

A Comprehensive Analysis of Constituent Coordination for Grammar Engineering

Agnieszka PATEJUK Adam PRZEPIÓRKOWSKI

Institute of Computer Science, Polish Academy of Sciences,
ul. Jana Kazimierza 5, 01-248 Warszawa, Poland
{aep, adamp}@ipipan.waw.pl

ABSTRACT

This paper provides an explicit, formal analysis of two not only interesting but also frequent coordination phenomena: the coordination of unlike categories and the coordination of distinct grammatical functions possibly belonging to entirely different levels of structure. The proposed account of the former makes it possible to take full advantage of new generation valence dictionaries which encode information about the possibility of such non-standard coordination, creating new possibilities for existing grammar implementations. Furthermore, interactions with complex phenomena such as case assignment are taken into consideration, providing a solution which may be adopted for modelling similar phenomena in other languages. The other variety of coordination addressed in this paper sheds some new light on coordination and how it should be modelled. It demonstrates on the basis of attested data that coordination is subject to fewer constraints than previously assumed and provides an elegant linguistically-motivated solution employing currently available LFG mechanisms.

TITLE AND ABSTRACT IN POLISH

Kompleksowa analiza koordynacji składnikowej na potrzeby inżynierii lingwistycznej

Niniejszy artykuł przedstawia sformalizowaną analizę dwóch zarówno interesujących, jak i częstych zjawisk związanych z koordynacją: koordynację różnych kategorii oraz koordynację różnych funkcji gramatycznych, które mogą również należeć do zupełnie odmiennych poziomów struktury. Zaproponowana analiza pozwala na wykorzystanie w pełni możliwości stwarzanych przez słowniki walencyjne nowej generacji, które zawierają informacje o możliwości wystąpienia niejednorodnej koordynacji, co otwiera zupełnie nowe perspektywy przed istniejącymi już implementacjami gramatyk. Ponadto brane są pod uwagę interakcje ze złożonymi zjawiskami takimi jak nadawanie przypadku, które to rozwiązanie może zostać przystosowane do opisu podobnych zjawisk w innych językach. Drugi rodzaj koordynacji opisany w artykule rzuca nowe światło na koordynację i jej formalny opis: na podstawie autentycznych danych pokazano, że koordynacja podlega znacznie mniejszej liczbie ograniczeń, niż wcześniej sądzono, oraz zaproponowano eleganckie rozwiązanie o solidnych podstawach lingwistycznych, wykorzystujące obecnie dostępne mechanizmy LFG.

KEYWORDS: coordination, unlikes, formal grammar, parsing, Polish, LFG.

KEYWORDS IN POLISH: koordynacja, gramatyka formalna, parsowanie, polski, LFG.

1 Introduction

During the last two decades of the previous millennium grammar engineering was solidly placed within the core of Natural Language Processing (NLP), as also witnessed by the number of papers devoted to this topic in COLING proceedings of 1980s and 1990s. While – in the age of inductive NLP – this theme has almost disappeared from major conferences, grammar engineering efforts, supported by new corpus-based methods, have continued, especially within the HPSG¹ and LFG² communities, where large-scale multilingual grammar engineering initiatives were set up: DELPH-IN (<http://www.delph-in.net/>) and PARGRAM (<http://pargram.b.uib.no/>). There has also been much progress during the last 15 years or so in the theoretical linguistic foundations of these efforts, with flourishing HPSG and LFG conferences and with analyses set up within these frameworks appearing in prestigious linguistic journals.

One of the topics that has received much attention in formal theoretical linguistics is coordination, a phenomenon which is not only theoretically challenging, but also – due to its textual frequency – crucial to grammar engineering. Unfortunately, as coordination remains difficult to describe accurately and exhaustively, this theoretical interest is not fully reflected in existing grammar implementations.

The aim of this paper is to present a comprehensive implementation of constituent coordination, a part of an ongoing effort to develop a wide-coverage LFG parser of Polish (Patejuk and Przepiórkowski, 2012b). Polish is a good test-bed for the task at hand, as it offers a wide range of interactions between coordination on one hand and various agreement, case assignment and valence phenomena on the other.

One aspect of coordination that has remained especially elusive in grammar engineering is the possibility to coordinate elements which are unlike in some sense. Polish offers a much wider range of unlike constituent coordination than has been discussed in NLP, including the so-called lexico-semantic coordination (Kallas 1993; Chaves and Paperno 2007; Gazdik 2010), i.e., the coordination of *very* unlikes, where coordinated items do not even represent the same grammatical function. While such cases go against any comprehensive analysis of coordination we are aware of, they are textually frequent and therefore should be taken into account in any wide-coverage grammar implementation effort.

Note that, for reasons of space, we do not provide here an analysis of non-constituent coordination, as in *give a teacher an apple and a policeman a flower* (Steedman, 2000, p.46).³ This subtype of coordination has an elegant analysis in different versions of categorial grammars, while it remains troublesome for many other grammatical theories. Within LFG, a comprehensive analysis of non-constituent coordination is offered in Maxwell and Manning 1996. In fact, while the LFG analysis seems to cover roughly the same patch as categorial grammar analyses, Maxwell and Manning (1996) claim that it is superior,⁴ as it provides more natural accounts of cases such as *You may call me directly or after 3pm through my secretary* and *She put a lamp on the table, and on the ledge a large antique punchbowl*, which are difficult because coordinated elements are not only non-constituents, but they also differ in the number and order of constituents. The analysis of the current paper may be extended to cover non-constituent

¹Head-driven Phrase Structure Grammar; cf. Pollard and Sag 1987, 1994.

²Lexical-Functional Grammar, cf. Bresnan 1982 and Dalrymple 2001.

³Steedman 2000 calls this phenomenon *argument cluster coordination*.

⁴See Levine 2011, § 2.4, for a refutation of this claim.

coordination along the lines of Maxwell and Manning 1996.

2 Basics: Coordination of ‘like’ categories

The usual intuition behind implementations of coordination is that, wherever some phrase type XP (e.g., a nominal phrase) may occur, a coordination of XP-like elements (i.e., of nominal phrases) may occur instead. In LFG, this intuition is expressed via constituency rules:

- $$(1) \quad \begin{array}{cccc} \text{XP} & \rightarrow & \text{XP} & \text{Conj} & \text{XP} \\ & & \downarrow \in \uparrow & & \downarrow \in \uparrow \end{array}$$

LFG makes a distinction between c(onstituent)-structure and f(unctional)-structure (among other linguistic levels). Annotations of elements of c-structure rules, such as $\downarrow \in \uparrow$, are used to construct corresponding f-structures. In this case, the two conjuncts are elements of the set representing the mother. For example, when used to parse the Polish sentence (2), the above rule constructs the set representation of the subject of the sentence, as in (3).⁵

- (2) *Idą Jan i Marysia.*
 walk.PL Jan.SG and Marysia.SG
 Jan and Marysia walk.’

