ARG-ST on Phrases: Evidence from Polish*

Adam Przepiórkowski
Ohio State University and Polish Academy of Sciences

1 Introduction

One of the major controversies in present-day HPSG is whether the information about a word's argument structure should also be available on this word's phrasal projections. Some works assume that ARG-ST is present on words only; this is the claim of, e.g., Pollard and Sag (1994), Miller and Sag (1997), Abeillé *et al.* (1998), and Bouma *et al.* (1999). The reason for this assumption is that it leads to more restrictive grammars: with this restriction, words cannot select their arguments on the basis of the argument structure of these arguments' heads (e.g., there seems to be no language in which a verb selects exactly VPs with an NP[*dat*] argument). On the other hand, various other works assume the presence of the complete information about a word's argument structure on this word's phrasal projections. This is the stance of, e.g., Grover (1995) (to formulate a fully nonconfigurational binding theory), Frank (1994) (to deal with verb second in German), Frank and Reyle (1995) (to account for the interactions between scope and word order in German), Calcagno and Pollard (1997) and Abeillé and Godard (2000, n. 9) (to analyze French causatives), Baxter (1999) (in an account of purpose infinitives in English), and Meurers (1999) (to deal with case assignment in German verb clusters).

Works assuming ARG-ST on phrases do not usually *argue* extensively for maintaining this assumption, i.e., they show that having ARG-ST on phrases makes the respective analysis possible or easier, without extensively arguing that giving up this assumption makes the analysis *im*possible.

The general aim of this paper is to show that, whether ARG-ST is present on phrases or not and, if it is present, what are the condition on this presence, is ultimately an empirical issue, one that perhaps should be resolved differently for different languages.[1]

To this end, I will examine two rather unusual constructions in Polish, not successfully analyzed in formal linguistics so far, and argue that they do call for the presence of ARG-ST on phrases. However, this is not taken as evidence that *all* Polish phrases must bear the ARG-ST of their heads; on the contrary, I show that these two constructions share a rather special property, i.e., they involve

* For useful comments and discussions on topics closely related to this paper I am grateful to Peter Culicover, Tibor Kiss, Carl Pollard, Ivan Sag, Gert Webelhuth and the Synners at the OSU Linguistics Department, as well as the audiences of the HPSG-2000 conference, the Formal Approaches to Slavic Linguistics 2000 Meeting, and the University of North Carolina Linguistics Colloquium.

[1] These issues are touched upon in Abeillé and Godard 2000, n. 9, where it is suggested that ARG-ST should be present on *word*s and a small subset of *phrase*s, namely, *lite phrase*s.

Proceedings of the 7th International HPSG Conference, UC Berkeley (22-23 July, 2000).
Dan Flickinger and Andreas Kathol (Editors).
2001. CSLI Publications. http://csli-publications.stanford.edu/

semantically vacuous words, in a sense made precise below. I provide an analysis which ties the presence of ARG-ST on phrases to the semantic emptiness of these phrases' heads. Because of the rarity of such constructions, the resulting grammar is not less restrictive than, say, a grammar which allows a verb to subcategorize for a lexical argument (and, hence, have access to this argument's ARG-ST), a possibility often taken advantage of in HPSG analyses of complex predicates in various languages.

2 Raising across Prepositions

Consider the raising constructions in (1)–(2).[2]

(1) Uważałem go [za szczerego / za studenta].
I considered him$_{acc}$ as sincere$_{acc}$ / as student$_{acc}$
'I considered him to be sincere / to be a student.'

(2) Miałem go [za szczerego / za studenta].
I had him$_{acc}$ as sincere$_{acc}$ / as student$_{acc}$
'I took him to be sincere / to be a student.'

The verbs *uważać* and *mieć* in (1)–(2) are raising verbs: the subject of the predicative argument of *za* (i.e., the subject of *szczerego* / *studenta* in (1)–(2)) must be structure-shared with the object of the verb (i.e., with *go* in (1)–(2)).

At first sight constructions like these do not seem to be particularly problematic: they could be analyzed as involving raising of the subject of the AP/NP predicate (*szczerego* / *studenta* in (1)–(2)) to the subject position of the preposition *za*, and then to the object position of the verb (*uważałem* / *miałem*). On this analysis, *za* would be a 2-argument raising preposition, as in (3), and *uważać* / *mieć* would be rather straightforward raising verbs with lexical entries as in (4).

(3) Incorrect (schematic) lexical entry for *za*:

$$\begin{bmatrix} \textit{word} \\ \text{PHON za} \\ \text{SYNSEM}|\text{LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD}|\text{PFORM za} \\ \text{ARG-ST } \langle \boxed{0}, \begin{bmatrix} \text{PRD } + \\ \text{SUBJ } \langle \boxed{0} \rangle \\ \text{CONT } \boxed{2} \end{bmatrix} \rangle \end{bmatrix} \\ \text{CONT } \boxed{2} \end{bmatrix} \end{bmatrix}$$

[2] These constructions are analyzed in greater detail in Przepiórkowski 2000a. It has long been noted that they are raising constructions, e.g., Tajsner 1990.

(4) Incorrect (schematic) lexical entry for *uważać* / *mieć*:

$$\begin{bmatrix} \textit{word} \\ \text{PHON}\ \text{uważać / mieć} \\ \text{SYNSEM|LOC} \begin{bmatrix} \text{CAT|ARG-ST}\ \langle \text{NP}_{\boxed{1}}, \boxed{0}, \text{PP} \begin{bmatrix} \text{PFORM}\ \text{za} \\ \text{SUBJ}\ \langle \boxed{0} \rangle \\ \text{CONT}\ \boxed{2} \end{bmatrix} \rangle \\ \text{CONT} \begin{bmatrix} \textit{consider} \\ \text{CONSIDERER}\ \boxed{1} \\ \text{SOA-ARG}\ \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

These lexical entries would lead to the structure of (1) as in Figure 1 below.

