rules governing the alignment of corresponding naturalness values. Parallel and chiasitic alignment are distinguished and related to Henning Andersen’s early work on markedness.

## 1. Introduction

Natural Syntax is a (developing) deductive linguistic theory that determines the presuppositions on the background of which a (morpho)syntactic state of affairs can be made predictable and thus synchronically explained.<sup>1</sup> The two basic kinds of presuppositions are what are known as naturalness scales and rules of alignment among corresponding values of any two scales. Every (morpho)syntactic state of affairs is represented by two comparable variants. Natural Syntax contains no generative component. The basic format of our naturalness scales is  $>\text{nat} (A, B)$ , where  $A$  is more natural than  $B$ . Two expanded scales are allowed, viz.,  $>\text{nat} (A + B, B)$  and  $>\text{nat} (A, A + B)$ ; they are valid if the corresponding scale of the format  $>\text{nat} (A, B)$  is valid. These are exemplified below. We begin by listing the criteria with which Natural Syntax substantiates naturalness scales:

---

<sup>1</sup> Some recent work related to Natural Syntax includes Orešnik 2007a–e; 2008a–c and Cvetko-Orešnik and Orešnik 2009. Only work written in English is mentioned.

- (a) The speaker/hearer parameter. In the scale  $>nat(A, B)$ , value A is natural for the speaker (and unnatural for the hearer); value B is unnatural for the speaker (and natural for the hearer). The basic naturalness scale is  $>nat$  (favorable for the speaker, favorable for the hearer). This view of naturalness is commonplace in linguistics (Havers 1931: 171), under the names of tendency to economize (utilized first of all by the speaker) and tendency to be accurate (mainly in the hearer's interest).<sup>2</sup>
- (b) The principle of least effort (Havers 1931: 171). What conforms better to this principle is more natural for the speaker. What is cognitively simple (for the speaker) is easy to produce, easy to retrieve from memory, etc.
- (c) Degree of integration into the construction. What is better integrated into its construction is more natural for the speaker.<sup>3</sup>
- (d) Frequency. What is more frequent (tokenwise) is more natural for the speaker. What is cognitively simpler (for the speaker) is used more.<sup>4</sup>
- (e) Small vs. large class. The use of (a unit pertaining to) a small class is more natural for the speaker than the use of (a unit pertaining to) a large class. During speech, small classes are easier for the speaker to choose from than are large classes. (This is frequency typewise.)
- (f) The process criterion. Any process is natural. Examples of processes include movement and agreement.

---

<sup>2</sup> We follow Mayerthaler (1981: 13ff.) in assuming that the speaker is the center of communication, and therefore most properties of the speaker are natural; for instance, being the first person and/or the subject and/or +human and/or +masculine and/or +singular and/or +definite and/or +referential, and so on.

What is favorable for the hearer may be less natural for the speaker. This is a pivotal point in Natural Syntax and will be maintained until some good counterexample nullifies it. By way of illustration, it can be pointed out that producing a longish noun phrase may be tiresome for the speaker (= less natural for the speaker) but may ease the hearer's decoding process considerably (= be more natural for the hearer).

<sup>3</sup> As a rule of thumb, what is located at the margin of a construction is less natural (for the speaker) than what is placed inside a construction.

<sup>4</sup> However, the reverse does not obtain: what is natural for the speaker is not necessarily more frequent.

- (g) Acceptable vs. non-acceptable use. What is acceptable is more natural for the speaker than what is not acceptable. The very reason for the acceptability of a syntactic unit is its greater naturalness for the speaker with respect to any corresponding non-acceptable unit.
- (h) What is more widespread in the languages of the world is more natural for the speaker (the typological criterion). What is cognitively simpler (for the speaker) is realized in more languages.

The above criteria of naturalness (henceforth also called axioms) are utilized to support our naturalness scales.<sup>5</sup> Normally it suffices to substantiate any scale with one criterion, which backs up either value A or value B of the scale; the non-supported value is allotted the only remaining position in the scale. Of course, a scale may be supported with more than one criterion. Any clash among the criteria applied to a scale is to be handled with constraints on the combinations of criteria.

### 1.1. Example of a Deduction

Naturalness scales are an essential part of what are known as deductions, in which Natural Syntax expresses its predictions about the state of affairs in language data. The following is an example of a deduction:

In English, the numerical indication of frequency normally consists of a cardinal number followed by the word *times* (e.g., *four times*) except that there are one-word expressions available for the lowest numbers: *once*, *twice*, and the archaic *thrice* (Collins Cobuild 1990: 270–71).<sup>6</sup>

---

<sup>5</sup> We have been applying these criteria to language material covering several languages and miscellaneous (morpho)syntactic states of affairs. Throughout our work, these criteria have compelled us to reject certain solutions and to give precedence to other solutions. Given this encouraging experience, we will retain the list in (a–h) until some convincing and irreparable counterexample casts doubt upon our axioms. The occurrence of such an event is in the overriding interest of Natural Syntax anyway. The only realistic aim of deductive theories is that they are eventually disproved. We are afraid that any improvement of the axioms would lead to a reduction of the chances for the desirable definitive outcome.

<sup>6</sup> Henning Andersen (p.c.) has pointed out to us that there is a parallel system covering numerical indications of frequency (*one additional time*, *two/three/four additional*

- The two variants: the type *once* and the type *four times*.
1. The assumptions of Natural Syntax:
    - 1.1. >nat (type *once*, type *four times*); i.e., the type *once* is more natural than the type *four times*. According to the criterion of least effort, item (b) in the list of axioms.<sup>7</sup>
    - 1.2. >nat (low, non-low) / number; i.e., any low number is more natural than any non-low number.<sup>8</sup> Low numbers are more easily accessible to the speaker. According to the speaker/hearer criterion, item (a) in the list of axioms.
  2. The rules of parallel alignment of corresponding values:
    - 2.1. value A tends to associate with value C,
    - 2.2. value B tends to associate with value D.<sup>9</sup>
  3. The consequences: if the language distinguishes between low and non-low numbers in numerical indications of frequency such that one kind of number uses the pattern *four times* and the other kind of number uses the pattern *once*, it is the low numbers that tend to use the pattern *once* and it is the non-low numbers that tend to use the pattern *four times*. Q.E.D.