- $$(3) \quad \left\{ \left[\begin{array}{cc} \text{PRED} & \text{‘JAN’} \\ \text{NUM} & \text{SG} \end{array} \right], \left[\begin{array}{cc} \text{PRED} & \text{‘MARYSIA’} \\ \text{NUM} & \text{SG} \end{array} \right] \right\}$$

This simple example also illustrates the immediate weakness of the basic intuition given above: neither the singular *Jan*, nor the singular *Marysia* alone can directly replace the coordination *Jan i Marysia*, as this would lead to number disagreement with the plural verb. Because of such agreement facts, coordinate structures are represented in LFG as hybrid feature structures, which contain sets such as (3), but may also contain their own features (Dalrymple and Kaplan, 2000); a fuller representation of the subject of (2) would be (4).

- $$(4) \quad \left[\begin{array}{c} \left\{ \left[\begin{array}{cc} \text{PRED} & \text{‘JAN’} \\ \text{NUM} & \text{SG} \end{array} \right], \left[\begin{array}{cc} \text{PRED} & \text{‘MARYSIA’} \\ \text{NUM} & \text{SG} \end{array} \right] \right\} \\ \text{NUM} \quad \text{PL} \end{array} \right]$$

3 Basics: Coordination of ‘unlike’ categories

An influential paper that demonstrated that the assumption of ‘likeness’ of coordinated elements is too strong is Sag et al. 1985, set within GPSG.⁶ Under their analysis, examples (5)–(7) are grammatical because both conjuncts satisfy the underspecified requirements which hold for the syntactic position they occupy.

- (5) That was a rude remark and in very bad taste.
 (6) We walked slowly and with great care.
 (7) Pat became a republican and quite conservative.

⁵The attribute PRED stands for PREDICATE and represents the basic predicate-argument structure of a given element. Here both predicates ‘JAN’ and ‘MARYSIA’ have no arguments.

⁶Generalized Phrase Structure Grammar; cf. Gazdar et al. 1985.

- (8) *Tracy has become a republican and of the opinion that we must place nuclear weapons in Europe.

For example, the verb *BE*, as in (5), only requires that its complement be predicative, [_{PRD} +], and both conjuncts *a rude remark* and *in very bad taste* are [_{PRD} +] (apart from being a noun phrase and a prepositional phrase, respectively). Similarly, the adjunct in (6) is specified as [_{MANNER} +], and again both conjuncts, although categorially unlike, are [_{MANNER} +]. On the other hand, the verb *BECOME* puts stronger restrictions on its complement: not only should it be predicative, but also nominal, [_N +], in the sense that nouns and adjectives are nominal, while prepositions and verbs are not (they are marked as [_N -]). Hence, (7), involving two predicative nominal conjuncts (a noun phrase and an adjectival phrase), is grammatical, while (8), involving a [_N -] conjunct (the prepositional *of the opinion*...) is not.

In order for this analysis to work, the theory needs to implement a notion of feature structure subsumption. In GPSG such a notion is hardwired into the Head Feature Convention (cf. Sag et al. 1985, § 2.4). Attempts have also been made to carry over these subsumption-based insights into HPSG, but they require certain extensions of the formal apparatus assumed within HPSG; two such analyses are proposed in Sag 2002 and Yatabe 2004.

Despite the fact that LFG offers interesting analyses of various aspects of coordination (Maxwell and Manning 1996 on non-constituent coordination; Dalrymple and Kaplan 2000 and Dalrymple et al. 2009 on how coordination influences the representation of features such as person, gender and case; Dalrymple and Nikolaeva 2006, Kuhn and Sadler 2007, and Dalrymple and Hristov 2010 on the interaction between coordination and agreement), there is surprisingly little explicit discussion of coordination of unlike constituents. The assumption seems to be that, at the level of c-structure, there are no categorial constraints on conjuncts, i.e., instead of (1) above, the relevant rule would rather resemble (9) below (cf. Peterson 2004, p. 652).

- (9) $XP \rightarrow YP \text{ Conj } ZP$
 $\downarrow \in \uparrow \qquad \downarrow \in \uparrow$

In LFG, any appearances of categorial ‘likeness’ are a side effect of constraints imposed at the functional level of representation. For example, in Polish, subjects are typically nominative (but see below), a fact which may be expressed via (10).⁷

- (10) $(\uparrow \text{ SUBJ CASE}) =_c \text{ NOM}$

In fact, on standard LFG assumptions, nothing more is required to ensure that *both conjuncts* in (2) are nominative. This is because LFG distinguishes between distributive features, such as *CASE*, and non-distributive features, such as *NUM(BER)* (Dalrymple and Kaplan, 2000). The former distribute over all elements of coordination, so the specification *CASE = NOM* for the outer hybrid feature structure in (11) results in *CASE = NOM* on all conjuncts within the set.

- (11) $\left[\left\{ \begin{bmatrix} \text{PRED} & \text{‘JAN’} \\ \text{NUM} & \text{SG} \\ \text{CASE} & \text{NOM} \end{bmatrix}, \begin{bmatrix} \text{PRED} & \text{‘MARYSIA’} \\ \text{NUM} & \text{SG} \\ \text{CASE} & \text{NOM} \end{bmatrix} \right\} \right]$
 NUM PL

⁷The constraining equation, expressed by ‘=_c’, verifies that the equality holds, rather than assigning a value, as in case of the operator ‘=’.

Using these mechanisms, it is simple to implement in LFG an analysis similar to that of Sag et al. 1985. For example, part of the lexical entry of the verb BE might be: (\uparrow OBJ PRD) = +. Assuming that PRD⁸ is a distributive feature, the requirement that the object be predicative will percolate to all conjuncts, without predetermining their categorial status, and thus accounting for (5).

4 Towards an LFG analysis of the coordination of unlikes

At the end of the paper, Sag et al. (1985) discuss cases of coordination of an NP with a clause (p. 165):

- (12) Pat remembered the appointment and that it was important to be on time.
 (13) That Himmler appointed Heydrich and the implications thereof frightened many observers.

The analysis they propose is far from elegant and it involves, *inter alia*, the assumption that such clauses are in a sense NPs. Apparently, they are attempting to avoid a disjunctive specification of the verb's argument, one that would state that, e.g., the object of the verb REMEMBER is either an NP or a clause (or a coordination of such elements).

