Institute of Computer Science Polish Academy of Sciences
and Université Paris 7

An HPSG Grammar of Polish Clitics

Anna Kupść

Doctoral dissertation
Warszawa 2000

Supervisors: prof. Leonard Bolc
prof. Anne Abeillé

Contents

Acknowledgements

There are many people who helped me to accomplish this complex Polish-French-German project.

I wanted to thank my supervisors, Prof. Anne Abeillé at Université Paris 7 and Prof. Leonard Bolc at the Institute of Computer Science, Polish Academy of Sciences, Warsaw, and my advisor Frank Richter at Universität Tübingen, for their support for this combined project and, of course, for comments and suggestions which helped me to improve my dissertation. Especially Frank's criticism provided a strong stimulus for many improvements at the final stage of this study. I am deeply indebted for his patient assistance in formalising my ideas in RSRL.

I also benefited from comments of other people who read earlier drafts of this study and/or my papers the work presented here is based on. In particular, I wanted to thank: Guergana Popova, Adam Przepiórkowski, Marek Świdziński but also Bob Borsley, Krzysztof Czuba, Magdalena Derwojedowa, Stefan Dyła, Frederik Fouvry, Tracy Holloway King, Valia Kordoni, Małgorzata Marciniak, Detmar Meurers, Paola Monachesi, Agnieszka Mykowiecka, Manfred Sailer, Kiril Simov, Gilbert Rappaport, Jesse Tseng.

I am grateful to those whose native speaker's intuitions I have repeatedly tested: Krzysztof Czuba, Elżbieta Hajnicz, Małgorzata Marciniak, Agnieszka Mykowiecka, Danusia i Adam Przepiórkowscy.

I had good luck to always find an enjoyable working environment and helpful people around me. I would like to thank my office-mates and colleagues from the Language Engenering Group in Warsaw (Elżbieta Hajnicz, Małgorzata Marciniak, Agnieszka Mykowiecka, Piotr Rychlik), the group TALaNa in Paris (first of all Lionel Clément, for his help to shorten the distance between Anne's and my desk but also Pascal Amsili, Marie-Hélène Candito, Patrick Caudal, Morgane Erenati, Veronique Gendner, Kim Gerdes, Sylvain Kahane, Alexandra Kinyon, Veronika Lux, Gaëlle Recoursé, Laurent Roussarie) and the CLaRK programme in Tübingen (Paul King, Atanas Kiryakov, Guergana Popova, Kiril Simov). My special thanks go to those who had to put up with me almost every day when I was writing and completing this work: Paul King, Frank Richter, Valia Kordoni, but particularly, Guergana Popova, Frederik Fouvry and Jesse Tseng.

Finally, I want to mention people who kept on reminding me that there is more to life than my dissertation: my Mum and Dad, my brother and his family, my friends Basia, Agnieszka, Przemek, Mikołaj and, last but not least, Krzysiek.

Résumé

L'objectif de la thèse est une description formelle du polonais, bien adaptée pour le traitement automatique. Les problèmes linguistiques abordés concernent surtout des clitiques (grosso-modo des formes atones ou non-accentués). Les clitiques posent un problème notoire pour l'analyse formelle d'une langue car ils partagent des propriétés avec des affixes et au même temps avec des mots et se placent à l'interface de la morpho(phono)logie et de la syntaxe. Nous avons surtout étudié la description des clitiques pronominaux (par example, *go* 'le', *mu* 'lui'), le clitique réfléchi *się* 'se' et le dit marqueur négatif *nie* 'ne (pas)'; les clitiques verbaux ne font pas partie de la thèse mais ils sont decrits dans le contexte d'autres clitiques. De point de vue linguistique, on a consideré, parmi d'autres, les phénomènes suivantes: le statut grammatical des clitiques, l'ordre des clitiques pronominaux, le comportement des clitiques dans les phrases infinitives et constructions à auxiliaire (dite 'montée des clitiques'), des phrases personnelles et impersonnelles au clitique réfléchi, l'omission du clitique réfléchi dans certains contextes verbaux (dite haplologie) mais aussi des propriétés morpho-syntaxique du marqueur négatif *nie* 'ne (pas)'. Comme l'examen soigneux de leurs propriétés le montre, les clitique polonais ne forment pas un groupe uniforme grammatical. Nous montrons que les clitiques pronominaux et le clitique réfléchi sont des unités syntagmatiques mais le marqueur negatif est un prefixe verbal. Quand au statut des clitiques verbaux, nous adoptons l'analyse de Borsley selon laquelle ce sont des items syntagmatiques.