This deduction maintains that the state of affairs cannot be the reverse; that is, that numerals above *two* (or *three*) would be one-word formations and that the numerals under *three* (or *four*) would be two-word formations. All predictions of our Natural Syntax are restricted to such modest claims about the unlikelihood of the reverse situation.

---

*times*, etc.), which does not make use of the dichotomy treated in the above deduction. Donald Reindl (p.c.) has added *one more time*, *two/three/four more times*, etc.

<sup>7</sup> Natural Syntax cannot predict the cutoff point between low and non-low numerals.

<sup>8</sup> Mayerthaler 1981: 15.

<sup>9</sup> Value A of scale 1.1 (= the type *once*) tends to combine with value C of scale 1.2 (= low number). Value B of scale 1.1 (= the type *four times*) tends to combine with value D of scale 1.2 (= non-low number). Similarly, it follows in the remaining deductions, with the proviso that the alignment (unlike here) is chiastic in most cases. Chiastic alignment is explained below.

## 1.2. Alignment

In every deduction, the rules of alignment play a prominent role; consider item 2 in the above deduction. The alignment rules regulate the combinations of corresponding values of the two naturalness scales mentioned in the deduction.

The alignment can be parallel or chiasitic. Suppose that the two scales are  $>\text{nat} (A, B)$  and  $>\text{nat} (C, D)$ . Parallel alignment pairs value A with value C, and value B with value D. Chiasitic alignment pairs A with D, and B with C.

A paramount question is whether the alignment is parallel or chiasitic. Parallel alignment is the default case. Chiasitic alignment is necessary whenever a given deduction is limited to the language data obtaining within an “unnatural environment”. This is defined as value B of the scale  $>\text{nat} (A, B)$ .

For example, in the scale  $>\text{nat} (\text{main}, \text{dependent}) / \text{clause}$ , the value “dependent clause” is an unnatural environment. This means: all deductions whose language data lie within the environment “dependent clause” require the implementation of chiasitic alignment.

Chiasitic alignment is prohibited when a naturalness scale is substantiated with an axiom. If, however, an axiom is engaged as one of the scales in a deduction, it obeys the usual distribution of the alignment rules.<sup>10</sup> The insistence of Natural Syntax on the distinction between parallel and chiasitic alignments stems indirectly from Henning Andersen’s work on markedness theory. Andersen observes situations such as the following in all human semiotic systems: on an everyday occasion casual wear is unmarked, and formal wear marked; on a festive occasion it is the formal wear that is unmarked, whereas casual wear is marked. See Andersen (1972: 45, esp. fn. 23). This example can be expressed with our scaling: (i)  $>\text{nat} (\text{casual}, \text{formal}) / \text{wear}$ ; (ii)  $>\text{nat} (-, +) / \text{marked}$ . A third scale is the source of the environment of the deduction:  $>\text{nat} (\text{everyday}, \text{festive}) / \text{occasion}$ . If the environment is an “everyday occasion”, the alignment within (i–ii)

---

<sup>10</sup> Although Natural Syntax in principle does not deal with semantic phenomena, it does happen in rare cases that semantics is involved in a deduction. Semantic phenomena never seem to be the cause of chiasitic alignment. See the deduction discussed in section 2.3 below.

is parallel; if the environment is “festive occasion”, the alignment within (i–ii) is chiasitic.

### 1.3. Goals

Our aim is to show that Natural Syntax can be successfully applied to Slavic language material. Specifically, the aim is to provide confirmation of results from other languages, to show that naturalness scales apply to languages other than those first tested and to confirm that languages that might have been expected to behave differently within the areas of coordination and subordination in fact follow the same pattern. Another important feature of Natural Syntax is the predictions of frequency phenomena. See axioms (d) (on token frequency) and (e) (on type frequency), and the deductions discussed in sections 2.1, 2.2, 3.6, 3.8, and 4.4 below.

We turn now to the topic of this paper. In our past work, the two values of any naturalness scale have never pertained to the immediate constituents of one and the same complex sentence; that is, the values of a single scale have never been conjoins of one and the same coordination or the matrix clause and the subordinate clause of one and the same subordination. In this paper it will be shown that Natural Syntax can process even such cases successfully, without any additional apparatus.

The basic naturalness scale for complex sentences is  $>nat$  (simple, complex) / sentence; that is, a simple sentence is more natural than a complex sentence. This is so because simple sentences are on average shorter than complex ones and have less internal structure. Therefore simple sentences are natural according to the criterion of least effort, item (b) in the list of axioms. Consequently complex sentences, coordination, and subordinations constitute an unnatural environment (and invariably require chiasitic alignment in corresponding deductions).<sup>11</sup>

The body of this paper is divided into three sections: coordination (section 2), subordination (section 3), and miscellaneous (section 4). As is customary in Natural Syntax, the examples are presented in the format of deductions.

---

<sup>11</sup> The terms coordination and subordination also apply to units smaller than a clause; for instance, *very, very, very (good)* (= asyndetic coordination). For reasons of space, subclausal material is not considered below.

## 2. Coordination

Exemplification is limited to coordination consisting of only two conjoins. As a further simplification, each conjoin will be a main clause (although this need not be relevant in all cases). The basic scale involving any pair of conjoins is >nat (first, second) / conjoin; that is, the first conjoin of a coordination is more natural than the second conjoin. The second conjoin usually contains the coordinator, and this additional material makes the second conjoin unnatural according to the criterion of least effort, item (b) in the list of axioms.