When other languages are taken into consideration, it turns out that such disjunctive specifications are inevitable. For example, Kosek 1999, pp. 43–44, cites the following:

- (14) Owinął dziecko w koc i ręcznikiem.
 wrapped baby in blanket.ACC and towel.INST
 'He wrapped the baby in a blanket and in a towel.'
 (15) Nadajesz się do tej pracy i na dyrektora.
 are fit for this job and for manager
 'You are fit for this job and for a manager.'

Prepositional phrases involved in these examples must be headed by the specific prepositions W, DO and NA, and the NP in (14) must be in the instrumental case. It seems highly unlikely that the intuitive disjunctive specifications 'PP[w] \vee NP[inst]' (for (14)) and 'PP[do] \vee PP[na]' (for (15)) could be replaced by non-disjunctive specifications capturing putative commonalities between, for instance, PP[w] and NP[inst] (to the exclusion of other types of phrases).

It turns out that such disjunctive specifications are also problematic for LFG. The obvious way to formalise the requirements of the verb OWINAĆ 'wrap', as in (14), would be:⁹

- (16) (\uparrow OBL CASE) = INST \vee (\uparrow OBL PFORM) = W

Unfortunately, the statement (16) does not have the intended meaning. Because of the distributivity of CASE (and, presumably, PFORM, which stands for 'prepositional form'), the statement, when applied to a coordinate complement, has the following effect: either all conjuncts are instrumental, or all conjuncts have the prepositional form w. Making (any of) these features non-distributive does not solve the problem, as it would only render the relevant constraints inapplicable to individual conjuncts.

⁸Not to be confused with 'PRED'.

⁹We use here ' \vee ' to mark disjunction, instead of the '|' usual in the LFG literature; similarly, we explicitly mark conjunction with '^' instead of leaving it implicit.

There is, however, a solution which does not require extending the formal apparatus of LFG, although it is based on a relatively rarely used LFG mechanism, namely, the so-called off-path constraints (Dalrymple, 2001, p. 148).¹⁰

Off-path constraints make it possible to restrict the path (or, more importantly, its part) used by other statements. For example, while the minimal feature structure satisfying (17) is that of (18), the statement (19), with an off-path constraint added to the attribute A, specifies (20).¹¹

$$\begin{array}{ll}
 (17) \quad (\uparrow A B C) =_c + & (18) \quad \left[\begin{array}{c} A \left[\begin{array}{c} B \left[\begin{array}{c} C \quad + \end{array} \right] \end{array} \right] \\ A \left[\begin{array}{c} B \left[\begin{array}{c} C \quad + \end{array} \right] \end{array} \right] \\ D \quad E \end{array} \right] \\
 (19) \quad (\uparrow \quad A \quad B C) =_c + & (20) \quad \left[\begin{array}{c} A \left[\begin{array}{c} B \left[\begin{array}{c} C \quad + \end{array} \right] \end{array} \right] \\ \leftarrow D \quad E \end{array} \right]
 \end{array}$$

More formally, ‘←’ denotes the f-structure which contains the attribute to which it is attached, while ‘→’ denotes the f-structure which is the value of the attribute to which it is attached. Hence, (21) (i.e., with ‘←’ above replaced by ‘→’) specifies the structure in (22).

$$\begin{array}{ll}
 (21) \quad (\uparrow \quad A \quad B C) =_c + & (22) \quad \left[\begin{array}{c} A \left[\begin{array}{c} B \left[\begin{array}{c} C \quad + \end{array} \right] \end{array} \right] \\ \rightarrow D \quad E \end{array} \right]
 \end{array}$$

The last piece of puzzle necessary to successfully analyse coordination of unlikes in LFG is the standard possibility to require the existence of a certain feature (whatever its value) in a feature structure: for example, (↑ PRED), without any equality sign, requires that the feature PRED be present in the feature structure being described. Given that PRED is a distributive feature, this specification forces all conjuncts to contain PRED (i.e., they must be semantically non-vacuous).

Returning to the failed attempt to formalise a disjunctive constraint in (16), let us note that the problem occurs because the disjunction is understood too early: instead of being interpreted as “for every conjunct, either... or...”, it means: “either for every conjunct... , or for every conjunct...”. What is needed is a means of ‘smuggling’ the disjunction into conjuncts before it is interpreted. Off-path constraints provide a mechanism to achieve this.

The relevant statement, replacing (16), is given below:

$$(23) \quad (\uparrow \text{OBL} \quad \text{PRED} \quad) \\
 (\leftarrow \text{CASE}) =_c \text{INST} \vee (\leftarrow \text{PFORM}) =_c \text{W}$$

(23) ensures that – in Polish – there are no semantically vacuous (expletive) oblique complements, i.e., each such complement has a PRED value. This part of the statement is trivial. The main import of the statement is given in the off-path constraint part: for each such PRED, either the value of the CASE attribute (at the same level as the PRED) is instrumental, or the value of PFORM (again, at the same level) is w. As a result, disjunction is interpreted independently for each conjunct.

¹⁰We would like to thank Mary Dalrymple for suggesting to us that off-path constraints could be used to account for the coordination of unlikes, namely, for Polish unlike category subjects. (The usual disclaimers apply.) Przepiórkowski and Patejuk 2012 presents this analysis – and an alternative account – in more detail.

¹¹Note that off-path constraints are placed *below* the attribute to which they apply.

5 Interactions with case assignment

Polish is a morphologically rich language with 7 grammatical cases. While case values of some arguments are stable and do not change with the syntactic environment (such case values are said to be assigned lexically or inherently), other case values depend on the syntactic context, e.g., the presence of negation or the categorial status of the argument. The syntactic (or structural) case assignment facts may be partially summarised as follows:¹²

- (24) a. **subjects** bearing structural case are in the nominative,
 b. with the exception of numeral phrase subjects, headed by so-called governing numerals (see below), which are in the accusative;
- (25) a. **objects** bearing structural case are in the accusative,
 b. unless they are in the syntactic scope of sentential negation, in which case they are in the genitive (so-called Genitive of Negation, GoN).

It should be possible to model these facts in a straightforward way by case assignment statements such as the following (for (24) above):¹³

- (26) $[(\uparrow \text{SUBJ ACM}) =_c \text{REC} \wedge (\uparrow \text{SUBJ CASE}) =_c \text{ACC}] \vee (\uparrow \text{SUBJ CASE}) =_c \text{NOM}$

In this statement, *ACM* represents *accommodability*, a lexical feature introduced for Polish by Bień and Saloni (1982) to distinguish numeral forms governing the genitive noun (the value of *ACM* in such cases is *REC*) from numeral forms agreeing with the following noun.

