$_{\rm No^{\rm n}}$ Configurational Case Assignment in HPSG

Adam Przepiórkowski adamp@sfs.nphil.uni-tuebingen.de

20th **Ma**996

1 Introduction

Pollard and Sag (1994), p. 30 (bold font mine):

[A]ssignment of case to complements, including nominative case assignment to the subject of finite verbs, is simply treated as part of subcategorisation... There is no separate theory of case (or Case). Nominative case assignment takes place directly within the lexical entry of the finite verb.

However, they add in a footnote (fn. 25, p. 30):

We do acknowledge, however, that for languages with more complex case systems, some sort of distinction analogous to the one characterized in GB work as 'inherent' vs. 'structural' is required.

The aims of this talk:

- to show that structural vs. lexical case distinction, as well as a Case Principle resolving structural cases is useful (if not necessary) for some languages;¹
- to give a general non-configurational formalization of these notions compatible with the traceless dialects of HPSG.

$_{\mathbf{2}}$ $_{\mathbf{Ge}}$ rman Data

2.1 Lexical Case Assignment Does Not Suffice

See (Pollard, 1994, p. 288):

(1) [Den Wagen (acc) zu reparieren] wurde versucht {The car}_{acc} to repair PASS-PAST tried.

'One tried to repair the car.'

In examples like (1), unlike in the example below (2), versucht doesn't raise the arguments of reparieren.

(2) [Zu reparieren versucht] wurde der Wagen (nom) lange Zeit. To repair tried PASS-PAST {the $\operatorname{car}\}_n^{om}$ long time. 'One has been trying to repair the car for a long time.'

The argument: What case should reparieren assign to its object? It can be neither nominative (see (1)) nor accusative (2). It also cannot be underspecified because it wouldn't get resolved by wurde in (1) (the object is not raised there).

A solution (mentioned by Pollard (1994) and formalized by Heinz and Matiasek (1994): a case principle saying that (in German) structural NPs which become complement daughters get accusative case, while the ones that become subject daughters get nominative.

2.2 Structural v. Lexical Dichotomy

Intuitively, structural cases are these cases, which vary according to the syntactical environment and, hence, should be resolved in the grammar. Lexical cases are assigned in the lexicon. (The examples below come from (Pollard, 1994, p. 274), and (Haider, 1984, p. 68) cited after (Haegeman, 1991, p. 175).)

- a. Ich wasche den Wagen (str → acc).
 I wash {the car}_{acc}.

 b. Der Wagen (str → nom) wird gewaschen.
 {The car}_n om PASS washed.
 The car is being washed.'
 - c. * Den Wagen (acc) wird gewaschen. {The car}_{acc} PASS washed.
- (4) a. Ich helfe dem Mann (dat).

 I help {the man}_{dat}

 wird asholt
 - b. * Der Mann (nom) wird geholfen.
 {The man}_n^{om} PASS helped.
 - c. Dem Mann (dat) wird geholfen. {The man}_{dat} PASS helped. 'The man is being helped.'
- (5) a. Sie gedachte vergangener Freuden (gen). She remembered $\{past joy\}_{gen}$.
 - b. $Vergangener\ Freuden\ (gen)\ wurde gedacht.$ {Past joy} $_{gen}$ PASS-PAST remembered.

A solution: lexical items specify their NP complements as either underspecified structural (str) or specific lexical (e.g., dat):

(6) a. waschen:
$$\begin{bmatrix} \text{SUBJ (NP[str])} \\ \text{COMPS (NP[str])} \end{bmatrix}$$
b. helfen:
$$\begin{bmatrix} \text{SUBJ (NP[str])} \\ \text{COMPS (NP[dat])} \end{bmatrix}$$

¹This has been argued for within HPSG for Korean (cf. Yoo (1993), German (cf. Heinz and Matiasek (1994), Pollard (1994) and Gerdemann (1994)), English (cf. ch. 2 of Grover (1995)) and Polish (cf. Przepiórkowski (1996)). Also the analysis of Icelandic case system of Sag et al. (1992) relies on a kind of structural vs. lexical distinction (encoded via features CASE and DCASE) and on a case principle (hard-wired into schema 1) assigning nominative case to structural subjects (p. 316).

