

ARG-ST on Phrases Headed by Semantically Vacuous Words: Evidence from Polish

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., 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) (scope and word order in German), Calcagno and Pollard (1997) and Abeillé and Godard (2000, n. 9) (French causatives), Baxter (1999) (purpose infinitives in English), and Meurers (1999) (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 *impossible*.

The aim of this paper is to examine two rather unusual constructions in Polish, not successfully analysed 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, we show that these two constructions share a rather special property, i.e., they involve semantically vacuous words, in a sense made precise below. We 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.

1 Predicative Modification of Numeral Phrases

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., (1)–(2); in (1) both the subject NP and the predicative AP are nominative, the usual situation, while in (2) both the Numeral Phrase (NumP) and the predicate are accusative. This reflects the fact that, in Polish, Numeral Phrases in subject position are accusative (Przepiórkowski, 1996; Franks, 1995) and that they are headed by the numeral (Saloni and Świdziński, 1985; Przepiórkowski, 1996).

- (1) Janek jest miły.
John_{nom} is nice_{nom}
'John is nice.'
- (2) Kilka drzew było wyrwane z ziemi.
a few_{acc} trees_{gen} were torn_{acc} from earth
'A few trees were uprooted.'

The situation in (1)–(2) is expected on standard HPSG assumptions: the subject of the copula, which is structure-shared with the subject of the adjectival predicate, receives its case according to general case assignment rules, and the predicate agrees with it because it agrees with its (predicate's!) subject.

What is unexpected is the possibility of the adjectival predicate agreeing with the NP argument of the numeral and occurring in the genitive, as in (3); in fact, many speakers prefer the genitive option to the accusative option.

- (3) Kilka drzew było wyrwanych z ziemi.
a few_{acc} trees_{gen} were torn_{gen} from earth
'A few trees were uprooted.'

Maintaining the standard assumption that case agreement in such cases is actually agreement between the predicate and the element on its SUBJ list, and adopting the assumption that ARG-ST is

present only on words, there is no way to account for (3): since the SUBJ list of the predicate *wyrwanych* contains only the *synsem* of the NumP *kilka drzew*, the information about the genitive argument is absent there. In other words, there is nothing genitive within that *synsem* that the predicate could agree with. On the other hand, if ARG-ST percolated from heads to their phrasal projections, the *synsem* of the genitive NP *drzew* would be available within the *synsem* of the whole numeral phrase (namely, on its ARG-ST), and the predicate could agree with it. This is our first argument for having ARG-ST on (some) phrases.

Of course, the strength of this argument rests on the strength of the assumptions it makes and on the unavailability of alternative accounts for (1)–(3).

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 (1)–(3) available. One putative alternative would be to say that the genitive on the predicate is not the result of agreement at all, but rather a ‘non-agreeing’ option; according to such an analysis, the predicate may either agree with the phrase it modifies or occur in the genitive. This analysis would make the wrong prediction that (1) should be possible with the genitive *młatego*, in addition to the nominative *młoty*. A refinement of this alternative, namely, that such a ‘genitive of predication’ would 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:

- (4) %(Te) *cztery tygodnie było mordercze* / **morderczych*.
 these_{acc} four_{acc} weeks_{acc} was murderous_{acc} / *murderous_{gen}*
 ‘These four weeks were murderous.’

Another putative alternative analysis could postulate 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 (2), or by the genitive noun, in which case they would occur with genitive predicates, as in (3). However, since attributive adjectives modifying NumPs show the same case optionality as predicative adjectives, this analysis would make the following prediction: when such a 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 (5) from Kopcińska (1997).

- (5) 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.’

We maintain, then, that examples such as (3) call for the presence of ARG-ST on (some) phrases.

2 Raising across Prepositions

In Polish (as, perhaps, also in other languages), there is a class of raising verbs which raise an argument ‘across a preposition’, as in (6).

- (6) *Uważałem go za szczerego.*
 I considered him_{acc} as sincere_{acc}
 ‘I considered him to be sincere.’