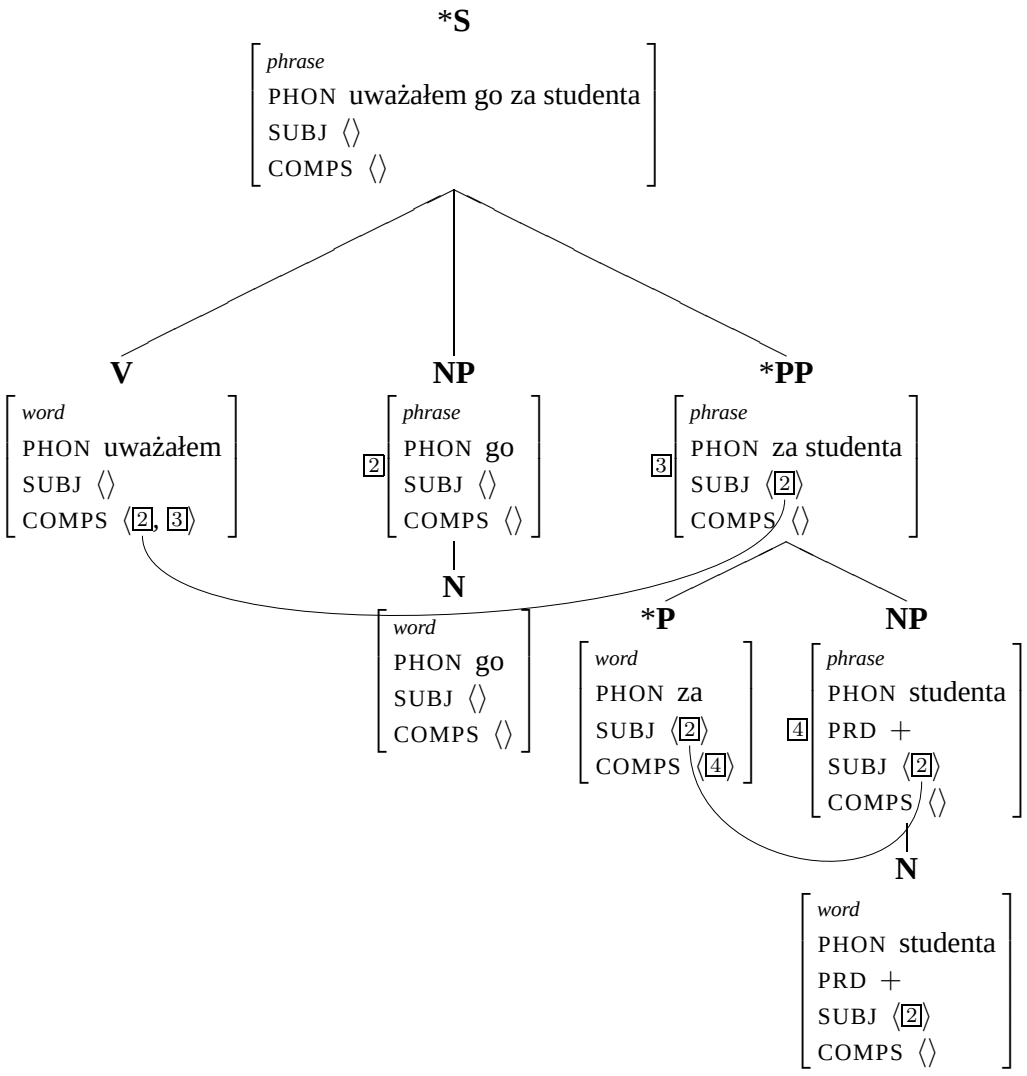


Figure 1: *Not* the structure of (1)

However, it turns out that this analysis cannot be maintained for reasons to do with binding. In Polish, as in many other languages, anaphors may be bound only by subjects (Reinders-Machowska,

1991; Marciniak, 1999). This is illustrated in (5), where the object *jej* cannot be a binder of the anaphor *sobie*.

(5) Mówiłem jej_i o sobie_{*i} samej / o niej_i samej.
I talked her about Self Emph_{fem} / about her Emph_{fem}
'I talked to her about herself.'

On the other hand, in case of 2-argument (predicative) prepositions, such apparent binding by object is possible:

(6) Nie można przecież położyć książki_i na $\text{sobie}_{?i}$ samej / na $\text{niej}_{??i}$ samej.
not may but lay book_{fem} on Self Emph_{fem} / on her Emph_{fem}
'But it is impossible to lay a book on itself.'

Of course, this binding by object in (6) is only apparent: this sentence is acceptable simply because the anaphor is bound by the unrealized subject of the 2-argument preposition, which is co-indexed with (in fact, raised to) the object of the verb. Similar binding differences between between 1-argument prepositions and 2-argument prepositions in English are discussed in Wechsler 1997.

Now, given this difference between 1-argument prepositions and 2-argument prepositions, *za* as used in (1) clearly patterns with 1-argument prepositions, such as *o* in (5).

(7) (Nie pomyliłem się,) uważałem go_i za siebie_{*i} samego / za niego_i samego.
I wasn't confused, $\text{considered}_{1st,sg,masc}$ him_{acc} as Self Emph_{masc} / as him Emph$_{masc}$
'(I wasn't confused,) I really considered him as himself.'

Despite pragmatic weirdness of (7), the coindexation between *go* and the pronoun *niego* is interpretable, while the coindexation between *go* and the anaphor *siebie* is completely uninterpretable.

This means that the preposition *za* in (1)–(2) is really a 1-argument preposition, as shown in (8).

(8) Correct (schematic) lexical entry for *za*:

$$\left[\begin{array}{l} \textit{word} \\ \text{PHON za} \\ \text{SYNSEM}|\text{LOC} \left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{l} \text{HEAD}|\text{PFORM za} \\ \text{ARG-ST} \left\langle \left[\begin{array}{l} \text{PRD} + \\ \text{SUBJ} \langle \boxed{0} \rangle \\ \text{CONT } \boxed{2} \end{array}\right] \right\rangle \end{array}\right] \\ \text{CONT } \boxed{2} \end{array}\right] \end{array}\right]$$

But if this is so, then the phrase PP[*za*] is a fully saturated phrase, with no information about the raised subject (0) of the predicate available in its *synsem*. This, in turn, means that there is no way of specifying lexical entries of verbs such as *uważać* and *mieć* in (1)–(2) that would make them raising verbs; such lexical entries would have no access to the argument of the preposition.[3]

This is the first example of an environment in which an item needs access to (already realized) arguments of its arguments: verbs such as *uważać* and *mieć* need to be able to look at the predicative argument of its PP[*za*] argument, even though that predicative argument is absent from the VALENCE (SUBJ or COMPS) features of that PP[*za*]. The most straightforward way of dealing with such environments is to allow the ARG-ST of the preposition *za* to percolate to the PP[*za*]; once this is allowed, raising verbs *uważać* and *mieć* (and other similar verbs) may have lexical entries as in (9), which lead to structures as in Figure 2.