Although, again, (26) does not have the intended meaning (its effect is that either all conjuncts are accusative numeral phrases or they are all nominative), the following version does the trick:

- (27) $(\uparrow \text{SUBJ} \quad \text{PRED})$
 $[(\leftarrow \text{ACM}) =_c \text{REC} \wedge (\leftarrow \text{CASE}) =_c \text{ACC}] \vee (\leftarrow \text{CASE}) =_c \text{NOM}$

Such statements are successful in handling coordinate arguments, as in the example below.¹⁴

- (28) [Pan Mirośław] i [czternastu ludzi] pracowało dzień i noc.
 Mr Mirośław.NOM.SG and fourteen.ACC.PL man.GEN.PL worked.3.SG.N day and night
 ‘Mr Mirośław and fourteen men worked night and day.’ (NKJP)

On the other hand, just as in English, some Polish verbs allow for either an NP or a clause (or their coordination) in an argument position (here: subject); a relevant example is given below:

- (29) Jana dziwiło, [że Maria wybiera Piotra], i [jej brak gustu].
 Jan.ACC puzzled.3.SG.N that Maria chooses Piotr and her lack of taste.NOM.SG
 ‘(The fact) that Maria prefers Piotr and her lack of taste puzzled Jan.’
 (Świdziński, 1992, 1993)

For such verbs, the following constraint holds:

¹²See Przepiórkowski 1999 for extensive justification. We limit our considerations to arguments of verbs here.

¹³Where necessary, square brackets are used for the purpose of grouping constraints.

¹⁴“NKJP” marks attested examples found in the National Corpus of Polish (<http://nkjp.p1/>; Przepiórkowski et al. 2010, 2012).

$$(30) \quad (\uparrow \text{SUBJ} \quad \text{PRED} \quad) \\
\begin{array}{l}
[(\leftarrow \text{ACM}) =_c \text{REC} \wedge (\leftarrow \text{CASE}) =_c \text{ACC}] \vee \\
(\leftarrow \text{CASE}) =_c \text{NOM} \vee (\leftarrow \text{COMP-FORM}) =_c \text{ZE}
\end{array}$$

A similar account can be offered for syntactically case-assigned objects, which – for some verbs – may be alternatively realised as clauses. Consider the following examples:

- (31) Doradził mu [wyjazd] i [żeby nie wracał].
 advised him leave.ACC and that NEG come back
 ‘He advised him to leave and not to come back.’ Kallas (1993)
- (32) (Wcale) nie doradził mu [wyjazdu] ani [żeby nie wracał].
 not at all NEG advised him leave.GEN nor that NEG come back
 ‘He did not advise him to leave nor not to come back.’

These examples illustrate the principles of structural case assignment in Polish referring to objects (see (25) above): verbs assign accusative case to nominal objects in the absence of negation, as in (31), while genitive is assigned if negation is present, see (32). This basic principle is formalised in (33).

$$(33) \quad (\uparrow \text{OBJ} \quad \text{PRED} \quad) \\
\begin{array}{l}
[\neg((\text{OBJ} \leftarrow) \text{NEG}) \wedge (\leftarrow \text{CASE}) =_c \text{ACC}] \vee \\
((\text{OBJ} \leftarrow) \text{NEG}) =_c + \wedge (\leftarrow \text{CASE}) =_c \text{GEN}
\end{array}$$

However, in case of verbs such as DORADZIĆ ‘advise’, as in (32), another disjunct must be added, allowing for a clausal object:

$$(34) \quad (\uparrow \text{OBJ} \quad \text{PRED} \quad) \\
\begin{array}{l}
[\neg((\text{OBJ} \leftarrow) \text{NEG}) \wedge (\leftarrow \text{CASE}) =_c \text{ACC}] \vee \\
((\text{OBJ} \leftarrow) \text{NEG}) =_c + \wedge (\leftarrow \text{CASE}) =_c \text{GEN}] \vee \\
(\leftarrow \text{COMP-FORM}) =_c \text{ZEBY}
\end{array}$$

Again, the distributive feature PRED is used as an anchor so that disjunctive off-path constraints are checked against each element of OBJ independently (so, each conjunct under coordination). Note that a combination of outside-in (used so far) and inside-out equations is employed in (33)–(34). The former designate a path leading downwards into the f-structure, the latter make it possible to construct a path leading in the opposite direction, namely to higher f-structures; ‘(OBJ ←)’ creates a path starting from the f-structure designated by ‘←’, leading to the one which contains OBJ. Paths constructed in this way may be used as designators in outside-in equations as in the first disjunct of (33)–(34) which requires that the particular OBJ f-structure be marked for accusative case and that there be no negation in the f-structure containing OBJ (see ‘¬((OBJ ←) NEG)’). The second disjunct uses the same mechanisms to ensure that a given OBJ f-structure be marked for genitive case and that there be negation in the f-structure which contains OBJ (see ‘((OBJ ←) NEG) =_c +’).¹⁵

6 Implications for valence lexicons

Any adequate account of coordination of unlike categories implies a certain organisation of the valence dictionary, i.e., a lexicon recording information about arguments of predicates. Such a

¹⁵This solution may be extended to handle optional GoN under transferred negation in verb chains. This is achieved using functional uncertainty in relevant off-path constraints.

lexicon should make it clear whether arguments in the same position can be coordinated.

Note that it is not enough to know that various categorial kinds of arguments bear the same grammatical function in relation to the predicate. For example, the Polish verb *mówić* ‘say, speak’ takes as its (passivisable) object the following categories, among others: *oratio recta* (direct speech; marked as *or* below), prepositional phrase headed by *o* and requiring a locative phrase (*prepnp* (*o*, *loc*)), subordinate finite clause introduced by the complementiser *że* (*cp* (*że*)), embedded question (*cp* (*intrel*)), etc. While various of these possible objects may be coordinated, apparently an *oratio recta* object cannot be coordinated with any other category. This calls for postulating two different subcategorisation frames: one involving an *oratio recta* object, and another involving other kinds of objects.

Valence dictionaries we are aware of, Polish or otherwise, do not include such ‘coordinability’ information. However, a new valence dictionary, *Walenty*,¹⁶ is being developed for Polish which encodes such information explicitly. In this valence dictionary, any frame allowing for the coordination of unlikes is accompanied by an attested example, usually from the National Corpus of Polish. Figure 1 shows a screenshot of the application assisting in the creation of this dictionary.