3 Polish Data

The data presented in this section is based on Przepiórkowski (1996).

3.1 Genitive of Negation

Example (7) shows that a case principle is useful also in Polish. In this language all otherwise accusative complements of verbs in scope of negation change its case into genitive. Without a case principle this would have to be stated in the lexicon by postulating two lexical entries of each such verb (admittedly, possibly related by a lexical rule). On the other hand, examples (8)–(9) show that, unlike nominative, accusative and the genitive of negation (we remain agnostic as to the status of other occurences of genitive), dative and instrumental are better off analysed as lexical cases; they are not affected by the GoN.

- (7) a. Janek wspiera Marię
 John_{nom} supports Mary_{acc}.
 'John is supporting Mary.'
 - b. Janek nie wspiera Marii John $_{nom}$ not supports Mary $_{gen}$. 'John is not supporting Mary.'
- (8) a. Janek pomaga Tomkowi. $John_{nom}$ helps Tom_{dat} . 'John is helping Tom.'
 - b. Janek nie pomaga Tomkowi.
 John_{nom} not helps Tom_{dat}.
 'John is not helping Tom.'
- (9) a. Janek pogardza Tomkiem $John_{nom}$ scorns Tom_{ins} .

 'John scorns Tom.'
 - Janek nie pogardza Tomkiem
 John_{nom} not scorns Tom_{ins}.
 'John doesn't scorn Tom.'

3.2 Nominalization

The examples below should be compared with (7)–(9) above. They independently confirm the observations we've made above.² (Similar facts can be observed in German, cf. (Heinz and Matiasek, 1994, p. 206).)

- (10) a. Janek wspiera Marię
 John_{nom} supports Mary_{acc}.
 'John is supporting Mary.'
 - b. wspieranie Marii (Janka)
 helping Mary_{gen} (John_{gen})
 'the help for/from Mary (John)'

- (11) a. Janek pomaga Tomkowi. John, helps Tom_{dat} . 'John is helping Tom.'
 - b. pomaganie Tomkowi
 helping Tom_{dat}
 'the help for/*from Tom'
 - c. pomaganie Janka
 helping John_{gen}
 'the help from/*for John'
- (12) a. Janek pogardza Tomkiem $John_{nom}$ scorns Tom_{ins} .
 - b. pogardzanie Tomkiem scorning Tom_{ins}
 'the scorn for/*from Tom'
 - c. pogardzanie Janka
 scorning John_{gen}
 'the scorn from/*for John'

3.3 Indefinite Numerals

In Polish, there is a class of indefinite numerals which are traditionally analyzed as having **only nominative and accusative** forms (cf. Doroszewski (1980), e.g., $du\dot{z}o$ ('a lot'), mato ('little'), troche ('a little'), sporo ('quite a lot'), etc. The puzzle concerning these numerals is that they are grammatical in **some positions which normally require genitive case**, but not in others:³

- (13) a. Nie mam w domu (zbyt) dużo chleba. Not have $_{1st,sing}$ in home (too) a lot of $_{nom/acc}$ bread $_{gen}$.

 'I don't have (too) much bread at home.'
 - Nie mam w domu chleba.
 Not have_{1st,sing} in home bread_{gen}.
 I don't have bread at home.'
 - c. * Nie mam w domu chleb. Not have_{1st,sing} in home bread_{acc}. 'I don't have bread at home.'
- (14) a. Nie lubię dużo osób. Not like $_{1st,sing}$ a lot of $_{nom/acc}$ people $_{gen}$. 'I don't like a lot of people.'
 - b. Nie lubię tych osób. Not like $_{1st,sing}$ these $_{gen}$ people $_{gen}$. 'I don't like these people.'
 - c. * Nie lubię te osoby. Not like $_{1st,sing}$ these $_{acc}$ people $_{acc}$. 'I don't like these people.'

²In Przepiórkowski (1996) we actually argue against structurality of adnominal genitive.

³In these examples b. and c. show that the positions in question are true genitive positions (b.), i.e., that accusative cannot occur there (c.).