(9) Correct (schematic) lexical entry for *uważać* / *mieć*:

$$\left[\begin{array}{l} \textit{word} \\ \text{PHON}\ \text{uważać / mieć} \\ \text{SYNSEM|LOC} \left[\begin{array}{l} \text{CAT|ARG-ST}\ \langle \text{NP}_{\boxed{1}}, \boxed{0}, \text{PP}\left[\begin{array}{l}\text{PFORM}\ \text{za} \\ \text{ARG-ST}\ \langle\left[\begin{array}{l}\text{SUBJ}\ \langle\boxed{0}\rangle \\ \text{CONT}\ \boxed{2}\end{array}\right]\rangle\end{array}\right]\rangle \\ \text{CONT}\left[\begin{array}{l}\textit{consider} \\ \text{CONSIDERER}\ \boxed{1} \\ \text{SOA-ARG}\ \boxed{2}\end{array}\right]\end{array}\right]\end{array}\right]$$

More on (Im)Possible Alternatives

We noted above that *za* as used in (1)–(2) is a 1-argument preposition, and we cited data involving binding as supporting this conjecture. In fact, there is another argument against the analysis of *za* as a 2-argument preposition: assuming, that 1-argument prepositions are 'case-marking' prepositions and 2-argument prepositions are predicative prepositions, such an analysis would predict that the *za* is a predicative preposition, i.e., that the PP[*za*] is a predicative phrase.

It is, however, relatively clear that PP[*za*] cannot be a predicative phrase: if it were, it should be able to occur in environments which allow any predicative phrases, such as complements of copula or exclamatives. (10) shows that this prediction is false:

(10) Janek jest szczery / prezydentem / w domu… / *za szczerego.
John$_{nom}$ is sincere$_{nom}$ / president$_{ins}$ / at home… / as sincere
'John is sincere / the president / at home… / *as sincere.'

[3] In fact, technically, lexical entries having access to arguments of their arguments could be posited in RSRL (Richter *et al.*, 1999; Richter, 2000), but such entries would introduce as much non-restrictiveness as ARG-ST on *phrases*, and — additionally — they would involve relations and in this sense would be more complex than the entries proposed below.

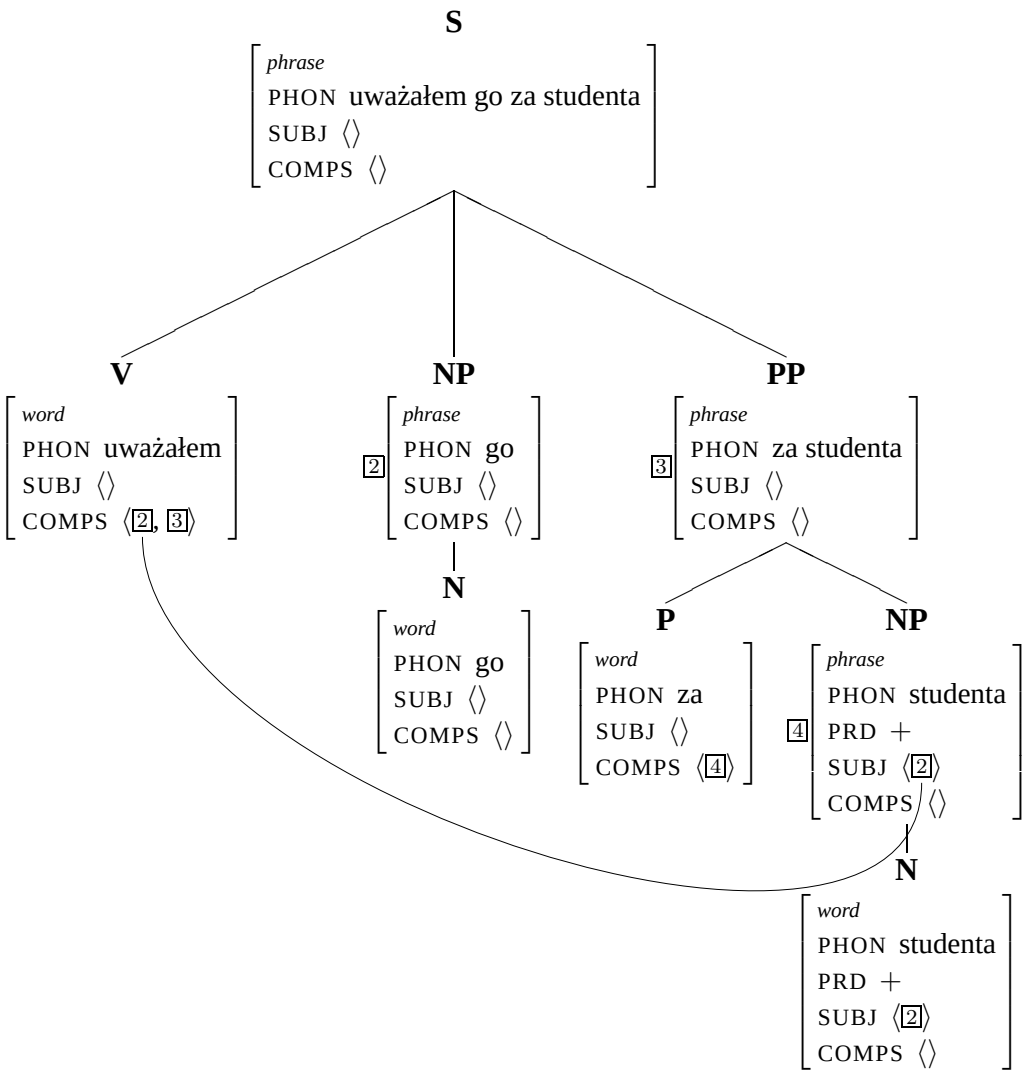


Figure 2: The structure of (1)

The copula *jest* in (10) could perhaps impose an idiosyncratic constraint to the effect that its complement cannot be marked with *za*, but such a constraint would violate the otherwise overwhelming generalization that the copula may combine with any predicative complement and, moreover, it is not clear that such a constraint could be imposed in case of (11), where there is no overt copula and no obvious reason to posit a phonologically empty one.