### 2.1. Slovenian: Asyndetic and Syndetic Coordination

In Slovenian, asyndetic coordination is rarer in texts than syndetic coordination.<sup>12</sup> The data were suggested to us by the situation in English (Quirk et al. 1985: 918) and obtain in Slovenian as well. Examples: (asyndetic) *prišel sem, videl sem, zmagal sem* 'I came, I saw, I conquered'; (syndetic) *Janez poje in Micka pleše* 'Janez is singing and Micka is dancing.'

- The two variants: syndetic and asyndetic coordination. The deduction proceeds in the unnatural environment "complex sentence".

#### 1. The assumptions of Natural Syntax:

- 1.1 >nat (asyndetic, syndetic) / coordination; i.e., asyndetic coordination is more natural than syndetic coordination. Unlike any syndetic coordination, any asyndetic coordination lacks a coordinator and is therefore natural.<sup>13</sup>
- 1.2 >nat (more, less) / frequent tokenwise; i.e., more frequent is more natural than less frequent.<sup>14</sup>

#### 2. The rules of chiastic alignment apply.

<sup>12</sup> Unlike syndetic coordination, asyndetic coordination lacks overt coordinators.

<sup>13</sup> According to the criterion of least effort, item (b) in the list of axioms.

<sup>14</sup> This is the frequency criterion itself, item (d) in the list of axioms.

3. The consequences: if the language distinguishes between syndetic and asyndetic coordination such that one type is more frequent and the other type is less frequent, then it is syndetic coordination that tends to be more frequent and it is asyndetic coordination that tends to be less frequent. Q.E.D.

## 2.2. Slovenian: Single and Correlative Coordinators

In Slovenian, most coordinators are single words; for instance, *in* 'and'. A small class are correlative coordinators, for instance, *ali pojdi z nami ali ostani doma* 'either go with us or stay at home'. The data were suggested to us by the situation in English (Carter and McCarthy 2006: 557) and obtain in Slovenian as well.

- The two variants: single and correlative coordinators. The deduction proceeds in the unnatural environment "complex sentence".
1. The assumptions of Natural Syntax:
    - 1.1.  $>\text{nat}$  (single, correlative) / coordinator; i.e., a single coordinator is more natural than a correlative coordinator; see fn. 13.
    - 1.2.  $>\text{nat}$  (small, large) / class of coordinators; i.e., a small class is more natural than a large class.<sup>15</sup>
  2. The rules of chiastic alignment apply.
  3. The consequences: if the language distinguishes between single and correlative coordinators such that one option constitutes a small class and the other option constitutes a large class, then it is the single coordinators that tend to constitute a large class and it is the correlative coordinators that tend to constitute a small class. Q.E.D.

---

<sup>15</sup> This is the very criterion of small vs. large class, item (e) in the list of axioms.



### 2.3. Slovenian: First + Second Conjoin

In Slovenian, in general, “first + second conjoin” (= “f + s”) does not mean the same as “s + f”; for instance, *Janez je šel domov in povečerjal* ‘Janez went home and had dinner’ is different from *Janez je povečerjal in šel domov* ‘Janez had dinner and went home’. However, sometimes the equation “f + s” = “s + f” does obtain; for instance, *Janez pije pivo in Micka pije vino* ‘Janez is drinking beer and Micka is drinking wine’ equals *Micka pije vino in Janez pije pivo* ‘Micka is drinking wine and Janez is drinking beer’. This is in principle possible with *in* ‘and’ and impossible with, say, *toda* ‘but’. Thus the equation “f + s” = “s + f” represents a smaller class of cases than the inequation “f + s” ≠ “s + f”.

- The two variants: “f + s” = “s + f”; “f + s” ≠ “s + f”. The deduction proceeds in the unnatural environment “complex sentence”. However, chiasitic alignment is not indicated because the deduction treats a semantic phenomenon.

#### 1. The assumptions of Natural Syntax:

- 1.1. >nat (“f + s” = “s + f”; “f + s” ≠ “s + f”); i.e., “f + s” = “s + f” is more natural than “f + s” ≠ “s + f”. The former value of the scale contains more (pure and simple) repetition than the latter value of the scale. Repetition is very natural; humans have inherited it from higher mammals. The former value of the scale is therefore natural.
- 1.2. >nat (small, large) / class; i.e., a small class is more natural than a large class; see fn. 15.

#### 2. The rules of parallel alignment apply.

3. The consequences: if the language distinguishes between the equation “f + s” = “s + f” and the inequation “f + s” ≠ “s + f” such that one option constitutes a small class and the other option constitutes a large class, then it is the inequation “f + s” ≠ “s + f” that tends to constitute a large class and it is the equation “f + s” = “s + f” that tends to constitute a small class. Q.E.D.

## 2.4. Slovenian: Ellipsis of Repeated Units

In Slovenian, in coordination, the ellipsis of the repeated unit is only possible in the second conjoin; for instance, *Janez pije pivo in Micka (pije) vino* ‘Janez is drinking beer and Micka (is drinking) wine’. The data were suggested to us by the situation in English (Biber et al. 1999: 156) and obtain in Slovenian as well.

- The two variants: the first and the second conjoin. The deduction proceeds in the unnatural environment “complex sentence”.
  1. The assumptions of Natural Syntax:
    - 1.1. >nat (first, second) / conjoin; i.e., the first conjoin is more natural than the second conjoin.<sup>16</sup>
    - 1.2. >nat (+, -) / ellipsis; i.e., realized ellipsis is more natural than non-realized ellipsis; see fn. 13. A special case of 1.2:
      - 1.2.1. >nat (+/-, -) / ellipsis; scale 1.2.1 assumes the permitted expanded format >nat (A + B, B) and is automatically valid because the corresponding basic scale 1.2 has been substantiated.
  2. The rules of chiastic alignment apply.
  3. The consequences: if the language distinguishes, in coordination, between the first and the second conjoin such that one conjoin allows ellipsis and the other conjoin does not allow it, then it is the first conjoin that tends not to allow ellipsis and it is the second conjoin that tends to allow ellipsis. Q.E.D.