La description théorique proposée pour les clitiques est formalisée dans le cadre de HPSG (Head-driven Phrase Structure Grammar, une grammaire syntagmatique guidée par la tête, cf. Pollard et Sag (1994)). Comme les clitiques pronominaux et le clitique réfléchi sont des items syntagmatiques, leur analyse est plutôt syntaxique. La spécificité de la langue polonaise et le comportement des clitiques en particulier exigent une modification (une généralisation) des principes proposés par Pollard et Sag (1994). Selon cette modification, la tête lexicale d'un syntagme se combine avec un nombre quelconque de ses compléments, realisés dans un ordre libre. Les clitiques pronominaux et le clitique réfléchi, comme les unités syntagmatiques, obéissent aux principes généraux syntagmatiques de dominance immédiate. Nous introduisons des règles specifiques au polonais, par exemple, pour l'ordre des clitiques syntaxiques. Le clitique réfléchi est polyfonctionel et il a certaines propriétés différentes des clitiques pronominaux. En particulier, nous montrons que c'est la sémantique lexicale des verbes qui autorise la présence du clitique réfléchi, pas seulement la syntaxe comme dans le cas des clitiques pronominaux. Finalement, le marqueur négatif est un préfix verbal et nous proposons une analyse morpho-syntaxique de verbes niés. Nous utilisons des contraintes et la sous-specification lexicale au lieu des règles lexicales, moins efficace, pour obtenir de formes niés.

Le noyau de l'analyse formelle sert comme base pour créer un analyseur simple automatique en ALE (the Attributte Logic Engine), la plat-forme d'implémentation choisie qui permet le traitement automatique de grammaires de type HPSG. L'implémentation comprend la montée

des clitiques aux infinitives et, malgré les règles syntagmatiques rigides d'ALE, on rend compte des différents ordres des clitiques et non-clitiques postverbaux.

Chapter 1

Introduction

The aim of this dissertation is to provide a formal computer-tractable account of certain Polish linguistic phenomena, in particular a formal account of Polish clitics. The theory developed for clitics is encoded within the linguistic framework of HPSG (Head-driven Phrase Structure Grammar), Pollard and Sag (1994). The core of the analysis is implemented in ALE (The Attribute Logic Engine, see Carpenter and Penn (1999)), a computational platform which allows for processing HPSG grammars.

The first two sections of this chapter are devoted to the presentation of the motivation and scope of this study. The last section contains the necessary background concerning the adopted formalism (§1.3.1) and main features of the Polish language (§1.3.2).

1.1 Motivation

The main motivation for this dissertation was the development of a formal grammar of a fragment of Polish well-suited for an efficient automatic treatment.

Polish lacks such a grammar so far. The first computer-oriented descriptions of Polish were pioneered in late seventies and early eighties, see Szpakowicz (1978, 1986), Szpakowicz and Świdziński (1981). This work was then continued by Marek Świdziński and resulted in a formal syntactic description of a broad fragment of Polish in Świdziński (1992). However, as admitted by the author, see Świdziński (1992, p. 58), a direct computational implementation even of a fragment of the proposed grammar will be difficult if not impossible. This is because the grammar of Świdziński (1992), encoded in the framework of Metamorphosis Grammar (MG), cf. Colmerauer (1978), employs a large number of partly redundant re-writing rules which do not express generalizations of the language. Not underestimating the value of this extensive formal description of Polish, it has to be stressed that it is purely syntactic. A more comprehensive account which integrates various aspects of language is necessary if the formal grammar is to be interpreted as a model of natural language.