(11) Janek szczery! / Wałęsa prezydentem! / Krokodyl w klatce! / Obiad o dziesiątej! / *Janek za szczerego! (Też pomysł!)
John sincere$_{nom}$ / Wałęsa president$_{ins}$ / crocodile in cage / dinner at ten / John as sincere also idea

‘John (being) sincere! / Wałęsa (as) the president! / A crocodile in a cage! / Dinner at 10! *John as sincere! (What an idea!)

Another obvious but only apparent alternative to the account involving ARG-ST on phrases would be to say that *za* in (1)–(2) is not a preposition at all, but rather a marker (cf. Pollard and Sag, 1987, p. 65). Since the AP/NP combining with *za* is a predicative phrase, according to this analysis, the PP[*za*] is really a predicative AP/NP[MARKING *za*].

An immediate problem faced by this account would be the same as the second problem with the "*za* as a 2-argument preposition" approach: if *za* is just a marker, then AP/NP[MARKING *za*] phrases are predicative and should be able to appear in predicative positions such as those in (10)–(11).

An even more serious problem is that markers are not supposed to be able to assign case, while in (1)–(2) *za* clearly assigns the accusative to the predicative AP/NP. Why *clearly*? One possible alternative would be that the accusative case on the AP/NP is the result of case agreement with the object of the verb. However, this alternative is not available for a number of reasons discussed in Przepiórkowski 2000a. One such reason is that various processes which trigger the change of case on the object should also trigger a similar case shift on the presumably agreeing predicative AP/NP. As examples (12), involving the genitive of negation, and (13), involving passivization, show, this prediction is false.

(12) Nie uważałem jej za szczerą / *szczerej / studentkę / *studentki.
not I considered her$_{gen}$ as sincere$_{acc}$ / sincere$_{gen}$ / student$_{acc}$ / student$_{gen}$
'I didn't consider her to be sincere / to be a student.'

(13) Ona jest uważana za studentkę / *studentka / zdolną / *zdolna.
she$_{nom}$ is considered as student$_{acc}$ / student$_{nom}$ / gifted$_{acc}$ / gifted$_{nom}$
'She is considered (as) a student / gifted.'

Example (12) falsifies also another putative analysis regarding the accusative case of the predicative AP/NP, namely, the analysis treating this accusative case as received directly from the governing verb. If *szczerą* / *studentkę* were assigned the case directly by the verb (or by general case assignment principles, as in Przepiórkowski 1999), then they should be able to occur in the genitive case when the verb is negated, contrary to (12).

Thus, in summary, no analysis of examples such as (1)–(2) seems available that would not require the presence of the ARG-ST of the preposition *za* on the maximal projection of that preposition.

3 Predicative Modification of Numeral Phrases

Another argument for the presence of ARG-ST on at least some phrases in Polish concerns case agreement with numeral phrases.[4]

In Polish, as in many other languages, predicative adjectives usually agree in case (and also in number and gender, but this will not concern us here) with the NP they are predicated of, e.g.,

[4] The issue of predicative case agreement in Polish is considered in much greater empirical and theoretical detail in Przepiórkowski 1999. An attempt at an explanation of the quirky behavior of numeral phrases considered here is made in Przepiórkowski 2000b.

(14)–(16);[5] in (14) both the subject NP and the predicative AP are nominative, while in (15) the object NP and the predicate are both accusative.

(14) Janek jest miły.
John$_{nom}$ is nice$_{nom}$
'John is nice.'

(15) Pamiętam go miłego.
I remember him$_{acc}$ nice$_{acc}$
'I remember him as nice.'

However, in the case of predication of a class of quantifier / numeral phrases, the predicate may agree either with the accusative numeral head or with the genitive NP:

(16) [Kilka drzew] było [wyrwane z ziemi].
a few$_{acc}$ trees$_{gen}$ was$_{3rd,sg,neut}$ torn$_{acc}$ from earth
'A few trees were uprooted.'

(17) [Kilka drzew] było [wyrwanych z ziemi].
a few$_{acc}$ trees$_{gen}$ was$_{3rd,sg,neut}$ torn$_{gen}$ from earth
'A few trees were uprooted.'

I assume here that, in Polish, Numeral Phrases (NumPs) in subject position are accusative (Szober, 1928, 1953; Schenker, 1971; Franks, 1995; Przepiórkowski, 1996, 1999) and that they are headed by the numeral (Saloni, 1976; Saloni and Świdziński, 1985; Przepiórkowski, 1996, 1999). On these assumptions, (16) is expected, but (17), where the predicate agrees with the genitive dependent of the numeral, is surprising.[6]

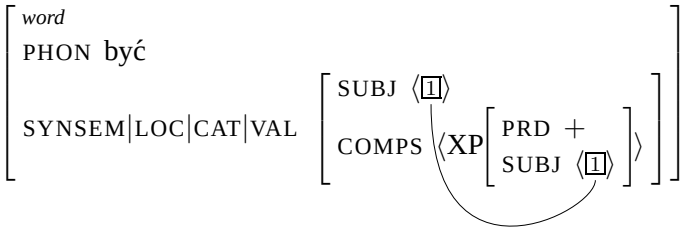
I also adopt the standard HPSG assumptions regarding the copula as a raising verb (cf. (18)), and predicative case agreement as a local phenomenon, i.e., essentially, as agreement between the predicate and *its* subject:[7]

[5] I ignore here so-called 'instrumental of predication'; cf. Pisarkowa 1965, Przepiórkowski 1999.

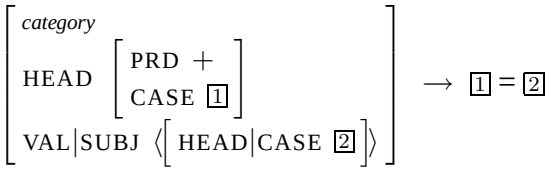
[6] It should be noted that one of (16)–(17) is surprising regardless of our assumptions: whether the numeral is accusative or nominative, as sometimes claimed, the predicate agrees either with the numeral or with the genitive NP. Moreover, if the noun were assumed to be the head of "numeral phrases" in (16)–(17), then case agreement with the numeral in (16) would be unexpected.

[7] See Przepiórkowski (1999) on locality and non-configurationality of case marking.