## 2.5. Slovenian: Subordinator Ellipsis

In Slovenian, in the example sentence *Janez pravi, [da Micka igra šah] in [(da) Tone plava]* ‘Janez says [that Micka is playing chess] and [(that) Tone is swimming]’ only the latter instance of the subordinator can be ellided. The data were suggested to us by the situation in English

---

<sup>16</sup>See the introduction to section 2.

(Radford 1997: 149) and obtain in Slovenian as well. In this particular case the coordination consists of two subordinate clauses.

- The two variants: the first and the second conjoin. The deduction proceeds in the unnatural environment “complex sentence”.
1. The assumptions of Natural Syntax:
    - 1.1.  $>\text{nat}$  (first, second) / conjoin; i.e., the first conjoin is more natural than the second conjoin.<sup>17</sup>
    - 1.2.  $>\text{nat}$  (–, +) / subordinator; i.e., the absence of a subordinator is more natural than its presence; see fn. 13. A special case of 1.2:
      - 1.2.1.  $>\text{nat}$  (+/–, +) / subordinator; scale 1.2.1 assumes the permitted expanded format  $>\text{nat}$  (A + B, B) and is automatically valid because the corresponding basic scale 1.2 has been substantiated.
  2. The rules of chiastic alignment apply.
  3. The consequences: if the language distinguishes, in coordination, between the first and the second conjoin such that one can contain a subordinator and the other must contain that subordinator, then it is the first conjoin that tends to contain an obligatory subordinator and it is the second conjoin that tends to contain an optional copy of that subordinator. Q.E.D.

### 3. Subordination

This exemplification is limited to subordinations consisting of only one superordinate clause and only one subordinate clause. As a further simplification, the superordinate clause will mostly be a matrix clause.

The basic naturalness scale governing subordination is  $>\text{nat}$  (matrix, subordinate) / clause; that is, a matrix clause is more natural than a subordinate clause. Matrix clauses are more frequent in texts than subordinate clauses and therefore natural. Unlike most matrix clauses,

---

<sup>17</sup> See the introduction to section 2.

subordinate clauses are introduced by a subordinator; because of this additional material they are unnatural.

At first, there seems to be a contradiction here if the criterion of integration into the construction, item (c) in the list of axioms, is applied: the subordinate clause is integrated into its matrix clause, not vice versa. Therefore, contrary to the wording of the scale in the preceding paragraph, the subordinate clause appears more natural than the matrix clause. However, closer inspection reveals that this contradiction is artificial. Disregarding for a moment our self-imposed limitation to one matrix and corresponding subordinate clause, it has to be recalled that superordinate clauses can also be embedded into even higher clauses. Consequently, the criterion of integration into the construction cannot differentiate between superordinate and corresponding subordinate clauses. It is another matter that the subordinate clause is typically better integrated and the matrix clause is typically less integrated into the complex clause consisting of that matrix clause and that subordinate clause.<sup>18</sup>

### 3.1. Universal: Conditional Constructions

Universally, the protasis is always the topic (also called the theme), being the speaker's and hearer's shared presupposition (Haiman 1978). Consequently the apodosis must be the rheme. For example, *če bo deževalo* (= protasis), *bom ostal doma* (= apodosis) 'if it rains, I'll stay at home.'

- The two variants: the protasis and the apodosis. The deduction proceeds in the unnatural environment "complex sentence".

#### 1. The assumptions of Natural Syntax:

- 1.1. >nat (apodosis, protasis); i.e., the apodosis is more natural than the protasis. In most cases the scale corresponds to >nat (matrix, subordinate) / clause.<sup>19</sup>

---

<sup>18</sup> This matter is dealt with in the deduction discussed in section 3.3 below. It will be seen that the apparent inconsistency arises because chiasmic alignment must be applied.

<sup>19</sup> See the introduction to section 3.

- 1.2. >nat (theme, rheme); i.e., the theme is more natural than the rheme (Mayerthaler 1981: 15). The rheme, representing new information, is in the special interest of the hearer, and therefore it must be mentioned in slot B of the scale.<sup>20</sup>
2. The rules of chiasitic alignment apply.
3. The consequences: if the language distinguishes, within conditional constructions, between the protasis and the apodosis such that one is the theme and the other is the rheme, then it is the apodosis that tends to be the rheme and it is the protasis that tends to be the theme. Q.E.D.

### 3.2. Slovenian: Finite or Non-Finite Subordinate Clause

In Slovenian, in subordination, the matrix clause is usually finite, whereas the subordinate clause can be finite or non-finite (usually depending on the kind of matrix clause). For instance, *Micka je dovolila Janezu, da odide* (the underlined part is a *that*-clause); *Micka je dovolila Janezu oditi* (the underlined part is an infinitive), both meaning 'Micka permitted Janez to leave'.

- The two variants: matrix and subordinate clause. The deduction proceeds in the unnatural environment "complex sentence".
1. The assumptions of Natural Syntax:
    - 1.1 >nat (matrix, subordinate) / clause; i.e., the matrix clause of a subordination is more natural than the corresponding subordinate clause.<sup>21</sup>
    - 1.2. >nat (–, +) / finite clause; i.e., a non-finite clause is more natural than a finite clause. Non-finite clauses often lack realized subjects or even further properties of finite clauses. Non-finite clauses are therefore natural; see fn. 13. A special case of 1.2:

<sup>20</sup> According to the speaker/hearer criterion, item (a) in the list of axioms.