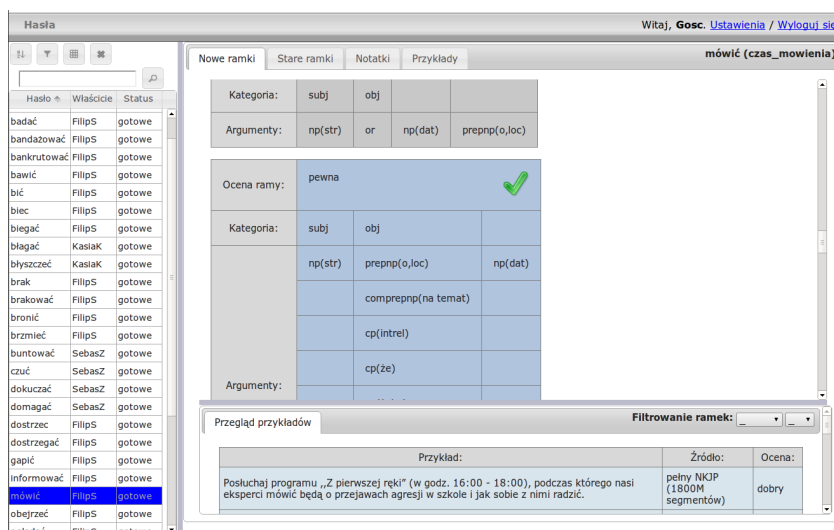


Figure 1: An application for creating *Walenty*, a new valence dictionary of Polish – a screenshot showing subcategorisation frames for *mówić* ‘say, speak’: one involving an *oratio recta* object (near the top), and another involving an object which may be realised as a prepositional phrase, an embedded clause, etc. (in the bottom). An example from the National Corpus of Polish, illustrating the possibility to coordinate *prepnp* (*o*, *loc*) and *cp* (*intrel*) is provided at the very bottom.

¹⁶The preliminary version is available at: <http://clip.ipipan.waw.pl/Walenty>.

7 Coordination of very unlikes

Standard coordination of unlike categories involves elements which bear the same grammatical function. For example, in (12) above, both *the appointment* and *that it was important to be on time* correspond to the direct object of *remembered*. However, in some languages and under certain circumstances it is possible to coordinate elements bearing starkly different grammatical functions, as in Polish (35), involving coordination of the subject and the direct object of the verb *uczył* ‘teach’.

- (35) Kto i kogo będzie uczył?
 who.NOM and who.ACC will teach
 ‘Who will teach whom?’ (NKJP)

To the best of our knowledge, such cases, called in English – after Mel’čuk 1988, p. 40 – *lexico-semantic coordination* or – as in Chaves and Paperno 2007, who provide a preliminary HPSG analysis – *hybrid coordination*, were first described for Russian (and in Russian) in Sannikov 1979, 1980, and for Polish (and in Polish) in Kallas 1993. Similar data are also discussed for French and Hungarian, and given the first LFG account, in Gazdik 2010 (only for *wh*-words, though).

Examination of Polish data reveals that such coordination is more robust than previously described in the literature. First of all, elements which enter lexico-semantic coordination are not only *wh*-phrases, as in (35), but also phrases containing a pronoun expressing an existential quantifier, a universal quantifier or phrases with *n*-words, as in (36), (37) and (38), respectively.

- (36) czy komukolwiek, kiedykolwiek i do czegokolwiek przydał się poradnik
 PART anybody.DAT anytime and for anything.GEN come in handy guide
 ‘Has a(ny) guide ever come in handy to anybody for anything?’ (NKJP)
- (37) Obiecać można wszystko i wszystkim.
 promise may everything.ACC and everyone.DAT
 ‘One may promise everything to everyone.’ (NKJP)
- (38) nikogo i nic nie może tłumaczyć.
 nobody.GEN and nothing.NOM NEG can excuse
 ‘Nothing may excuse anybody.’ (NKJP)

Second, as already evidenced in (38), and contra theoretical assumptions of the HPSG and LFG analyses mentioned above, the coordinated elements may be dependents of different predicates. This is more clearly illustrated in (39), where *skąd* ‘where from’ is an adjunct of the verb *otrzymujemy* ‘receive’ and *jakie* ‘what kind’ is a modifier of the noun *informacje* ‘information’.

- (39) Skąd i jakie otrzymujemy informacje?
 whence and what receive information
 ‘What information and where from do we receive?’ (NKJP)

While coordinated elements may be dependents of different heads, the semantics of (35)–(39) is that of a single clause; for example, (35) may be translated into ‘Who will teach whom?’, and not into ‘Who will teach (somebody) and whom will (somebody) teach?’. Moreover, when the conjunction is *i* ‘and’, it may often be omitted, without any obvious change of meaning:

- (40) Kto kogo będzie uczył?
 who.NOM who.ACC will teach
 ‘Who will teach whom?’

This is reflected in the implemented LFG analysis, where the main c-structure rule handling lexico-semantic coordination is (41), where *type* is a parameter which may assume the following values: WH (question word; cf. (35) and (39)), ANY (existential quantifier; cf. (36)), ALL (universal quantifier; cf. (37)), NEG (n-word; cf. (38)).

$$(41) \text{XPlexsem-mono}_{type} \rightarrow \text{XPextr}_{type} \left[\text{XPextr}_{type} \right]^* \text{CONJ} \text{XPextr}_{type}$$

$\uparrow=\downarrow$ $\uparrow=\downarrow$ $\uparrow=\downarrow$

According to this rule, the f-structure of each extracted phrase of a given type is identified with the f-structure of the mother. This is achieved using the ‘ $\uparrow=\downarrow$ ’ annotation on all conjuncts which treats them as co-heads: f-structure fragments built by particular conjuncts, see (42), are unified in one top-level f-structure, resulting in (43). The latter, once it is unified with the f-structure of the rest of the utterance, renders (44).

$$(42) \text{ a. } \left[\text{ADJ} \left\{ \left[\text{PRED} \text{'WHENCE'} \right] \right\} \right]$$

$$\text{ b. } \left[\text{OBJ} \left[\text{ADJ} \left\{ \left[\text{PRED} \text{'WHAT'} \right] \right\} \right] \right]$$

$$(43) \left[\text{ADJ} \left\{ \left[\text{PRED} \text{'WHENCE'} \right] \right\} \right]$$

$$\left[\text{OBJ} \left[\text{ADJ} \left\{ \left[\text{PRED} \text{'WHAT'} \right] \right\} \right] \right]$$

$$(44) \left[\begin{array}{l} \text{PRED} \text{'RECEIVE'}(\boxed{1}, \boxed{2}) \\ \text{SUBJ } \boxed{1} \left[\text{PRED} \text{'PRO'} \right] \\ \text{OBJ } \boxed{2} \left[\begin{array}{l} \text{PRED} \text{'INFORMATION'} \\ \text{ADJ} \left\{ \left[\text{PRED} \text{'WHAT'} \right] \right\} \end{array} \right] \\ \text{ADJ} \left\{ \left[\text{PRED} \text{'WHENCE'} \right] \right\} \end{array} \right]$$