(18) The predicative copula *być* (schematic and simplified):



(19) Case agreement (simplified):

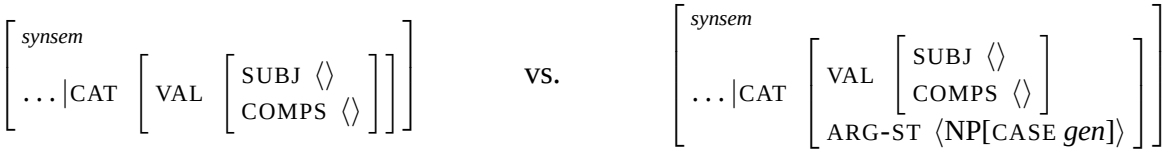


On these assumptions, in (16), the accusative predicative phrase *wyrwanych z ziemi* agrees with the accusative numeral *drzew* in the sense that it agrees with its (predicate's) SUBJECT element, i.e., it agrees with the *synsem* of the NumP *kilka drzew*.

Now, the problem that examples like (17) pose is: how can the predicate agree with the genitive NP *drzew* if the *synsem* of the NumP *kilka drzew* does not contain anything genitive? Note that *kilka drzew* is a fully saturated phrase, i.e., the NP[*gen*] *drzew* has been 'cancelled' from VALENCE features.

Again, the *synsem* of the NumP does not contain anything genitive *unless* the numeral's ARG-ST percolates to the NumP — in such a case, the *synsem* of the genitive NP *drzew* is present within the *synsem* of the NumP *kilka drzew*, namely, in its ARG-ST:

(20) *synsem* of the NumP *kilka drzew*:



If ARG-ST is present on the NumP, then predicative agreement with that NumP could be defined as either agreement with the NumP's CASE or agreement with CASE of NumP's argument. Such an analysis may be ugly, but it correctly deals with the data and, after all, Slavic numeral / quantifier phrases are well known to be ugly and idiosyncratic in many (Corbett, 1978; Franks, 1994; Przepiórkowski, 1999), so it is possible that no elegant analysis for such data is available at all.

See Appendix A for a sketch of an analysis of such constructions.

(Im)Possible Alternatives

Of course, the strength of the argument above rests on the strength of the assumptions it makes and on the unavailability of alternative accounts for (14)–(17).

The assumption that agreement between the modifier (whether attributive or predicative) and the phrase it modifies happens *within* the modifier, i.e., that it is a relationship between the modifier's own morphosyntactic features and those of its argument (selected via VALENCE or MOD) is well-entrenched in HPSG. An alternative would require positing global constraints powerful enough to look into the constituent structure of the subject of the copula (to get hold of the genitive NP); this would violate the overwhelming generalization that case assignment and case agreement is a strictly local phenomenon.

Moreover, there does not seem to be any alternative account of (14)–(17) available. Various such putative alternatives are considered below.

Genitive of Predication One putative alternative would be to say that the genitive on the predicate in (17) is not the result of agreement at all, but rather a 'non-agreeing' option very much like the 'instrumental of predication' in Slavic. According to such an analysis, the predicate may either agree with the phrase it modifies or occur in the genitive.

This analysis makes a blatantly wrong prediction that examples such as (21)–(22) below, to be compared with (14)–(15) above, should be grammatical.

(21) *Janek jest miłego.
John$_{nom}$ is nice$_{gen}$
'John is nice.'

(22) *Pamiętam ją miłej.
I remember her$_{acc}$ nice$_{gen}$
'I remember her as nice.'

A refinement of this alternative, namely, that such a 'genitive of predication' be restricted to numeral subjects would not work either: in colloquial Polish, the paucal numerals *dwa* 'two' to *cztery* 'four', which have all the syntactic properties of numerals (e.g., triggering the 3rd singular neutral 'default' agreement features on the verb) but combine with the agreeing (i.e., accusative) NP argument, do not occur with a genitive predicate:

(23) %(Te) cztery tygodnie było mordercze / *morderczych.
these$_{acc}$ four$_{acc}$ weeks$_{acc}$ was murderous$_{acc}$ / murderous$_{gen}$
'These four weeks were murderous.'

NumPs as NumP/NP Ambiguous Assume that there is a structural ambiguity of numeral phrases: they could be headed either by the accusative numeral, in which case they would occur with accusative predicates, as in (16), or by the genitive noun, in which case they would occur with genitive predicates, as in (17).

However, since attributive adjectives modifying NumPs show the same case optionality as predicative adjectives, this analysis would make the following prediction: when such an NumP/NP is

modified both by an attributive adjective and by a predicative adjective, these adjectives should be either both accusative (in case the NumP/NP is headed by the numeral) or both genitive (in case it is headed by the noun). This prediction is false; see (24) from Kopcińska (1997).

(24) a. Leniwe siedem kotów było śpiących.
lazy$_{acc}$ seven$_{acc}$ cats$_{gen}$ was sleepy$_{gen}$
'Seven lazy cats were sleepy.'

b. Leniwych siedem kotów było śpiące.
lazy$_{gen}$ seven$_{acc}$ cats$_{gen}$ was sleepy$_{acc}$
'Seven lazy cats were sleepy.'

NumPs as NumP+NP Appositions Another possibility would be to treat numeral phrases as NumP + NP appositions. However, this possibility is immediately refuted by the fact that, in Polish, appositions agree in case, while the numeral does not agree in case with the genitive NP.

NumPs as Bi-Headed Another, rather far-fetched hypothesis would be that NumPs are simultaneously headed by the quantifier and by the noun. The NumP would then be, in some sense, accusative and genitive at the same time, and agreeing APs could pick any of these values for the purpose of case agreement. This would account not only for our initial data (1)–(2), but also for (24), problematic for the previous alternative.

A technical problem that this analysis faces is that it is not clear how bi-headedness could be formalized in such a way that both heads contribute their case values. Previous analyses of bi-headedness assume that different heads of a construction contribute *different* sets of features or, when two heads do attempt to contribute values of the same feature, only one of them wins and the other one is suppressed.

There is also an empirical problem, namely: if such bi-headed NumPs were accusative and genitive at the same time, they should be able to occur in both accusative and genitive environments. This prediction is false: NumPs headed by accusative quantifiers cannot occur in genitive environments.

(25) Bałem się tych kilku drzew / *kilka drzew.
feared RM these$_{gen}$ a few$_{gen}$ trees$_{gen}$ / a few$_{acc}$ trees$_{gen}$
'I was afraid of (these) several trees.'