The framework of HPSG, adopted in this dissertation, is a modern linguistic formalism designed to deal with various aspects of language simultaneously. This means that HPSG not only allows a linguist to work on syntax, morphology, semantics as well as phonology or pragmatics but also to account for phenomena at grammatical interfaces between these levels. HPSG is a lexicalist constraint-based declarative formalism. A theory formulated in terms of HPSG consists of a small set of general principles whereas its main body resides in the lexicon. These features of HPSG allow one to obtain a concise linguistic description. The sound logical grounds of HPSG (King (1989), Richter (2000)) allow for a high degree of formal precision and, consequently, provide theoretical grounds well-prepared for an automatic treatment.

These aspects of the formalism make it particularly attractive both for the investigation of theoretical issues (to mention only a few, see collections of papers in Nerbonne *et al.* (1994) (on German), Balari and Dini (1998) (on Romance) and Borsley and Przepiórkowski (1999) (on Slavic)) and for the application in practical projects (e.g., a large-scale speech to speech translation system developed in the Verbmobil project, the SOAS project on an automatic resolution of ellipsis at King's College, London, an implementation of a large HPSG grammar of German in the SFB340 project, Tübingen).

The success of HPSG as a tool for the formal description of a variety of languages gives a good prospect for its application to Polish as well, which has been confirmed, e.g., by two dissertations on Polish couched in this framework: Mykowiecka (1998) and Przepiórkowski (1999). It should be stressed, however, that the existing HPSG accounts proposed for other languages cannot be directly applied to Polish. A rich inflectional system and free word order distinguish Polish from more constrained Germanic or Romance languages. Therefore, although the HPSG description of Polish presented in this dissertation is inspired by existing approaches, new solutions are proposed as well. Since some of the modifications outlined in this study are not language-specific, they can in turn be re-used for analyses of phenomena in other languages.

A formal description of a language requires a careful classification of its elements. Clitics (roughly speaking unstressed elements) are a notoriously difficult problem for any formal linguistic account. They share properties with both affixes and words and belong to a complex syntax-morpho(phono)logy interface. As observed in Spencer (1991), the system of Polish clitics "is unlike that of other Slav languages in a number of respects, and at the same time it differs from that of [...] Romance languages" (p. 367) and "poses very special problems for a general theory of clitics" (p. 375). One of the sources of these problems is the peculiar distribution of Polish clitics not only with respect to clitics in other languages but also Polish word order in general. Witkoś (1998, p. 1) writes that "it seems puzzling that so many well behaved and very regular word order properties in a notoriously free phrase / word order language such as Polish should concern clitics". In spite of their exceptional behaviour, Polish clitics attracted much less attention than their counterparts in other languages; moreover, not all clitics were investigated equally well and some issues have been recognized only recently. For example, the discussion of the clitic *się* 'self' is conspicuously absent from Witkoś's (1998) monograph on Polish clitics while the haplology of *się* has been more extensively addressed for the first time in Rappaport (1998). Our goal in this dissertation is to address less known issues concerning Polish clitics specified in the next section.

1.2 Scope

In Polish, the term 'clitic' comprises weak (unstressed) forms of personal pronouns or auxiliary verbs but also the negative marker *nie* 'not' and certain particles, see Dukiewicz and Sawicka (1995). As noted in Franks (1998, p. 76), while the special behaviour of auxiliary clitics "has long been recognized and inspired a considerable debate, surprisingly little attention has been paid to the idiosyncrasies of pronominal clitics in Polish". The main concern of this dissertation are pronominal clitics, the reflexive clitic *się* 'self' (which we call the reflexive marker, RM) and the negative marker whereas the auxiliary clitics are considered only marginally.