The complete analysis, which is not given here in detail for lack of space (e.g., the definition of XP_{type} is omitted), also includes the rule (45) which, together with (46)–(47), encodes appropriate locality (or island) constraints on the provenance of coordinated phrases.¹⁷ While (46) defines the set of grammatical functions which may be assigned to a given element, (47) defines the main extraction path of such elements: they may be extracted across any number of infinitival complements (XCOMP) and, below that, across any number of grammatical functions GF, but – crucially – not across finite sentential complements, etc.¹⁸ Since each conjunct resolves its functional annotation independently, it is possible for every conjunct to bear entirely different grammatical function and depend on different heads, as in (42a) vs (42b).

$$(45) \text{XPextr}_{type} \rightarrow \text{XP}_{type} \quad (46) \text{GF} \equiv \{ \text{SUBJ} | \text{OBJ} | \text{OBJ}_{\theta} | \text{OBL} | \text{ADJ} \in \}$$

$(\uparrow \text{PATH GF}^+) = \downarrow$ (47) $\text{PATH} \equiv \text{XCOMP}^*$

Let us finish this section with the observation that there is a subtype of lexico-semantic coordination which is not amenable to the above analysis:

¹⁷The complete analysis is presented in Patejuk and Przepiórkowski 2012a.

¹⁸For *wh*-phrases (47) is extended to allow for extraction from sentential complements: $\text{PATH} \equiv \text{COMP}^* \text{XCOMP}^*$, to account for examples such as *Kto i kogo chcecieś, żeby zaprosi?* ‘Who did you want to invite whom?’, lit.: ‘who and whom you-wanted that invite.FIN’.

- (48) Nie wiadomo było, czy **i* kiedy wróci.
 NEG know was whether and when returns
 ‘It was not clear whether and when (s)he would return.’ (NKJP)

As indicated by the asterisk and the parentheses, this example is ungrammatical once the conjunction *i* is omitted. Moreover, the meaning of the embedded question is biclausal: ‘It was not clear whether (s)he would return, and – if (s)he did – when (s)he would return.’ Such cases, where the first conjunct in the lexico-semantic coordination is the question particle *czy*, are handled by the following c-structure rule:¹⁹

- (49) $XPlexsem\text{-}bi_{wh} \rightarrow PART_{wh} [, XPextr_{wh}]^* CONJ XPextr_{wh}$
 $\downarrow \in \uparrow \qquad \qquad \qquad \downarrow \in \uparrow \qquad \qquad \qquad \downarrow \in \uparrow$

The crucial difference between (41) and (49) is that f-structures of the conjuncts in the latter are not identified with the mother, but become elements of the set representation of the mother; cf. the representation of the conjuncts in (50) and the resulting coordination in (51).

- (50) a. $\left[\begin{array}{l} \text{CLAUSE-TYPE} \\ \text{INT} \end{array} \right]$ (51) $\left\{ \left[\begin{array}{l} \text{CLAUSE-TYPE} \\ \text{INT} \end{array} \right], \left[\begin{array}{l} \text{ADJ} \\ \left\{ \left[\text{PRED} \text{ 'WHEN'} \right] \right\} \end{array} \right] \right\}$
 b. $\left[\begin{array}{l} \text{ADJ} \\ \left\{ \left[\text{PRED} \text{ 'WHEN'} \right] \right\} \right]$

When (51) is unified with the rest of the utterance, a biclausal coordinate structure results, as illustrated in (52):

- (52) $\left\{ \left[\begin{array}{l} \text{PRED} \text{ 'RETURN'}(\overline{\text{I}}) \\ \text{SUBJ } \overline{\text{I}} \left[\begin{array}{l} \text{PRED} \text{ 'PRO'} \end{array} \right] \\ \text{CLAUSE-TYPE} \\ \text{INT} \end{array} \right], \left[\begin{array}{l} \text{PRED} \text{ 'RETURN'}(\overline{\text{I}}) \\ \text{SUBJ } \overline{\text{I}} \\ \text{ADJ} \left\{ \left[\text{PRED} \text{ 'WHEN'} \right] \right\} \end{array} \right] \right\}$

8 Towards evaluation

As mentioned in the Introduction, coordination is textually frequent, so any parser should be able to handle it. But exactly how frequent is it?

There are two balanced subcorpora within the National Corpus of Polish: one – let us call it NKJP1M – is manually annotated and contains around 1 million words, another one – NKJP250M – contains 250 million words tagged automatically. As shown in Table 1, depending on which subcorpus is used to gather statistics, between 37.5% and 38.7% of all sentences involve coordination, which amply justifies the claim advanced in the Introduction.

It is considerably more difficult to estimate how many of these coordinate structures involve unlike categories. Instead, let us concentrate on cases of lexico-semantic coordination, which are easier to find automatically. Of these, undoubtedly the most frequent – and the ones most readily noticed in the linguistic literature – are cases of coordination of *wh*-phrases, as in (35) and (39) above. This subtype of lexico-semantic coordination is especially important in NLP, for example, in question answering, where Polish users will naturally use such constructions.

¹⁹There are examples where the question particle is the last conjunct. Handling these requires simple modifications.

corpus	# of sentences	# of conjunctions	# of sentences containing conjunctions	% of sentences containing conjunctions
NKJP1M	85 663	44 841	32 147	37.5
NKJP250M	18 625 185	10 455 657	7 210 648	38.7

Table 1: The number of conjunctions and the percentage of sentences containing conjunctions in balanced subcorpora of the National Corpus of Polish.

In NKJP250M, there are 272 results of the query ‘Kto [orth="i|lub|oraz|albo"] [orth!=co/i]’, where *kto* means ‘who’ (nominative), *i*, *lub*, *oraz* and *albo* are the most frequent conjunctions, and *co* means ‘what’ (nominative or accusative); the last part of this query ensures that structures like *Kto i co*, i.e., possibly involving two nominative phrases, are not among the reported results.²⁰ Compared to over 18 million sentences of NKJP250M, this number seems to be infinitesimal, but note that this is only one subtype of *wh*-lexico-semantic coordination; *Gdzie i...* ‘where and...’ occurs 157 times, *Komu i...* ‘who.DAT and’ occurs 47 times, etc.²¹ These numbers are sufficiently large to show that this is not a marginal construction in Polish.