Thus, again, no alternative to an analysis assuming the presence of ARG-ST on (some) phrases seems available.

4 ARG-ST on Phrases with Semantically Vacuous Heads

One conclusion that could be drawn from the previous considerations is that ARG-ST should be present on all phrases, for example, as a HEAD feature. This is a position adopted in Przepiórkowski

1999, as well as in most of the works assuming ARG-ST on phrases cited in §1, with the notable exception of Abeillé and Godard (2000).

However, as noted by many HPSG practitioners, the unconstrained presence of ARG-ST on phrases would go against those locality considerations which led to the introduction of the attribute SYNSEM (Pollard and Sag, 1994, p. 23), so it is desirable to constrain the presence of ARG-ST on phrases to those very few cases where this presence is really necessary. The most elegant such constraint would be based on a single underlying property common for all constructions which require ARG-ST on phrases.[8]

The question that should be answered here is, thus: How to uniformly characterize prepositional phrases of §2 and numeral phrases of §3?

A tentative proposal that I would like to adduce here is this: only ARG-ST of *semantically vacuous* heads 'percolates' to phrases (in Polish).[9]

(26) ARG-ST *on Phrases* (Polish):

The value of the ARG-ST on a headed phrase is structure-shared with the value of ARG-ST of its head daughter if the head daughter is semantically vacuous, and it is the empty list otherwise.

(27) A sign is *semantically vacuous* iff its CONT value is structure-shared with that of one of its arguments. (Pollard and Yoo, 1998; Przepiórkowski, 1997, 1998)

Note that the class of semantically vacuous words is very limited. Typical examples of semantically vacuous words mentioned in the HPSG literature are: non-predicative ('case-assigning') prepositions (Pollard and Sag, 1994, pp. 255, 347ff.), predicative copula (Pollard and Sag, 1994, p. 147), and English *to* and the auxiliary *be* (Pollard and Yoo, 1998).

Since *za* is a non-predicative preposition, the presence of its ARG-ST on its projections is predicted by (26). What about numeral phrases, though? *Can* they be classified as semantically vacuous (from a technical point of view)? If so, *should* they be so analyzed (from an empirical point of view)?

***Can* Numerals be Semantically Vacuous?** According to the analysis of quantification in Przepiórkowski (1997, 1998) (which builds on Pollard and Yoo (1998) and Manning *et al.* (1999)), quantifiers are introduced as values of the attribute NEW-QS[10] (appropriate for *word*) and collected

[8] For Abeillé and Godard (2000) such an underlying property is *liteness*. Another possible restriction has been proposed by Ivan Sag (p.c.), who notes that, cross-linguistically, arguments that must be visible outside the immediate phrases in which they are realized are usually subjects, so — instead of making the whole ARG-ST available on phrases — it should suffice to make the subjects available, perhaps by requiring that SUBJ be a HEAD feature, as in some earlier HPSG work. This proposal cannot be directly applied to the data considered above because, in cases of long raising across a preposition, the *complement* of a preposition must be visible at the PP. While this proposal might be modified in terms of *the first argument on* ARG-ST instead of *the subject*, the solution proposed below is more restrictive than such a modification and thus should be preferred.

[9] See Appendix B for an RSRL formalization of (26)–(27).

[10] This proposal was incorporated into the analysis of Manning *et al.* (1999).

into the QSTORE value via general principles. If so, technically, it is plausible to analyze quantifier words as introducing the *quant* value *only* in NEW-QS and sharing (the rest of) their CONTENT value with that of their NP argument, as in (28).

(28)
$$\begin{bmatrix} \textit{word} \\ \text{PHON}\ \text{pięć} \\ \text{SS}|\text{LOC} \begin{bmatrix} \text{CONT}\ \boxed{0} \\ \text{CAT}|\text{ARG-ST}\ \langle[\text{CONT}\ \boxed{0}]\rangle \end{bmatrix} \\ \text{NEW-QS}\ \{5(\boxed{0})\} \end{bmatrix}$$

***Should* Numerals Share Their CONTENT Value with That of Their Argument?** There is evidence that, in Polish, the numeral and its NP argument share at least INDEX values. Consider the binding examples (29)–(30) below.

(29) [Pięciu facetów] zobaczyło [siebie samych] w lustrze.
five$_{acc}$ guys$_{gen}$ saw$_{3rd,sg,neut}$ Self Emph$_{pl,masc}$ in mirror
'Five guys saw themselves in a mirror.'

(30) [Pięć kobiet] zobaczyło [siebie same] w lustrze.
five$_{acc}$ women$_{gen}$ saw$_{3rd,sg,neut}$ Self Emph$_{pl,fem}$ in mirror
'Five women saw themselves in a mirror.'

The argument rests on the assumption that, in Polish, Numeral Phrases are really headed by the numeral. Since binding involves co-indexation, the index of the NumP in (29)–(30) is the same as the index of the anaphor *siebie*. The index of the anaphor *siebie* in (29) is masculine plural, as evidenced by the morphology of the emphatic modifier of the anaphor, *samych*.[11] This means that the index of the NumP and, hence, the index of the numeral *pięciu* in (29) is masculine plural. But also the index of the NP argument of the numeral, i.e., that of the NP *facetów* is masculine plural, as is clear on the basis of both morphology and reference, so the numeral and the NP seem to have the same index values.

By the same reasoning, the index of the numeral *pięć* in (30) is *feminine plural*, the same as the index of the NP *kobiet*.

Since the *index* of the numeral and that of its NP argument systematically co-vary, I assume that the numeral and the NP actually agree in *index*, i.e., that they share their INDEX values. Note that this argument implies in particular that the CONTENT value of numerals must be *nom-obj* and not *quant*, supporting the structure (28) above.

Now, since 1) the CONTENT of the numeral and of its NP argument are both *nom-obj* and 2) they have the same INDEX value, and 3) the numeral does not introduce any meaning apart from the quantificational force captured in NEW-QS, it makes sense to postulate that the numeral and the NP actually share their CONTENT values, i.e., that numerals in fact are *semantically vacuous* in the sense of (27) above.

[11] The anaphor itself does not overtly inflect for number or gender.