<sup>21</sup> See the introduction to section 3.

1.2.1. >nat (+/−, +) / finite clause; scale 1.2.1 assumes the permitted expanded format >nat (A + B, B) and is automatically valid because the corresponding basic scale 1.2 has been substantiated.

2. The rules of chiastic alignment apply.
3. The consequences: if the language distinguishes, within subordinations, between a matrix and a subordinate clause such that one is finite and the other is finite or non-finite, then it is the matrix clause that tends to be finite, and it is the subordinate clause that tends to be finite or non-finite. Q.E.D.

### 3.3. Slovenian: Matrix Clause of a Subordination

In Slovenian, the subordinate clause is better integrated into the complex clause than the matrix clause. In fact, the subordinate clause is usually an element of the corresponding matrix clause, which makes the matrix clause a more independent element of the subordination (less integrated) than the subordinate clause (better integrated). For example, *Janez pravi, da ga zebe* 'Janez says that he is cold.'

- The two variants: the matrix and the subordinate clause. The deduction proceeds in the unnatural environment "complex sentence".

1. The assumptions of Natural Syntax:
  - 1.1. >nat (matrix, subordinate) / clause; i.e., the matrix clause of a subordination is more natural than the corresponding subordinate clause.<sup>22</sup>
  - 1.2. >nat (more, less) / integrated; i.e., what is better integrated is more natural than what is less integrated.<sup>23</sup>

2. The rules of chiastic alignment apply.

---

<sup>22</sup> See the introduction to section 3.

<sup>23</sup> This is the very criterion of integration into the construction, item (c) in the list of axioms.

3. The consequences: if the language distinguishes between matrix and subordinate clauses such that one kind of clause is better integrated into the complex sentence and the other kind of clause is less integrated into the complex sentence, then it is the matrix clause that tends to be less integrated into the complex sentence and it is the subordinate clause that tends to be better integrated into the complex sentence. Q.E.D.

### 3.4. Slovenian: Conditional Constructions

In Slovenian, one way of expressing a conditional construction referring to an unreal situation is to use the subordinator *da* and the present or past indicative in the subordinate clause; the matrix clause contains the conditional mood (i.e., the particle *bi* and the *l*-participle). For instance, *da si me vprašal, bi ti dal* ‘if you had asked me I would have given [it to] you’ (Toporišič 2000: 394).

- The two variants: (within a subordination whose subordinator is *da*) the matrix and the subordinate clause. The deduction proceeds in the unnatural environment “complex sentence”.

1. The assumptions of Natural Syntax:
  - 1.1. >nat (matrix, subordinate) / clause; i.e., the matrix clause of a subordination is more natural than the corresponding subordinate clause.<sup>24</sup>
  - 1.2. >nat (indicative, *bi* + *l*-participle); i.e., the indicative is more natural than other moods. In many languages the indicative is zero coded, and therefore natural,<sup>25</sup> see fn. 13.
2. The rules of chiastic alignment apply.

---

<sup>24</sup> See the introduction to section 3.

<sup>25</sup> Very likely the imperative mood is even more natural than the indicative mood, but this circumstance is not relevant here.

3. The consequences: if the language distinguishes (within a subordination whose subordinator is *da*) between a matrix and a subordinate clause such that one clause contains an indicative and the other clause contains a *bi + l*-participle, then it is the matrix clause that tends to contain a *bi + l*-participle and it is the subordinate clause that tends to contain an indicative. Q.E.D.

### 3.5. Slovenian: Prolepsis

In Slovenian, in subordination consisting of a matrix clause and a subordinate clause (in that order) it often happens that an NP is mentioned as the subject of the subordinate clause; for instance, *glejte, kako lilije rastejo* ‘behold how the lilies grow’. Prolepsis happens less often. Under prolepsis, the first mention of that NP occurs as the object of the matrix clause, and then the subordinate clause contains (in the subject) a second mention of that NP (usually pronominalized); for instance, *glejte lilije<sub>v</sub>, kako Ø<sub>i</sub> rastejo* ‘behold the lilies<sub>i</sub> how they<sub>i</sub> grow’.

- The two variants: (within prolepsis) the object of the matrix clause and the subject of the corresponding subordinate clause. The deduction proceeds in the unnatural environment “complex sentence”.
1. The assumptions of Natural Syntax:
    - 1.1. >nat (matrix, subordinate) / clause; i.e., the matrix clause of a subordination is more natural than the corresponding subordinate clause.<sup>26</sup>
    - 1.2. >nat (subject, object) / co-referential; i.e., a subject is more natural than an object. The speaker (as the center of communication) is usually the subject, and therefore the subject is natural; see fn. 20. Moreover, the subject is more frequent textwise than the object and therefore the subject is natural; see fn. 14.
  2. The rules of chiastic alignment apply.

---

<sup>26</sup> See the introduction to section 3.



3. The consequences: if the language distinguishes (within prolepsis) between co-referential subject and object such that one clause element is located in the matrix clause and the other clause element is located in the subordinate clause, then it is the subject that tends to be in the subordinate clause and it is the object that tends to be in the matrix clause. Q.E.D.

### 3.6. Slovenian: Initial and Final Position of the Matrix Clause

In Slovenian, a subordinate clause can precede (less often) or follow (more often) its matrix clause (Toporišič 2000: 636).<sup>27</sup> For instance: (subordinate clause preceding) *da je škoda, je jasno* 'it is clear that it is a pity', (subordinate clause following) *jasno je, da je škoda* 'idem'.