On the assumption (defended below) that *uważać* ‘consider’ takes three arguments, i.e., a subject (*pro*-dropped in (6)), a raised object, and a *fully saturated* PP[*za*] complement, and that the PP involves a preposition and an *almost saturated* predicative phrase, whose subject is raised to the object position

of the verb, there is no way to encode this raising at the level of *uwazać*: the *synsem* of the PP[*za*], which is present on the ARG-ST of *uwazać*, does not contain any information about the subject of the predicative complement of the preposition. That is, unless we allow ARG-ST of the preposition to ‘percolate’ to the PP level; if we allow such ‘percolation’, *uwazać* may have the following (partial and simplified) lexical entry:

$$(7) \left[\begin{array}{l} \text{word} \\ \text{PHON } \text{uwazać} \\ \\ \text{ARG-ST } \langle \text{NP } \boxed{1}, \boxed{0}, \text{PP} \left[\begin{array}{l} \text{PFORM } \text{za} \\ \text{ARG-ST } \langle \left[\begin{array}{l} \text{SUBJ } \langle \boxed{0} \rangle \\ \text{CONT } \boxed{2} \end{array} \right] \rangle \right] \rangle \\ \\ \text{CONT } \left[\begin{array}{l} \text{consider} \\ \text{CONSIDERER } \boxed{1} \\ \text{SOA-ARG } \boxed{2} \end{array} \right] \end{array} \right]$$

This argument for ARG-ST on phrases rests heavily on the assumption that the PP[*za*] on the ARG-ST of *uwazać* is fully saturated. An alternative that is readily available is that the preposition *za* is a raising word itself, i.e., that it raises the subject of its predicative complement to its own subject position. If this were the case then *uwazać* could subcategorize for an *almost* saturated PP[*za*] and raise the unrealized subject of the preposition to its own object.

There is, however, a strong argument against the analysis of *za* as taking two arguments, a subject and an object. As discussed by Wechsler (1997), such 2-argument prepositions have different binding properties in English than ordinary 1-argument prepositions; a similar difference can be observed in Polish, e.g., (8), involving a 2-argument locative preposition, vs. (9), involving a 1-argument ‘case marking’ preposition.

(8) Nie można przecież położyć książki_i na sobie_{?i} samej / na 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.’

(9) 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.’

Although the judgments in (8) are not very clear, the contrast between binding across a predicative preposition (8) and across a ‘case marking’ preposition (9) is clear: since in Polish only subjects can be binders, binding by the object in (9) is impossible, while binding by the object (8) is acceptable, but only because it controls the subject of the predicative preposition.

Now, if *za* were a 2-argument preposition with the first argument controlled by the object of the verb, then it should pattern with predicative prepositions such as *na* in (8). As (10) shows, this prediction is false; *za* as used in (6) clearly patterns with ‘case marking’ prepositions, such as *o* in (9).

(10) (Nie pomyliłem się,) uważałem go_i za siebie_{*i} samego / za niego_i samego.
I wasn’t confused, considered_{1st,sg,masc} him_{acc} as Self Emph_{masc} / as him Emph_{masc}
‘(I wasn’t confused,) I really considered him as himself.’

Another alternative treatment of (6) would assume that *za* is not really a preposition, but rather a marker (cf. Pollard and Sag, 1987, p. 65). In this case, the apparent PP *za szczerego* would actually be an almost saturated predicative AP marked as [MARKING *za*]. Again, the subject of the adjectival predicate would be available at the level of *uwazać*. However, this analysis would predict that phrases such as *za szczerego*, being predicative, should be able to appear in other predicative environments, especially in those environments which do not posit particular constraints on the categorial makeup of the predicative phrase, such the complement of the copula, or exclamations. (11)–(12) show that this prediction is false:

(11) Janek jest szczerzy / 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.’

(12) Janek szczerzy! / Wałęsa prezydentem! / Krokodyl w klatce! / Obiad o dziesiątej! / *Janek
John sincere_{nom} / Wałęsa president_{ins} / crocodile in cage / dinner at ten / John
za szczerego! (Też pomysł!)
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!

Although the copula in (11) could impose an idiosyncratic constraint to the effect that its complement cannot be marked with *za*, such a constraint would violate the otherwise overwhelming generalization that the copula may combine with any predicative complement. Moreover, it is not clear that such a constraint could be imposed in case of (12), where there is no overt copula and no obvious reason to posit a phonologically empty one.

The final alternative would be to treat the PP[*za*] as a thematic predicate in the sense of Wechsler (1997). For lack of space we do not consider it at length here but, in the full paper, we reject it on two grounds: first, PP[*za*] is an obligatory complement of *uwagać*, contrary to what such an analysis would predict, and, second, according to such an analysis, *uwagać* would assign a semantic role to its object, contrary to the received wisdom that it is a raising verb (e.g., Tajsner 1990).