5 Summary

In this paper, I endeavored to make linguistic contributions of two kinds. On the theory-internal side, I argued that the issue whether ARG-ST or any such attribute should be present at the level of possibly saturated *phrase*s, in addition to its presence on *word*s, is not an "all or nothing" issue. Although I showed that there are some environments in Polish which do seem to require the presence of ARG-ST on *phrase*s, I also linked this presence to the common feature of such environments, namely, to their semantic vacuity. Although no formal proof can be given that this is the only possible analysis, I tried to proceed carefully by examining a variety of possible alternatives and showing that all of them fail in one way or another.

Since semantically vacuous environments are extremely rare, the resulting grammar is not less restrictive than, say, a grammar which allows a verb to subcategorize for a lexical argument (and, hence, have access to this argument's ARG-ST), a possibility often taken advantage of in HPSG analyses of complex predicates in various languages.

On the empirical side, I looked at two rarely considered and ill-understood constructions in Polish, namely, at "long raising" across a preposition, and at case agreement with predicative phrases. Neither of these constructions had been successfully analyzed so far: previous analyses of raising constructions such as (1)–(2) wrongly assume that the argument of *za* agrees in case with the object of the raising verb (e.g., Bailyn and Citko 1999; see Przepiórkowski 2000a for extensive discussion), while previous analyses of case agreement in Polish fail to account for (or even notice) the optionality of such case agreement with predicative phrases (e.g., Franks 1994, 1995). Although the analyses proposed here may be perceived as less than satisfactory on the aesthetical side, they constitute the first formal and uniform account of these phenomena.

Appendices

A Predicative Case Agreement with Quantifier Phrases

Full analysis is given in Przepiórkowski (1999, 2000b). The analysis encodes the observation that case agreement between a phrase XP and its modifier YP normally means the identity of XP's and YP's CASE values, but — in case XP is a numeral phrase — it might also mean the identity of CASE values of YP and XP's first *argument*.

In Przepiórkowski (2000b), I extensively argue for linking this case agreement optionality when NumPs are involved to the fact that quantifiers share indices with their NP arguments. On this basis, I propose the following case agreement principles (assuming RSRL as the underlying logic, but not adopting the RSRL notation here):

(31) Attributive case agreement:

$$\left[\begin{array}{l} \textit{head} \\ \text{CASE } \boxed{1} \\ \text{MOD}|\text{LOC } \boxed{2}\left[\text{CAT}|\text{HEAD}|\text{CASE } \boxed{0}\right] \end{array}\right] \rightarrow \texttt{case-agreement(}\boxed{1}\texttt{, }\boxed{2}\texttt{)}$$

(32) Predicative case agreement:

$$\begin{bmatrix} \textit{category} \\ \text{HEAD} \begin{bmatrix} \text{CASE } \boxed{1} \\ \text{PRD } + \end{bmatrix} \\ \text{SUBJ } \langle \begin{bmatrix} \text{LOC } \boxed{2} \begin{bmatrix} \text{CAT}|\text{HEAD}|\text{CASE } \boxed{0} \end{bmatrix} \end{bmatrix} \rangle \end{bmatrix} \rightarrow \texttt{case-agreement}(\boxed{1}, \boxed{2})$$

(33) Definition of case agreement:

$$\texttt{case-agreement}(\boxed{1}\textit{case}, \boxed{2}\textit{local}) \leftrightarrow \begin{array}{l} (\boxed{2} = \begin{bmatrix} \text{CAT}|\text{HEAD}|\text{CASE } \boxed{1} \end{bmatrix} \vee \\ \boxed{2} = \begin{bmatrix} \text{CAT}|\text{ARG-ST } \langle \begin{bmatrix} \text{CASE } \boxed{1} \\ \text{INDEX } \boxed{3} \end{bmatrix}, \ldots \rangle \\ \text{CONT}|\text{INDEX } \boxed{3} \end{bmatrix}) \end{array}$$

Note that these principles preserve the overwhelming generalization that case marking is a local (and possibly non-configurational) phenomenon.

B ARG-ST on Phrases — RSRL Formalization

Technically, I assume the presence of the *list(synsem)*-valued ARG-ST attribute on all *category* objects, and formalize (26)–(27) within RSRL (Richter *et al.*, 1999; Richter, 2000) as follows:

(34) $\forall x \, [\, x \approx : \text{HEAD-DTR} \rightarrow$
$[\, [\texttt{sem-empty}(x) \rightarrow : \text{SS LOC CAT ARG-ST} \approx x\text{SS LOC CAT ARG-ST}]$
$\wedge \, [\neg\texttt{sem-empty}(x) \rightarrow : \text{SS LOC CAT ARG-ST} \sim \textit{elist}] \,] \,]$

(35) $\texttt{sem-empty}(x) \overset{\forall}{\Longleftarrow}$
$\exists y \, \exists z \, [$
$y \approx x\text{SS LOC CAT ARG-ST}$
$\wedge \, \texttt{member}(z, y)$
$\wedge \, z\text{LOC CONT} \approx x\text{SS LOC CONT} \,]$

(See Richter *et al.* (1999) or Richter (2000) for the definition of `member`, as used in (35).)

References

Abeillé, A. and Godard, D. (2000). French word order and lexical weight. In R. D. Borsley, editor, *The Nature and Function of Syntactic Categories*, volume 32 of *Syntax and Semantics*, pages 325–360. Academic Press, San Diego.

Abeillé, A., Godard, D., and Sag, I. A. (1998). Two kinds of composition in French complex predicates. In E. Hinrichs, A. Kathol, and T. Nakazawa, editors, *Complex Predicates in Non-derivational Syntax*, volume 30 of *Syntax and Semantics*, pages 1–41. Academic Press, New York.

Bailyn, J. F. and Citko, B. (1999). Case and agreement in Slavic predicates. In K. Dziwirek, H. Coats, and C. M. Vakareliyska, editors, *Annual Workshop on Formal Approaches to Slavic Linguistics: The Seattle Meeting 1998*, volume 44 of *Michigan Slavic Materials*, pages 17–37, Ann Arbor. Michigan Slavic Publications.

Baxter, D. P. (1999). *English Goal Infinitives*. Ph. D. dissertation, University of Illinois at Urbana-Champaign.

Bouma, G., Malouf, R., and Sag, I. (1999). Satisfying constraints on extraction and adjunction. To appear in *Natural Language and Linguistic Theory*.