The LFG grammar of Polish (Patejuk and Przepiórkowski, 2012b), which includes the analysis of coordination sketched above, is implemented in the XLE system developed at PARC.²² It is based on two previous implemented grammars of Polish: its *c*-structure is based on a DCG (Warren and Pereira, 1980) grammar used by the parser Świgr (Świdziński, 1992; Woliński, 2004, 2005), while its *f*-structure is inspired by an HPSG grammar (Przepiórkowski et al., 2002; Marciniak et al., 2003). The quality of the grammar is ensured in a two-fold way: using treebank testing, but also verifying the linguistic coverage against manually constructed test suites.

The former takes the form of reparsing *Składnica* (Woliński et al., 2011), a treebank of Polish containing parses for sentences extracted from the manually annotated subcorpus of the National Corpus of Polish and parsed using Świgr. The treebank coverage of the LFG grammar amounts to 90%. Unfortunately, the current version of *Składnica* contains sentences which were relatively unproblematic for Świgr and for human annotators, i.e., it is skewed towards simple and short sentences. In particular, coordination is underrepresented in general (1869 out of 8227 sentences, or 22.7%, contain conjunctions), and no cases of lexico-semantic coordination were found by the authors.²³ The complete NKJP1M, from which *Składnica* draws its sentences, does contain instances of lexico-semantic coordination, e.g., *Kto i dlaczego boi się prywatyzacji?* ‘Who and why is afraid of privatisation?’.

While *Składnica* currently contains good parses for 8227 sentences, it contains many more sentences for which human annotators could not identify a good parse among the trees

²⁰The query syntax is described in more detail at <http://nkjp.pl/poliqarp/help/en.html>.

²¹Also note that such queries only find cases of coordination of two *wh*-phrases, while – as expected – it is possible to coordinate more of them, as in the attested (NKJP) *Gdzie, jak i za ile będą się bawić*, lit. ‘Where, how and for how much will they have fun’. Since the first word is capitalised, such queries also do not target embedded questions.

²²Maxwell and Kaplan 1996; <http://www2.parc.com/is1/groups/nl1t/xle/>

²³The treebank search engine available at <http://nlp.ipipan.waw.pl:8000/ui.xhtml> has been employed, and queries such as ‘[base = /kto/ & orth = /K.*/]’ were given, which should find any capitalised form of the pronoun *kto* ‘who’. Out of 39 sentences beginning with a form of *kto*, 57 beginning with a form of *co* ‘what’, 22 beginning with *kiedy* ‘when’, 21 beginning with *dlaczego* ‘why’ and 5 beginning with *gdzie* ‘where’, none contained lexico-semantic coordination.

generated by the DCG grammar. It is hoped that extensions related to coordination, including the two described above, make it possible to parse sentences which have been rejected so far due to the limitations of the previous grammar and its valence dictionary, which did not take coordination of unlikes into consideration.

The other method of testing relies on constructed sentences: currently there are over 1200 items which provide a means of comprehensive grammar testing, making it possible to test a wide range of phenomena, ensuring proper handling of fundamental issues, but also giving an opportunity to test very sophisticated phenomena where complex interactions between various areas of the grammar are involved. Also, while treebank testing is limited to positive examples exclusively, there are numerous negative examples among constructed test suite items. This test suite is currently being extended with examples of unlike and lexico-semantic coordination, so no quantitative results can be cited here (they should be available by the time of COLING 2012), but the grammar correctly parses any sentences involving such difficult instances of coordination that we have come across.

Conclusion

While currently a niche activity in NLP, grammar engineering is actively pursued and finds high-profile applications, such as the use of the English LFG grammar in Microsoft's Bing search engine.²⁴ Since coordination is textually very frequent, any self-respecting grammar should allow for a comprehensive treatment of this phenomenon. The analysis presented above, taking into account not only run of the mill cases of coordination, but also coordination of unlike and very unlike constituents, is implemented within a large LFG grammar of Polish.

It must be stressed, however, that – although the details of the interaction of coordination with case assignment and other phenomena vary from language to language – the general mechanisms described in sections 4–7 seem to be applicable to any language. In particular, coordination of unlikes has been by now reported for many languages, syntactic case assignment mechanisms which interact with coordination can be observed in Slavic, Baltic and Finno-Ugric languages (to limit ourselves to European languages), and lexico-semantic coordination has been described for French, Hungarian, Romanian and a number of Slavic languages.

Due to space constraints, many aspects of the implementation of coordination in the LFG grammar of Polish have been omitted here. Probably the most important is agreement between the verb (or, more generally, a predicate) and its coordinated subject, but also the interesting phenomenon of single conjunct agreement. However, unlike coordination of unlikes, these phenomena already have established accounts in the LFG literature, e.g., Dalrymple and Nikolaeva 2006, Kuhn and Sadler 2007 and Dalrymple et al. 2009, and their standard analyses are implemented in the current grammar. The first stable version of the grammar will be made publicly available by the end of January 2013 at <http://zil.ipipan.waw.pl/LFG>.

Acknowledgements

This research is supported by the POIG.01.01.02-14-013/09 project which is co-financed by the European Union under the European Regional Development Fund. The publication of the resulting implemented grammar is supported by the CESAR project (European project CIP ICT-PSP 271022, part of META-NET).

²⁴See <http://www.essex.ac.uk/linguistics/external/LFG/Bulletin/Jul09.html> and Ahmed and Hauti 2011, p. 9.