The first issue we address is the grammatical status of clitics. As the careful consideration of their properties shows, Polish clitics do not form a uniform grammatical class. As we argue, pronominal clitics and the RM are syntactic items whereas the negative marker *nie* is a verbal prefix. As far as the grammatical status of auxiliary clitics is concerned, we follow Borsley (1999) and treat them as syntactic items.

The behaviour of clitics is discussed in a wide range of syntactic environments: simple clauses, verbal complexes, auxiliary constructions, coordinated structures as well as participle or deverbal nominal phrases. The range of clitic-related phenomena covered by this dissertation includes the so-called clitic climbing (CC), haplology of the RM and linear positions of clitics.

As mentioned above, Polish pronominal clitics and the RM are syntactic items and, hence, the account provided for them is mostly syntactic. Prosody and phonology are present in the discussion of the grammatical status of clitics but are absent in the formal analysis. We show that the presence of the RM is related to the lexical semantics of verbs, which is also reflected in our analysis. Finally, as the negative marker is a verbal prefix, a morphosyntactic analysis of negated verbs is provided.

The account of clitics proposed in this dissertation is not demoted to one module of the grammar but is implemented as a close interaction of syntax, semantics and morphology.

1.3 Background

1.3.1 HPSG

In this subsection, we explain the basic ideas and mechanisms of HPSG. It should be stressed that this brief and informal introduction is by no means exhaustive and its main objective is to introduce the notions which are indispensable for understanding the present study.[1] Necessary modifications and extensions will be provided later on in the text. As a formal foundation for this study we assume RSRL (Relational Speciate Re-entrant Language) of Richter (2000), which allows not only for an interpretation of the classical concepts of typed feature logics such as sorts and attributes but also relations.

HPSG is an eclectic generative formalism influenced by the GB (Government and Binding) tradition (e.g., Chomsky (1981)) but most closely related to GPSG (Generalized Phrase Structure Grammar, Gazdar *et al.* (1985)), LFG (Lexical Functional Grammar, Bresnan (1982)) and CG (Categorial Grammar, cf. Ajdukiewicz (1935)). HPSG is a comprehensive and monostratal linguistic framework in which various aspects of a language can be dealt with simultaneously. HPSG belongs to a class of nonderivational formalisms, i.e., there is no ordering of the application of grammatical principles.

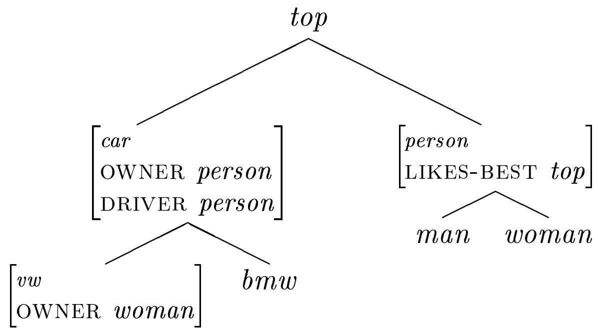
An HPSG grammar consists of a *signature* (a set of sorts, usually organized in a sort hierarchy, with specifications of attributes and their values at each sort) and a *theory* (a set of constraints); cf. King (1989), Pollard (1999), Richter (2000). Roughly speaking, the signature specifies potential objects of a grammar whereas the theory states which of these objects are licensed by the grammar.

[1] The traditional reference for HPSG is Pollard and Sag (1994). An informal presentation of an HPSG-like (but not really HPSG) formalism is given in Sag and Wasow (1999). For a brief outline of the main ideas of HPSG, as well as an overview of current research see Przepiórkowski (2000b).