(15) a. * $B \circ j \epsilon$ supply a lot of $a \circ j \epsilon$ supply $a \circ j \epsilon$ supply $a \circ j \epsilon$ b. $B \circ j \epsilon$ supply $a \circ j$

Feaths a fraid of these people.' osoby

c. * Boje sign these nom/acc people nom/acc. Familatring of these people.'

With structural vs. lexical distinction in hand this behaviour of the $du\dot{z}o$ -phrases can be easily accounted for by postulating that $du\dot{z}o$ can bear only structural case:

(16)
$$\begin{bmatrix} word \\ PHON & (du\dot{z}o) \\ SYNSEM & LOC & CAT \end{bmatrix} \begin{bmatrix} c^{at} \\ HEAD & AGR & CASE & Str \\ COMPS & NP[AGR & CASE & Gen] \end{bmatrix}$$

3.4 Cases in Polish

On the basis of the foregoing examples we postulate the following case hierarchy for Polish:⁴

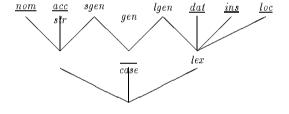


Figure 1: Case Hierarchy for Polish

4 Formaliz^{ation}

4.1 Summary of the Problems

The problems with the previous accounts are following:

• Pollard and Sag (1994): no account of case alternation;

- Pollard (1994): uses structural vs. lexical case distinction but doesn't formalize structural case assignment;
- Heinz and Matiasek (1994), Przepiórkowski (1996) and Grover (1995) assign case configurationally (via DTRS attribute);
- Heinz and Matiasek (1994) and Przepiórkowski (1996) assume traces;
- Grover (1995): redundant case principle.

The aim of this section is to try to find a solution which would be non-configurational, non-redundant and traceless. 5

4.2 Solution

Initial assumptions:

- UDCs via CELR and SELR, hence no traces;
- VALENCE features (SUBJ, COMPS and SPR) appropriate for signs' category, ARG-S appropriate for words' category.

4.2.1 In the Seaarch of a Single Case Assignment Locus

There are four candidates for case assignment locus: SLASH, DTRS, VALENCE and ARG-S because values of these attributes contain relevant *sunsems*.

SLASH Only extracted NPs appear there; in order to make SLASH a single domain of case assignment all complements would have to appear in SLASH at some point. This would be very non-intuitive and rather far-fetched.

DTRS Extracted complements appear in DTRS only in *head-filler-structures*. In order to assign case there, they would have to carry some local information, e.g., subject vs. complement status, whether in the scope of negation, etc. Such formalization would also be redundant.

VALENCE Extracted NPs not present in the VALENCE features, so this is not a good candidate.

ARG-S All arguments (according to some accounts, also adjuncts, cf. Kasper (1994), van Noord and Bouma (1994)) present here, but only at the lexical level.

4.2.2 Bringing Global Information to ARG-S

Structural case cannot be resolved on the basis of local information only, cf. e.g. (1)-(2). However, it can be resolved if the information is added whether a given complement has been 'cancelled' (realized) from the given lexical item.

The trick: let the values of ARG-S and of the VALENCE features be not lists of synsems, but lists

of structures $\begin{bmatrix} arg \\ ARG \text{ synsem} \end{bmatrix}$, where bin has + and - as its maximal subtypes. In the lexicon,

the value of REALIZED is normally underspecified.

⁴The underlined types correspond to the observable (distributional) cases in Polish with the exception of vocative, whose case status is doubtful (cf. Willim (1990)).

⁵The non-configurational case assignment technique presented here is, however, compatible also with the traditional, 'traced', UDC account of Pollard and Sag (1994), as well as with the account of Avgustinova and Oliva (1996), which assumes neither traces nor lexical rules.

The way it is resolved is following:

- SELR and CELR, which remove arguments from the VALENCE features mark these arguments as REALIZED+;
- 2. the (slightly modified) valence principle marks the arguments which are being cancelled as REALIZED+;
- 3. lexical items taking unsaturated complements mark all unrealized arguments of these complements as REALIZED-.