References

- Ahmed, T. and Hautli, A. (2011). A first approach towards an Urdu WordNet. *Linguistics and Literature Review*, 1(1):1–14.
- Bień, J. S. and Saloni, Z. (1982). Pojęcie wyrazu morfologicznego i jego zastosowanie do opisu fleksji polskiej (wersja wstępna). *Prace Filologiczne*, XXXI:31–45.
- Bresnan, J., editor (1982). *The Mental Representation of Grammatical Relations*. MIT Press Series on Cognitive Theory and Mental Representation. The MIT Press, Cambridge, MA.
- Butt, M. and King, T. H., editors (1996). *The Proceedings of the LFG'96 Conference*, Stanford, CA. CSLI Publications.
- Butt, M. and King, T. H., editors (2010). *The Proceedings of the LFG'10 Conference*, Stanford, CA. CSLI Publications.
- Butt, M. and King, T. H., editors (2012). *The Proceedings of the LFG'12 Conference*, Stanford, CA. CSLI Publications.
- Chaves, R. P. and Paperno, D. (2007). On the Russian hybrid coordination construction. In Müller, S., editor, *The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, pages 46–64, Stanford. CSLI Publications.
- Dalrymple, M. (2001). *Lexical-Functional Grammar*. Academic Press.
- Dalrymple, M. and Hristov, B. (2010). Agreement Patterns and Coordination in Lexical Functional Grammar. In (Butt and King, 2010), pages 186–206.
- Dalrymple, M. and Kaplan, R. M. (2000). Feature indeterminacy and feature resolution. *Language*, 76(4):759–798.
- Dalrymple, M., King, T. H., and Sadler, L. (2009). Indeterminacy by underspecification. *Journal of Linguistics*, 45:31–68.
- Dalrymple, M. and Nikolaeva, I. (2006). Syntax of natural and accidental coordination: Evidence from agreement. *Language*, 82:824–849.
- Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Blackwell / Harvard University Press, Cambridge / Cambridge, MA.
- Gazdik, A. (2010). Multiple Questions in French and Hungarian: An LFG Account. In (Butt and King, 2010), pages 249–269.
- Kallas, K. (1993). *Składnia współczesnych polskich konstrukcji współrzędnych*. Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Kosek, I. (1999). *Przyczasownikowe frazy przyimkowo-nominalne w zdaniach współczesnego języka polskiego*. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.
- Kuhn, J. and Sadler, L. (2007). Single conjunct agreement and the formal treatment of coordination in LFG. In Butt, M. and King, T. H., editors, *The Proceedings of the LFG'07 Conference*, pages 302–322, University of Stanford, California, USA.

- Levine, R. D. (2011). Linearization and its discontents. In Müller, S., editor, *Proceedings of the HPSG 2011 Conference*, pages 126–146, Stanford, CA. CSLI Publications.
- Marciniak, M., Mykowiecka, A., Przepiórkowski, A., and Kupś, A. (2003). An HPSG-annotated test suite for Polish. In Abeillé, A., editor, *Treebanks: Building and Using Parsed Corpora*, volume 20 of *Text, Speech and Language Technology*, pages 129–146. Kluwer, Dordrecht.
- Maxwell, III, J. T. and Kaplan, R. M. (1996). An efficient parser for LFG. In (Butt and King, 1996).
- Maxwell, III, J. T. and Manning, C. D. (1996). A theory of non-constituent coordination based on finite-state rules. In (Butt and King, 1996).
- Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, NY.
- Patejuk, A. and Przepiórkowski, A. (2012a). Lexico-semantic coordination in Polish. In (Butt and King, 2012).
- Patejuk, A. and Przepiórkowski, A. (2012b). Towards an LFG parser for Polish: An exercise in parasitic grammar development. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012*, pages 3849–3852, Istanbul, Turkey. ELRA.
- Peterson, P. G. (2004). Coordination: Consequences of a lexical-functional account. *Natural Language and Linguistic Theory*, 22:643–679.
- Pollard, C. and Sag, I. A. (1987). *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. Number 13 in CSLI Lecture Notes. CSLI Publications, Stanford, CA.
- Pollard, C. and Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. Chicago University Press / CSLI Publications, Chicago, IL.
- Przepiórkowski, A. (1999). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph.D. dissertation, Universität Tübingen, Germany.
- Przepiórkowski, A., Bańko, M., Górski, R. L., and Lewandowska-Tomaszczyk, B., editors (2012). *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw.
- Przepiórkowski, A., Górski, R. L., Łaziński, M., and Pęzik, P. (2010). Recent developments in the National Corpus of Polish. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC 2010*, Valletta, Malta. ELRA.
- Przepiórkowski, A., Kupś, A., Marciniak, M., and Mykowiecka, A. (2002). *Formalny opis języka polskiego: Teoria i implementacja*. Akademicka Oficyna Wydawnicza EXIT, Warsaw.
- Przepiórkowski, A. and Patejuk, A. (2012). On case assignment and the coordination of unlikes: The limits of distributive features. In (Butt and King, 2012).
- Sag, I. A. (2002). Coordination and underspecification. In Kim, J.-B. and Wechsler, S., editors, *Proceedings of the HPSG 2002 Conference*, pages 267–291. CSLI Publications, Stanford, CA.
- Sag, I. A., Gazdar, G., Wasow, T., and Weisler, S. (1985). Coordination and how to distinguish categories. *Natural Language and Linguistic Theory*, 3:117–171.

- Sannikov, V. Z. (1979). Sočinitel'nye i sravnitel'nye konstrukcii: ix blizost', ix sintaksičeskoe predstavlenie I. *WSA*, 4:413–432.
- Sannikov, V. Z. (1980). Sočinitel'nye i sravnitel'nye konstrukcii: ix blizost', ix sintaksičeskoe predstavlenie II. *WSA*, 5:211–242.
- Steedman, M. (2000). *The Syntactic Process*. The MIT Press, Cambridge, MA.
- Świdziński, M. (1992). *Gramatyka formalna języka polskiego*, volume 349 of *Rozprawy Uniwersytetu Warszawskiego*. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw.
- Świdziński, M. (1993). Dalsze kłopoty z bezokolicznikiem. In Sambor, J., Linde-Usiekniewicz, J., and Huszcza, R., editors, *Językoznawstwo synchroniczne i diachroniczne*, pages 303–314. Wydawnictwa Uniwersytetu Warszawskiego, Warsaw.
- Warren, D. H. D. and Pereira, F. C. N. (1980). Definite clause grammars for language analysis — a survey of the formalism and a comparison with augmented transition networks. *Artificial Intelligence*, 13:231–278.
- Woliński, M. (2004). *Komputerowa weryfikacja gramatyki Świdzińskiego*. Ph.D. dissertation, Instytut Podstaw Informatyki, Polska Akademia Nauk, Warsaw.
- Woliński, M. (2005). An efficient implementation of a large grammar of Polish. *Archives of Control Sciences*, 15(3):251–258.
- Woliński, M., Głowińska, K., and Świdziński, M. (2011). A preliminary version of Składnica—a treebank of Polish. In Vetulani, Z., editor, *Proceedings of the 5th Language & Technology Conference*, pages 299–303, Poznań.
- Yatabe, S. (2004). A comprehensive theory of coordination of unlikes. In Müller, S., editor, *Proceedings of the HPSG 2004 Conference*, pages 335–355, Stanford, CA. CSLI Publications.