Signature A signature provides a collection of (non-logical) symbols used in the formal language (e.g., in the principles of an HPSG grammar): sorts, a sort hierarchy (a finite partial order), attributes as well as the appropriateness function and relation symbols, cf. Richter (2000, pp. 155ff) for a formal presentation. The appropriateness function specifies which attributes are introduced at a given sort and which values the attributes can have. Moreover, appropriateness expresses attribute inheritance in sort hierarchies: if an attribute is appropriate to a given sort, it is also appropriate to all its subsorts; the value of an attribute at a subsort is at least as specific as the value of the attribute at the sort.

Let us consider a simple example of a signature (based on Richter (2000, p. 32)).

(1.1)



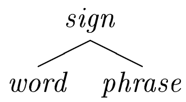
In (1.1), there are seven sorts (*top*, *car*, *vw*, *bmw*, *person*, *man*, *woman*) and three attributes (OWNER, DRIVER and LIKES-BEST). The attributes OWNER and DRIVER, introduced at the sort *car*, are appropriate to the sort *car* but also to its subsorts, i.e., *vw* and *bmw*. (Analogously, the attribute LIKES-BEST introduced at the sort *person* is appropriate also to its subsorts, i.e., *man* and *woman*.) The sort *person* is the appropriate value of both attributes at the sort *car*. In (1.1), the two attributes appropriate to *car* are not repeated on the *bmw* subsort. This means that the values of the attributes OWNER and DRIVER inherited from the sort *car* are not more specific on the sort *bmw*, i.e., they remain *person*. Note, however, that the attribute OWNER is repeated on the subsort *vw* since its value is more specific: *woman* is a subsort of *person*. The appropriate value of the attribute LIKES-BEST at the sort *person* (as well as its subsorts since the attribute is not repeated) is *top*.

A signature is just a set of symbols and by itself does not define any entities. An interpretation of the signature, however, "provides a set of entities and assigns each symbol of the signature a meaning in terms of these entities", Richter (2000, p. 155). In case of sorts, an interpretation is assigned directly only to maximally specific sorts (species, i.e., sorts with no proper subsorts in the sort hierarchy). The denotation of sorts which are not maximally specific is the union of sets of entities denoted by the species that are subsumed by these sorts, cf. Richter (2000, p. 158). For example, the sort *person* in the signature in (1.1) does not have a direct interpretation since it is not a species (*person* has two subsorts: *man* and *woman*). In the natural interpretation of the signature in (1.1), the species *woman* denotes a set of women and the species *man* denotes a set of men. The denotation of the sort *person* is the union of sets of entities denoted by its species, i.e., women and men. The sort *woman* assigned to the attribute OWNER at the sort *vw* means that VWs are owned by women (according to the signature in (1.1)). In what follows, if a symbol of the signature has a straightforward linguistic interpretation, the distinction between a symbol and its interpretation will be sometimes ignored.

After this brief general introduction, let us proceed to the presentation of the relevant parts of the signature of Pollard and Sag (1994).

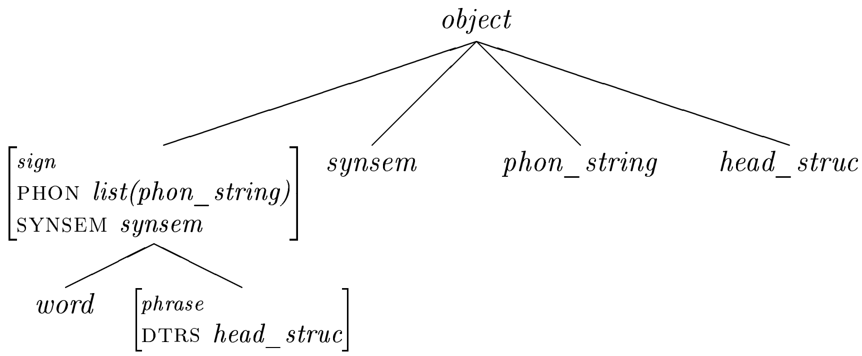
In HPSG, all linguistic expressions are in the denotation of the sorts *word* or *phrase* (but not both of them) which have a common supersort *sign*. The sort *phrase* is assigned not only to phrases but also to sentences or discourses, Pollard and Sag (1994, p. 15). The part of the sort hierarchy for the sort *sign* is given in (1.2).