Calcagno, M. and Pollard, C. (1997). Argument structure, structural case, and French causatives. Paper deliverd during the 4th International Conference on HPSG, 18–20 July 1997, Ithaca, New York.

Corbett, G. (1978). Numerous squishes and squishy numerals in Slavonic. In B. Comrie, editor, *Classification of Grammatical Categories*, pages 43–73. Linguistic Research, Inc., Edmonton.

Frank, A. (1994). Verb second by underspecification. In H. Trost, editor, *KONVENS'94*, pages 121–130, Vienna.

Frank, A. and Reyle, U. (1995). Principle based semantics for HPSG. In *7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 9–16, University College Dublin, Belfield, Dublin, Ireland. Association for Computational Linguistics.

Franks, S. (1994). Parametric properties of numeral phrases in Slavic. *Natural Language and Linguistic Theory*, **12**(4), 597–674.

Franks, S. (1995). *Parameters of Slavic Morphosyntax*. Oxford University Press, New York and Oxford.

Grover, C. (1995). *Rethinking Some Empty Categories: Missing Objects and Parasitic Gaps in HPSG*. Ph. D. dissertation, University of Essex.

Kopcińska, D. (1997). *Strukturalny Opis Składniowy Zdań z Podmiotem-Mianownikiem we Współczesnej Polszczyźnie*. Dom Wydawniczy Elipsa, Warszawa.

Manning, C. D., Sag, I. A., and Iida, M. (1999). The lexical integrity of Japanese causatives. In R. D. Levine and G. Green, editors, *Studies in Contemporary Phrase Structure Grammar*, pages 39–79. Cambridge University Press, Cambridge.

Marciniak, M. (1999). Toward a binding theory for Polish. In R. D. Borsley and A. Przepiórkowski, editors, *Slavic in Head-Driven Phrase Structure Grammar*, pages 125–147. CSLI Publications, Stanford, CA.

Meurers, W. D. (1999). *Lexical Generalizations in the Syntax of German Non-Finite Constructions*. Ph. D. dissertation, Universität Tübingen.

Miller, P. H. and Sag, I. A. (1997). French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, **15**, 573–639.

Pisarkowa, K. (1965). *Predykatywność Określeń w Polskim Zdaniu*. Zakład narodowy im. Ossolińskich, Wrocław.

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.

Pollard, C. and Yoo, E. J. (1998). A unified theory of scope for quantifiers and *wh*-phrases. *Journal of Linguistics*, **34**, 415–445.

Przepiórkowski, A. (1996). Case assignment in Polish: Towards an HPSG analysis. In C. Grover and E. Vallduví, editors, *Studies in HPSG*, volume 12 of *Edinburgh Working Papers in Cognitive Science*, pages 191–228. Centre for Cognitive Science, University of Edinburgh.

Przepiórkowski, A. (1997). Quantifiers, adjuncts as complements, and scope ambiguities. To appear in *Journal of Linguistics*. Draft of December 2, 1997. Available from: http://www.ipipan.waw.pl/mmgroup/ap-papers.html.

Przepiórkowski, A. (1998). 'A Unified Theory of Scope' revisited: Quantifier retrieval without spurious ambiguities. In G. Bouma, G.-J. M. Kruijff, and R. T. Oehrle, editors, *Proceedings of the Joint Conference on Formal Grammar, Head-Driven Phrase Structure Grammar, and Categorial Grammar, 14–16 August 1998, Saarbrücken*, pages 185–195.

Przepiórkowski, A. (1999). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. Ph. D. dissertation, Universität Tübingen, Germany. Available from: http://www.ipipan.waw.pl/mmgroup/ap-papers.html.

Przepiórkowski, A. (2000a). Case and agreement in Polish predicates. To appear in Steven Franks, Tracy Holloway King, and Michael Yadroff, eds., *Annual workshop on Formal Approaches to Slavic Linguistics: The Bloomington Meeting 2000*. Ann Arbor: Michigan Slavic Publications. Available from: http://www.ipipan.waw.pl/mmgroup/ap-papers.html.

Przepiórkowski, A. (2000b). Predicative case agreement with Quantifier Phrases in Polish. Ms., Ohio State University and Polish Academy of Sciences. Available from: http://www.ipipan.waw.pl/mmgroup/ap-papers.html.

Reinders-Machowska, E. (1991). Binding in Polish. In J. Koster and E. Reuland, editors, *Long-Distance Anaphora*, pages 137–150. Cambridge University Press, Cambridge.

Richter, F. (2000). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Ph. D. dissertation, Universität Tübingen. April 28, 2000.

Richter, F., Sailer, M., and Penn, G. (1999). A formal interpretation of relations and quantification in HPSG. In G. Bouma, E. Hinrichs, G.-J. M. Kruijff, and R. T. Oehrle, editors, *Constraints and Resources in Natural Language Syntax and Semantics*, pages 281–298. CSLI Publications, Stanford, CA.

Saloni, Z. (1976). *Cechy Składniowe Polskiego Czasownika*. Ossolineum, Wrocław.

Saloni, Z. and Świdziński, M. (1985). *Składnia Współczesnego Języka Polskiego*. Państwowe Wydawnictwo Naukowe, Warszawa, 2nd edition.

Schenker, A. M. (1971). Some remarks on Polish quantifiers. *Slavic and East European Journal*, **XV**(1), 54–59.

Szober, S. (1928). *Trzy piękne córki było nas u matki*: Formy podmiotu i orzeczenia w zdaniach z podmiotem logicznym, określonym przydawką liczebnikową. *Język Polski*, **XIII**(4), 97–106.

Szober, S. (1953). *Gramatyka Języka Polskiego*. Nasza Księgarnia, Warszawa, 3rd edition.

Tajsner, P. (1990). *Case Marking in English and Polish: A Government and Binding Study*. Ph. D. dissertation, Adam Mickiewicz University, Poznań.

Wechsler, S. (1997). Prepositional phrases from the twilight zone. *Nordic Journal of Linguistics*, **20**, 127–154.

Ohio State University
Department of Linguistics
222 Oxley Hall
1712 Neil Ave.
Columbus OH 43210
USA
adamp@ling.ohio-state.edu

and

Polish Academy of Sciences
Institute of Computer Science
ul. Ordona 21
01-237 Warszawa
Poland
adamp@ipipan.waw.pl

URL: http://www.ipipan.waw.pl/mmgroup/ap.html