- The two variants: initial and final position of the matrix clause. The deduction does NOT proceed in the unnatural environment "complex sentence" because the deduction does not affect the interior of the clauses involved. Chiastic alignment is not indicated.
1. The assumptions of Natural Syntax:
    - 1.1. >nat (matrix + subordinate, subordinate + matrix) / sentence; i.e., matrix + subordinate sentence is more natural than subordinate + matrix sentence. The subordinate + matrix sentence offers the hearer, at the very outset, the most important signal (the conjunction) for decoding the sentence, and therefore the initial subordinate clause is in the special interest of the hearer and must be mentioned in slot B of the scale; see fn. 20.
    - 1.2. >nat (more, less) / common; i.e., what is more common is more natural than what is less common; see fn. 14.
  2. The rules of parallel alignment apply.

---

<sup>27</sup> The case of the subordinate clause interrupting another clause is treated in the deductions discussed in 3.7–3.9.

3. The consequences: if the language distinguishes between matrix + subordinate sentence and subordinate + matrix sentence such that one option is more common and the other option is less common, then it is the matrix + subordinate sentence that tends to be more common and it is the subordinate + matrix sentence that tends to be less common. Q.E.D.

### 3.7. Slovenian: Interrupting Subordinate Clause (Style)

In Slovenian, a subordinate clause interrupting another clause occurs in more formal styles, for instance, *šli bomo, če bo treba, do konca* 'we will go, if need be, to the end'. The data were suggested to us by the situation in English (Carter and McCarthy 2006: 555) and obtain in Slovenian as well.

- The two variants: middle and initial/final position of subordinate clause. The deduction proceeds in the unnatural environment "complex sentence".
1. The assumptions of Natural Syntax:
    - 1.1. >nat (middle, initial/final) / position of subordinate clause; i.e., the middle position is more natural than the initial/final position. The middle position is better integrated into its construction than the initial/final position (the latter are located at the margins of the construction). Therefore the middle position is natural; see fn. 23.
    - 1.2. >nat (–, +) / formal language; i.e., non-formal language is more natural than formal language. Many speech communities use no formal language or very little. Therefore non-formal language is natural.<sup>28</sup> A special case of 1.2:
      - 1.2.1. >nat (+/–, +) / formal language; scale 1.2.1 assumes the permitted expanded format >nat (A + B, B) and is automatically valid because the corresponding basic scale 1.2 has been substantiated.

2. The rules of chiasitic alignment apply.

---

<sup>28</sup> According to the typological criterion, item (h) in the list of axioms.

3. The consequences: if the language distinguishes between the middle and initial/final position of subordinate clauses such that one position obtains in formal styles and the other position obtains in any style, then it is the middle position that tends to obtain in formal styles and it is the initial/final position that tends to obtain in any style. Q.E.D.

### 3.8. Slovenian: Interrupting Subordinate Clause (Position)

In Slovenian, a subordinate clause interrupting another clause occurs less often than a subordinate clause in initial or final position. This frequency relation can be deduced from the circumstance that, unlike subordinate clauses in initial or final position, subordinate clauses in middle position are limited to formal styles. For instance, *šli bomo, če bo treba, do konca* 'we will go, if need be, to the end'. The data were suggested to us by the situation in English (Carter and McCarthy 2006: 555) and obtain in Slovenian as well.

- The two variants: middle and initial/final position. The deduction proceeds in the unnatural environment "complex sentence".
1. The assumptions of Natural Syntax:
    - 1.1. >nat (middle, initial/final) / position of subordinate clause; i.e., the middle position is more natural than the initial/final position. The middle position is better integrated into its construction than the initial/final position (the latter are located at the margins of the construction). Therefore the middle position is natural; see fn. 23.
    - 1.2. >nat (more, less) / frequent tokenwise; i.e., more frequent is more natural than less frequent; see fn. 14.
  2. The rules of chiasitic alignment apply.

3. The consequences: if the language distinguishes between the middle and the initial/final position of subordinate clauses such that one position is more common and the other position is less common, then it is the middle position that tends to be less common and it is the initial/final position that tends to be more common. Q.E.D.

### 3.9. Slovenian: Interrupting Subordinate Clause (Markedness)

In Slovenian, a subordinate clause interrupting another clause is stylistically marked (creating emphasis, signaling contrast, etc.), for instance, *šli bomo, če bo treba, do konca* 'we will go, if need be, to the end'. The data were suggested to us by the situation in English (Carter and McCarthy 2006: 555) and obtains in Slovenian as well.

- The two variants: middle and initial/final position. The deduction proceeds in the unnatural environment "complex sentence".
1. The assumptions of Natural Syntax:
    - 1.1. >nat (middle, initial/final) / position of subordinate clause; i.e., the middle position is more natural than the initial/final position. The middle position is better integrated into its construction than the initial/final position (the latter are located at the margins of the construction). Therefore the middle position is natural; see fn. 23.
    - 1.2. >nat (–, +) / stylistically marked; i.e., absence of stylistic marking is more natural than its presence (in the spirit of Mayerthaler 1981: 15).
  2. The rules of chiastic alignment apply.
  3. The consequences: if the language distinguishes between the middle and the initial/final position of a subordinate clause such that one is stylistically marked and the other is not stylistically marked, then it is the middle position that tends to be stylistically marked and it is the initial/final position that tends not to be stylistically marked. Q.E.D.

## 4. Miscellaneous

This section concerns additional language data connected with coordination and/or subordination. The aim is to illustrate what has not been under consideration in this paper, and thus to preclude some possible misunderstandings.

### 4.1. Slovenian: Order of Clauses

Deductions applying simultaneously to coordination and subordinations were excluded from sections 2 and 3. The following deduction is an example of this kind.

In Slovenian, the order of clauses in a coordination is  $a + b$ , where  $b$  contains the coordinator. The order of clauses in a subordination is either  $a + b$  or  $b + a$ , where  $b$  contains the subordinator. Examples: *Janez poje (a) in Micka pleše (b)* 'Janez is singing and Micka is dancing'; *jasno je (a), da je škoda (b)* 'it is clear that it is a pity'; *da je škoda (b), je jasno (a)* 'idem'.