(1.2)



A more elaborate part of the signature which comprises the sort hierarchy in (1.2) and contains a declaration of attributes is depicted in (1.3).

(1.3)



As (1.3) indicates, every sort in Pollard and Sag (1994) is (directly or indirectly) a subsort of *object*. This means that an *object* is either a *sign* (hence, a *word* or a *phrase*), a *synsem*, a *phon_string*, etc.

As (1.3) shows, there are two attributes appropriate to the sort *sign* (hence, *word* and *phrase* as well): SYNSEM (SYNTAX-SEMANTICS) and PHON (PHONOLOGY).[2] The value of the former is used to specify syntactic and semantic properties of a linguistic object whereas the latter is dedicated to phonology. The values of the attributes appropriate to *sign* are specified as a *synsem* object and a list of *phon_string* objects, respectively.[3,4] Note that subsorts may introduce their own attributes. In addition to PHON and SYNSEM, the attribute DTRS (DAUGHTERS) is appropriate to the sort *phrase*. The value of the DTRS attribute encodes the constituent structure of a phrase. Since we will consider only headed structures, the value of DTRS in (1.3) is an object of the sort *head_struc* (*head_structure*). The subsorts of *sign* with all appropriate attributes are exemplified in (1.4).

(1.4) a. $\begin{bmatrix} \textit{word} \\ \textsc{phon}\ \textit{list(phon\_string)} \\ \textsc{synsem}\ \textit{synsem} \end{bmatrix}$ b. $\begin{bmatrix} \textit{phrase} \\ \textsc{phon}\ \textit{list(phon\_string)} \\ \textsc{synsem}\ \textit{synsem} \\ \textsc{dtrs}\ \textit{head\_struc} \end{bmatrix}$

[2]For simplicity, we do not introduce all sorts and attributes. For example, the attribute QSTORE, appropriate to *sign* in Pollard and Sag (1994), is omitted here.

[3]The representation of phonology is very simplistic here. See Bird and Klein (1994) and especially Höhle (1999) for an elaborate architecture for phonology in HPSG.

[4]Technically, a sort *list* is also a subsort of *object*. Minimally, it has two subsorts: *ne_list* (a non-empty list) and *e_list* (an empty list). The former (unlike the latter) has two attributes: FIRST (for the head of the list) and REST (for the tail of the list). Formally, parametric sorts, such as *list(phon_string)* in (1.3), are not sorts, cf. Penn (1998). Throughout this study, the notation *list(σ)* is used as a convenient abbreviation for a specification that uses relations to restrict values of the respective attribute to be a list of objects of the sort σ (see Richter (2000§4.5.2)).

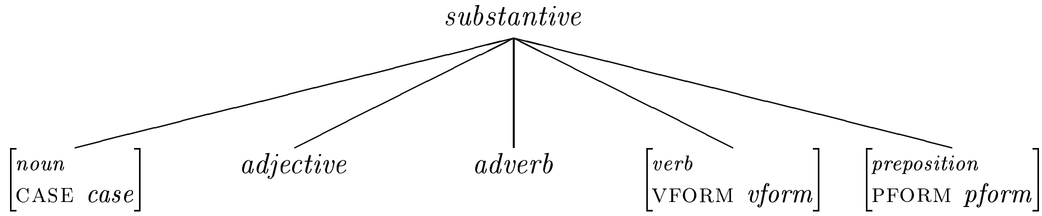


Figure 1.1: A part of the signature for *substantive* (English)