4.3 Two Examples

The examples below illustrate the foregoing considerations.

4.3.1 A Polish Example: Genitive of Negation

Formalizing the Case Principle for Polish In a head-complement-structure of category

- verb[-neg]: the structural object (if any) has a CASE value of acc,
- verb[+neg]: the structural object (if any) has a CASE value of sgen,

The above clauses of the Case Principle for Polish (cf. Przepiórkowski (1996)) can be formalized via the following constraints:

$$(17) \qquad \begin{bmatrix} cat \\ \text{NEG} - \\ \text{ARG-S} \boxed{1}_{ne_list} \oplus \langle \begin{bmatrix} \text{ARG NP}[str] \\ \text{REALIZED} + \end{bmatrix} \rangle \oplus \boxed{2}_{list} \end{bmatrix} \rightarrow \begin{bmatrix} \text{ARG-S} \boxed{1} \oplus \langle \begin{bmatrix} \text{ARG NP}[acc] \end{bmatrix} \rangle \oplus \boxed{2} \end{bmatrix}$$

$$(18) \qquad \begin{bmatrix} \frac{cat}{\text{NEG +}} \\ \text{ARG-S } \boxed{1}_{ne_list} \oplus \langle \begin{bmatrix} \text{ARG NP}[str] \\ \text{REALIZED +} \end{bmatrix} \rangle \oplus \boxed{2}_{list} \end{bmatrix} \rightarrow \begin{bmatrix} \text{ARG-S } \boxed{1} \oplus \langle \begin{bmatrix} \text{ARG NP}[sgen] \end{bmatrix} \rangle \oplus \boxed{2} \end{bmatrix}$$

The Example

(19) Marii Janek nie lubi. Mary_{aen} John_{nom} NEG likes.

(20)
$$lubi$$
 (lexical entry):⁶
$$\begin{bmatrix} \text{HEAD verb[fin]} \\ \text{SUBJ } \left(\boxed{1} \begin{bmatrix} \text{ARG NP[str]} \\ \text{REALIZED bin} \end{bmatrix} \right) \\ \text{COMPS } \left(\boxed{2} \begin{bmatrix} \text{ARG NP[str]} \\ \text{REALIZED bin} \end{bmatrix} \right) \\ \text{ARG-S } \left(\boxed{1}, \boxed{2} \right) \end{bmatrix}$$

(21)
$$lubi$$
 (lexical entry, after CELR):
$$\begin{bmatrix} HEAD \text{ verb[fin]} \\ SUBJ & \left(1 \right) & ARG \text{ NP[str]} \\ REALIZED \text{ bin]} & COMPS & \left(1 \right) \\ ARG-S & \left(1 \right) & \left[2 \right] & ARG \text{ NP[str]} \\ REALIZED & + \right] & NHER|SLASH & \left\{ 2 \right\} & COMPS & \left\{ 2 \right\} & CO$$

$$lubi \text{ (inside the phrase): } \begin{bmatrix} \text{NEG } + \\ \text{ARG-S } \\ \\ \end{bmatrix} \begin{bmatrix} \text{ARG } \boxed{7} \text{NP[str]} \\ \text{REALIZED } + \end{bmatrix}, \boxed{2} \begin{bmatrix} \text{ARG NP[str]} \\ \text{REALIZED } + \\ \end{bmatrix}$$

Now the constraint (18) applies resulting in the following *category* value of the embedded *word*:

$$lubi \ (inside the phrase, partially resolved): ^{8} \ \left[\begin{array}{c} \text{NEG +} \\ \text{ARG-S} \ \left(\boxed{1} \left[\begin{array}{c} \text{ARG} \ \boxed{7} \ \text{NP[str]} \\ \text{REALIZED} \ + \end{array} \right], \boxed{2} \left[\begin{array}{c} \text{ARG NP[sgen]} \\ \text{REALIZED} \ + \end{array} \right] \right) \right]$$

4.3.2 A German Example

Since there are no AcI constructions in Polish, we illustrate the interactions between this phenomenon and case assignment with a German ${\rm AcI}$ example.