- The two variants: coordination and subordination. The deduction does NOT proceed in the unnatural environment "complex sentence" because the deduction does not affect the interior of the clauses involved. Chiastic alignment is not indicated.

#### 1. The assumptions of Natural Syntax:

- 1.1.  $>nat(a + b, b + a)$ ; i.e., the complex sentence  $a + b$  is more natural than the complex sentence  $b + a$ .  $B + a$  gives the hearer, at the very outset, the most important signal (the conjunction) for decoding the sentence, and therefore  $b + a$  is in the special interest of the hearer and must be mentioned in slot B of the scale; see fn. 20. A special case of 1.1:
  - 1.1.1.  $>nat(a + b, a + b \text{ and } b + a)$ ; scale 1.1.1 assumes the permitted expanded format  $>nat(A, A + B)$  and is automatically valid because the corresponding basic scale 1.1 has been substantiated.

- 1.2. >nat (coordination, subordination); i.e., a coordination is more natural than a subordination. Coordinations are clearly more common textwise than subordinations. Compare the relevant statistics for English in Biber et al. 1999 (92). Thus, coordinations are natural; see fn. 14.
2. The rules of parallel alignment apply.
3. The consequences: if the language distinguishes between subordination and coordination such that one consists of a + b and the other consists either of a + b or of b + a, then it is coordination that tends to consist of a + b, and it is subordination that tends to consist either of a + b or of b + a. Q.E.D.

#### 4.2. Proto-Indo-European: Finite Verb Accentuation

Also excluded from sections 2 and 3 were cases in which the complex sentence is only a special instance of a broader phenomenon obtaining separately in main clauses and subordinate clauses. Consider the following deduction of this kind.

In Proto-Indo-European, it is usually assumed that the finite verb was accented in subordinate clauses and unaccented in main clauses (Meier-Brügger 2000: 228).

- The two variants: the accented and unaccented finite verb. The deduction is not limited to subordinations, but is valid even for independent main clauses and independent subordinate clauses. Chiastic alignment is not indicated.
1. The assumptions of Natural Syntax:
    - 1.1. >nat (main, subordinate) / clause; i.e., a main clause is more natural than a subordinate clause.<sup>29</sup>
    - 1.2. >nat (-, +) / accented; i.e., unaccented is more natural than accented; see fn. 13.
  2. The rules of parallel alignment apply.

---

<sup>29</sup> See the introductory part of this paper.

3. The consequences: if the language distinguishes between accented and unaccented finite verbs such that one option obtains in main clauses and the other option obtains in subordinate clauses, then it is the accented finite verb that tends to obtain in subordinate clauses and it is the unaccented finite verb that tends to obtain in main clauses. Q.E.D.

### 4.3. Slovenian: Reporting Clauses (Length)

Also excluded from sections 2 and 3 were the deductions discussed in 4.3 and 4.4, which concern reporting clauses. They are chiefly main clauses (not infrequently, although irrelevant for what follows, entering into a complex clause.)

In Slovenian, most reporting clauses (typically referring to the speaker and the act of communication) are shorter (on average) than other main clauses. For instance, *Janez pravi (, da je prepozno)* 'Janez says (that it is too late)'.

- The two variants: reporting and "other" clauses. This deduction does NOT proceed in the unnatural environment "complex sentence" or in any other unnatural environment. Chiastic alignment is not indicated.

The assumptions of Natural Syntax:

- 1.1. >nat (+, -) / reporting clause; i.e., a reporting clause is more natural than other kinds of clauses.<sup>30</sup>
  - 1.2. >nat (short, less short) / clause; i.e., a short clause is more natural than a less short clause; see fn. 13.
2. The rules of parallel alignment apply.

---

<sup>30</sup> As far as we can judge, the reporting clause is the specialized kind of clause most widespread across languages and therefore natural according to the typological criterion, item (h) in the list of axioms.

3. The consequences: if the language distinguishes between reporting and other clauses such that one kind of clause is short (on average) and the other kind of clause is less short (on average), then it is the reporting clause that tends to be short (on average) and it is other clauses that tend to be less short (on average).<sup>31</sup> Q.E.D.

#### 4.4. Slovenian: Reporting Clauses (Reporting Verb)

In Slovenian, most reporting clauses contain a reporting verb in addition to the subject (obligatory in non-pro-drop languages); this is the common variant of reporting clauses. Less frequently, reporting clauses contain some additional material. For instance, *Janez je (hitro) rekel (Micki)* 'Janez said it (to Micka quickly)'.

- The two variants: (within reporting clauses) +/-additional material in addition to (the subject and) the reporting verb. This deduction does NOT proceed in the unnatural environment "complex clause" or in any other unnatural environment. Chiastic alignment is not indicated.

1. The assumptions of Natural Syntax:
  - 1.1. >nat (-, +) / additional material; i.e., the absence of additional material is more natural than the presence of additional material; see fn. 13.
  - 1.2. >nat (more, less) / common; i.e., what is more common is more natural than what is less common.<sup>32</sup>
2. The rules of parallel alignment apply.

---

<sup>31</sup> This deduction predicts that most reporting clauses will be relatively short. See also the deduction discussed in section 4.4.

<sup>32</sup> This is a paraphrase of the frequency criterion, item (d) in the list of axioms.



3. The consequences: if the language distinguishes (within reporting clauses) between the presence and absence of additional material (in addition to any subject and the reporting verb) such that one option is more common and the other option is less common, then it is the absence of additional material that tends to be more common and it is the presence of additional material that tends to be less common. Q.E.D.