In summary, no analysis of constructions such as (6) is in sight, which would not assume the presence of the ARG-ST of the preposition *za* on its maximal projection.

3 ARG-ST on Phrases with Semantically Vacuous Heads

We have just considered two constructions in Polish which call for the presence of ARG-ST on certain phrases. It is striking that, in both cases, heads of the phrases assumed to bear the ARG-ST are semantically vacuous. This is clear in case of *za* in (6), but perhaps not so clear in case of numerals in sentences such as (3); after all, numerals introduce quantifiers. Nevertheless, numerals are semantically vacuous in the sense that their CONT values are structure-shared with CONT values of their arguments: assuming the analysis of Przepiórkowski (1997, 1998) (see also Manning *et al.* (1999)), quantifiers are introduced as values of the attribute NEW-QS, appropriate for *word* objects, and not within CONT; the CONT value of the quantifier is structure-shared with the CONT value of its NP argument (we discuss empirical arguments for that stance in the full paper). For example, *pięć* ‘five’ may be schematically represented as in (13):

$$(13) \left[\begin{array}{l} \textit{word} \\ \text{PHON pięć} \\ \text{SS|LOC} \left[\begin{array}{l} \text{CONT } \boxed{\emptyset} \\ \text{CAT|ARG-ST } \langle \text{CONT } \boxed{\emptyset} \rangle \end{array} \right] \\ \text{NEW-QS } \{5(\boxed{\emptyset})\} \end{array} \right]$$

It also makes sense to assume that the CONT of *za* is structure-shared with that of its argument: this would slightly simplify the lexical entry (7).

On the basis of these observations, we would like to posit a very restricted version of the ‘ARG-ST on phrases’ hypothesis for Polish (we assume that restrictions on the presence of ARG-ST on phrases may be subject to cross-linguistic variation):

(14) 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.

(15) 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)

Technically, we assume the presence of the *list(synsem)*-valued ARG-ST attribute on all *category* objects, and formalize (14)–(15) within RSRL (Richter *et al.*, 1999; Richter, 2000) as follows:

$$(16) \forall x [x \approx \text{HEAD-DTR} \rightarrow \\ [[\text{sem-empty}(x) \rightarrow \text{SS LOC CAT ARG-ST} \approx x\text{SS LOC CAT ARG-ST}] \\ \wedge [\neg \text{sem-empty}(x) \rightarrow \text{SS LOC CAT ARG-ST} \sim \textit{elist}]]]$$

$$(17) \text{sem-empty}(x) \stackrel{\forall}{\leftarrow} \\ \exists y \exists z [\\ y \approx x\text{SS LOC CAT ARG-ST} \\ \wedge \text{member}(z, y) \\ \wedge z\text{LOC CONT} \approx x\text{SS LOC CONT}]$$

(See Richter *et al.* (1999) or Richter (2000) for the definition of *member*, as used in (17).)

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 Nonderivational Syntax*, volume 30 of *Syntax and Semantics*, pages 1–41. Academic Press, New York.
- Baxter, D. P. (1999). *English Goal Infinitives*. Ph.D. thesis, 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 delivered during the 4th International Conference on HPSG, 18–20 July 1997, Ithaca, New York.
- 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. (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. thesis, 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.
- Meurers, W. D. (1999). *Lexical Generalizations in the Syntax of German Non-Finite Constructions*. Ph.D. thesis, 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.
- Pollard, C. and Sag, I. A. (1987). *Information-Based Syntax and Semantics, Volume 1: Fundamentals*. Number 13 in CSLI Lecture Notes. CSLI Publications, Stanford.
- 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.ling.ohio-state.edu/~adamp/Drafts/>.
- 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 Categorical Grammar, 14–16 August 1998, Saarbrücken*, pages 185–195.
- Richter, F. (2000). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Ph.D. thesis, Universität Tübingen. In progress.
- 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.
- Saloni, Z. and Świdziński, M. (1985). *Składnia Współczesnego Języka Polskiego*. Państwowe Wydawnictwo Naukowe, Warszawa, second edition.
- Tajnsner, P. (1990). *Case Marking in English and Polish: A Government and Binding Study*. Ph.D. thesis, Uniwersytet im. Adama Mickiewicza, Poznań.
- Wechsler, S. (1997). Prepositional phrases from the twilight zone. *Nordic Journal of Linguistics*, **20**, 127–154.