Case Principle for German

(25)
$$\begin{bmatrix} at \\ ARG-S & \langle ARG & NP[str] \\ REALIZED & + \end{bmatrix} \rangle \oplus 2_{list} \rightarrow [ARG-S & \langle ARG & NP[nom]] \rangle \oplus 2_{list}$$

(26)
$$\begin{bmatrix} at \\ ARG-S & \boxed{1}_{ne_list} \oplus \langle \begin{bmatrix} ARG & NP[str] \\ REALIZED & + \end{bmatrix} \rangle \oplus \boxed{2}_{list} \end{bmatrix} \rightarrow \begin{bmatrix} ARG-S & \boxed{1} \oplus \langle \begin{bmatrix} ARG & NP[acc] \end{bmatrix} \rangle \oplus \boxed{2}_{list} \end{bmatrix}$$

The Example

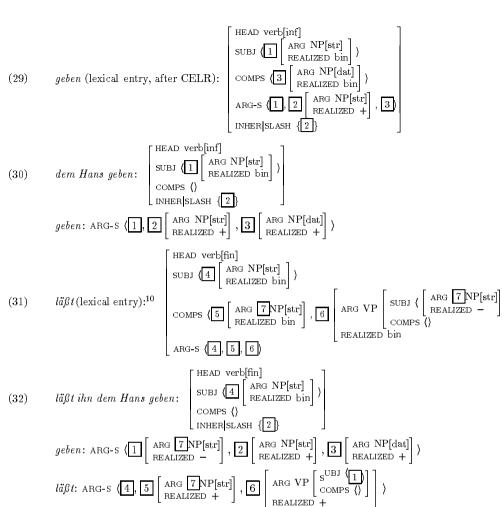
- (27) Den Wagen (acc) $l\ddot{a}\beta t$ sie (nom) ihn (acc) dem Hans (dat) geben. {The car}_{acc} CAUS she_{nom} he_{acc} {the Hans}_{dat} give. 'She made him give the car to Hans.'
- (28) geben (lexical entry): $\begin{bmatrix} \text{HEAD verb[inf]} \\ \text{SUBJ } \left(\boxed{1} \begin{bmatrix} \text{ARG NP[str]} \\ \text{REALIZED bin} \end{bmatrix} \right) \\ \text{COMPS } \left(\boxed{2} \begin{bmatrix} \text{ARG NP[str]} \\ \text{REALIZED bin} \end{bmatrix}, \boxed{3} \begin{bmatrix} \text{ARG NP[dat]} \\ \text{REALIZED bin} \end{bmatrix} \right) \\ \text{ARG-S } \left(\boxed{1}, \boxed{2}, \boxed{3} \right) \end{bmatrix}$

⁶ All the feature structures in these examples are partial and schematic,

⁷We don't commit ourselves to any particular analysis of negation in Polish here. We assume, though, that verbs in the scope of negation bear the NEG+ feature, while non-negated verbs are NEG+.

⁸Similar constraint (not stated here) resolves the case of the subject to nom.

⁹We don't commit ourselves to any particular grammar of German here.



(33) läßt sie ihn dem Hans geben:

[REALIZED + 1], D REALIZED + 1]

[REALIZED + 2], D REALIZED + 2]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D REALIZED + 3], D REALIZED + 3]

[REALIZED + 3], D REALIZED + 3], D

$$l\ddot{a}\beta t: \text{ arg-s } \langle \boxed{4} \left[\begin{array}{c} \text{arg NP[str]} \\ \text{REALIZED } + \end{array} \right], \boxed{5} \left[\begin{array}{c} \text{arg } \boxed{7} \text{NP[str]} \\ \text{REALIZED } + \end{array} \right], \boxed{6} \left[\begin{array}{c} \text{arg VP} \left[\begin{array}{c} \text{S}^{\text{UBJ}} \left\langle \boxed{1} \right\rangle \\ \text{COMPS } \left\langle \right\rangle \end{array} \right] \right] \rangle$$

(34) Den Wagen läßt sie ihn dem Hans geben:

[HEAD verb[fin] SUBJ ()
COMPS ()
INHER|SLASH { }

ARG-S
$$\left(1\right)\begin{bmatrix} \text{ARG } 7 \text{NP[str]} \\ \text{REALIZED } - \end{bmatrix}$$
, $\left[2\right]\begin{bmatrix} \text{ARG NP[acc]} \\ \text{REALIZED } + \end{bmatrix}$, $\left[3\right]\begin{bmatrix} \text{ARG NP[dat]} \\ \text{REALIZED } + \end{bmatrix}$ $\left[3\right]$

$$\begin{array}{c} \text{ARG-S} \; \langle \; \boxed{4} \left[\begin{array}{c} \text{ARG NP[nom]} \\ \text{REALIZED} \; + \end{array} \right], \boxed{5} \left[\begin{array}{c} \text{ARG } \boxed{7} \, \text{NP[acc]} \\ \text{REALIZED} \; + \end{array} \right], \boxed{6} \left[\begin{array}{c} \text{ARG VP} \left[\begin{array}{c} \text{SUBJ } \left\langle \boxed{1} \right\rangle \\ \text{COMPS } \left\langle \right\rangle \end{array} \right] \right] \rangle \end{array}$$

References

- Avgustinova, T. and Oliva, K. (1996). Unbounded dependencies in HPSG without traces or lexical rules. Technical Report CLAUS 70, Universität des Saarlandes, Saarbrücken.
- Doroszewski, W., editor (1980). Słownik Poprawnej Polszczyzny, Warszawa. Państwowe Wydawnictwo Naukowe.
- Gerdemann, D. (1994). Complement inheritance as subcategorization inheritance. In Nerbonne et al. (1994), pages 341-363.
- Grover, C. (1995). Rethinking Some Empty Categories: Missing Objects and Parasitic Gaps in HPSG. PhD dissertation, University of Essex. Available from ftp://ftp.cogsci.ed.ac.uk/pub/grover/thesis.ps.gz.
- Haegeman, L. (1991). Introduction to Government and Binding Theory. Blackwell Textbooks in Linguistics. Blackwell, Oxford.
- Haider, H. (1984). The case in German. In J. Toman, editor, Studies on German Grammar, pages 65–102. Foris, Dodrecht.
- Heinz, W. and Matiasek, J. (1994). Argument structure and case assignment in German. In Nerbonne *et al.* (1994), pages 199–236.
- Kasper, R. (1994). Adjuncts in the mittelfeld. In Nerbonne et al. (1994), pages 39-69.
- Nerbonne, J., Netter, K., and Pollard, C., editors (1994). German in Head-Driven Phrase Structure Grammar, Stanford, CA. Center for the Study of Language and Information.
- Pollard, C. (1994). Toward a unified account of passive in German. In Nerbonne et al. (1994), pages 273-296.
- Pollard, C. and Sag, I. A. (1994). Head-driven Phrase Structure Grammar. Chicago University Press.
- Przepiórkowski, A. (1996). Case assignment in Polish: Towards an HPSG analysis. In C. Grover and E. Vallduví, editors, *Edinburgh Working Papers in Cognitive Science: Studies in HPSG*, pages 195–233. Centre for Cognitive Science, University of Edinburgh.
- Sag, I. A., Karttunen, L., and Goldberg, J. (1992). A lexical analysis of Icelandic case. In I. A. Sag and A. Szabolcsi, editors, *Lexical Matters*, number 24 in Lecture Notes, pages 301–318. Center for the Study of Language and Information, Stanford.
- van Noord, G. and Bouma, G. (1994). Adjuncts and the processing of lexical rules. In Fifteenth International Conference on Computational Linguistics (COLING '94), Kyoto, Japan.
- Willim, E. (1990). On case-marking in Polish. Papers and Studies in Contrastive Linguistics, XXV, 203-220.
- Yoo, E.-J. (1993). Subcategorization and case marking in Korean. In A. Kathol and C. Pollard, editors, *Papers in Syntax*, number 42 in OSUWPL, pages 178–198. Department of Linguistics, Ohio State University.

¹⁰For presentation reasons we assume here that *lassen* raises only the subject of the embedded verb (rather than all its arguments). Nothing hinges on this assumption.