#### 4.5. Slovenian: Copula Ellipsis in Subordinate Clauses

The deduction discussed in section 4.5, which concerns subordinate clauses only, was excluded from sections 2 and 3. Chiastic alignment is indicated because the subordinate clause, unlike the main clause, constitutes an unnatural environment.

In Slovenian, in a subordinate clause containing a copula, the copula can be elided if the subject is also elided. For instance, *čeprav je bil Janez bolan* ‘although Janez was ill’ can be shortened to *čeprav bolan* ‘although ill’. The data were suggested to us by the situation in English (Carter and McCarthy 2006: 559–60) and obtain in Slovenian as well.

- The two variants: (in a subordinate clause) optional and obligatory ellipsis of the copula. The deduction proceeds in the unnatural environment “subordinate clause”.
1. The assumptions of Natural Syntax:
    - 1.1. >nat (+, -) / ellipsis of subject + copula; i.e., realized ellipsis of subject + copula is more natural than the non-realized ellipsis of subject + copula.<sup>33</sup> A special case of 1.1:
      - 1.1.1. >nat (+, +/-) / ellipsis of subject + copula; scale 1.1.1 assumes the permitted expanded format >nat (A, A + B) and is automatically valid because the corresponding basic scale 1.1 has been substantiated. (NB. +ellipsis = obligatory ellipsis; +/-ellipsis = optional ellipsis.)
    - 1.2. >nat (+, -) / acceptable; i.e., what is acceptable is more natural than what is not acceptable; see fn. 13.

---

<sup>33</sup> According to the criterion of least effort, item (b) in the list of axioms.

2. The rules of chiastic alignment apply.
3. The consequences: if the language distinguishes (in subordinate clauses) between obligatory and optional ellipsis of subject + copula such that one option is acceptable and the other option is not acceptable, then it is the optional ellipsis of subject + copula that tends to be acceptable and it is the obligatory ellipsis of subject + copula that tends to be unacceptable. Q.E.D.

## 5. Conclusion

The above illustrations testify that Natural Syntax is capable of predicting a significant amount of language situations, given a small number of presuppositions (i.e., the particular description of language data adopted, the choice of variants, the naturalness scales, the alignment rules) and a modest apparatus (namely, the deduction format). As in most of morphology and unlike in generative syntax, the predictions are shallow in the sense that no prediction follows from any other prediction.

The special trait of Natural Syntax is its insistence on comparing two variants (mostly variant constructions) in each deduction. Thus Natural Syntax has something to say only about those areas of a language that happen to display variants. For instance, the fact that Slovenian uses one construction with the cardinal numerals 1–4 and another from 5 on is of interest for Natural Syntax; the fact that English makes no such difference cannot be accounted for in Natural Syntax terms.

Ignoring this limitation, it is still impossible to compare the predictive power of Natural Syntax and, say, generative grammar because the presuppositions of the predictions in one school and the other are so different.

The development of Natural Syntax is to be continued, exploiting as variegated language material as possible. Several Slovenian (and other Slavic) constructions come to mind that would deserve further exploration within Natural Syntax, such as case, aspect, clitic ordering and placement, relative clauses, and infinitive versus finite complements.

## References

- Andersen, Henning. (1972) "Diphthongization". *Language* 48: 11–50.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. (1999) *Longman grammar of spoken and written English*. London: Longman.
- Carter, Ronald and Michael McCarthy. (2006) *Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Collins Cobuild. (1990) *Collins Cobuild English grammar*. London: HarperCollins.
- Cvetko-Orešnik, Varja and Janez Orešnik. (2009) "Natural Syntax: Expressions of future time". Steven Franks, Vrinda Chidambaram, and Brian Joseph, eds. *A linguist's linguist: Studies in South Slavic linguistics in honor of E. Wayles Browne*. Bloomington, IN: Slavica, 167–87.
- Haiman, John. (1978) "Conditionals are topics". *Language* 54: 564–89.
- Havers, Wilhelm. (1931) *Handbuch der erklärenden Syntax*. Heidelberg: Winter.
- Mayerthaler, Willi. (1981) *Morphologische Natürlichkeit*. Wiesbaden: Athenaion.
- Meier-Brügger, Michael. (2000) *Indogermanische Sprachwissenschaft*. 7th ed. with Matthias Fritz and Manfred Mayrhofer. Berlin: de Gruyter.
- Orešnik, Janez. (2007a) "Natural syntax: Negation in English". *Poznań studies in contemporary linguistics* 43: 97–111.
- . (2007b) "Natural syntax: The grammatical person of personal pronouns". *Sprachtypologie und Universalienforschung* 60: 293–313.
- . (2007c) "Natural syntax: English interrogative dependent clauses". *Razprave* 20: 191–208.
- . (2007d) "Natural syntax: English dependent clauses". *Studia anglica posnaniensia* 43: 219–36.
- . (2007e) "Natural syntax: English interrogative main clauses". *Linguistica* 47: 35–48.
- . (2008a) "Natural syntax: English relative clauses". *Poznań studies in contemporary linguistics* 44: 61–101. [Available at <http://versita.metapress.com/content/120709>.]
- . (2008b) "Natural syntax: English reported speech". *Studia anglica posnaniensia* 44: 217–52.

- Orešnik, Janez. (2008c) "Standard French liaison and Natural Syntax". *Linguistica* 48: 33–48.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. (1985) *A comprehensive grammar of the English language*. London: Longman.
- Radford, Andrew. (1997) *Syntactic theory and the structure of English: A Minimalist approach*. Cambridge: Cambridge University Press.
- Toporišič, Jože. (2000) *Slovenska slovnica*. 4th ed. Maribor: Obzorja.

Filozofska fakulteta  
University of Ljubljana  
1001 Ljubljana  
Slovenia  
cvetko@zrc-sazu.si  
Janez.Oresnik@sazu.si

Received: October 2009  
Revised: July 2010