There are two attributes appropriate to *synsem*: LOCAL (with an object of the sort *local* as its value) and NONLOCAL (its value is *nonlocal*). The latter, which will not be employed in this thesis, is used in analyses of unbounded dependencies such as *wh*-extraction, questions, etc. There are three attributes appropriate to the sort *local*: CATEGORY (with an object of the sort *category* as its value), CONTENT and CONTEXT with values *content* and *context*, respectively. The last two sorts denote objects containing (roughly) semantic and pragmatic information, see Pollard and Sag (1994, p. 24). The sort *category*, with the appropriate attributes HEAD (with the value *head*), VALENCE (with the value *valence*) and a list-valued ARG-ST (ARGUMENT-STRUCTURE), denotes objects which encode morphosyntactic properties of a sign. The general architecture of an object of the sort *synsem* is given in (1.5).

$$(1.5)\qquad \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL} & \begin{bmatrix} \textit{local} & \\ \text{CATEGORY} & \begin{bmatrix} \textit{category} \\ \text{HEAD}\ \textit{head} \\ \text{VALENCE}\ \textit{valence} \\ \text{ARG-ST}\ \textit{list(synsem)} \end{bmatrix} \\ \text{CONTENT}\ \textit{content} & \\ \text{CONTEXT}\ \textit{context} & \end{bmatrix} \\ \text{NONLOCAL}\ \textit{nonlocal} & \end{bmatrix}$$

All sorts implicitly introduced above, i.e., *local*, *nonlocal*, *category*, etc., along with the specification of appropriate attributes and their values, must be explicitly added to the signature in (1.3). The sorts which occur in (1.5) are immediate subsorts of *object*.[5]

The sort *head*, which roughly corresponds to a part of speech, is split into *substantive* (*subst*) and *functional* (*funct*) subsorts. The latter, which will not appear in this study, is introduced in Pollard and Sag (1994) as a supersort of functional elements such as a *determiner* or a *marker* (e.g., a *complementizer*). The sort *substantive* has several subsorts: *noun*, *adjective*, *adverb*, *verb* and *preposition*. These subsorts along with the appropriate attributes are given in fig. 1.1 (for a more realistic signature envisaged for *substantive* in Polish see Przepiórkowski (1999, p. 420)).

Subsorts of *case* as well as *vform* and *pform* are language-specific. For example, for English the *nominative* and *accusative* subsorts of *case* are sufficient but for Polish, a language with a richer case system (cf. §1.3.2), more subsorts are necessary. In particular, in addition to the classification of *case* based on morphology, we adopt the *lexical* (inherent) vs. *structural* (context-sensitive) case dichotomy proposed for Polish in Przepiórkowski (1996, 1999). The subsorts of *case*, relevant for our analysis, will be introduced in §1.3.2. As far as subsorts of *vform* are concerned, we will mostly use the subsorts *personal* (for inflected verbal forms) and

[5]The *list(synsem)* technically is not a sort, see fn. 4

impersonal (for non-inflected verb forms, e.g., an *infinitive*, based on the hierarchy of verbal sorts for Polish in Przepiórkowski (1999, p. 420)). A few subsorts of *pform* will appear in Chapter 3 for the prepositions *do* 'to', *w* 'in', etc.

In Pollard and Sag (1994, ch.1–8), the values of the SUBCAT attribute (SUBCATEGORIZATION) encode the subcategorization frame of a sign. However, in subsequent HPSG work, the function of the attribute is distributed between SUBJ (SUBJECT) and COMPS (COMPLEMENTS) attributes[6] (each of them takes a list of *synsem* objects as value) whereas SUBCAT is renamed ARG-S(T) (ARGUMENT-STRUCTURE) and often retained only for lexical signs, e.g., Sag and Godard (1993), Sag and Fodor (1994), Manning and Sag (1998).[7] In this study, we assume that ARG-ST is present only in lexical signs and its value is the concatenation of SUBJ and COMPS lists (see (1.12) below) (however, see Sag (1997), Miller and Sag (1997), Manning and Sag (1998), for possible dissociations between ARG-ST and valence attributes). The SUBJ and COMPS attributes are appropriate to the *valence* sort, cf. (1.6). The theory additionally requires that the SUBJ value is a list which contains at most one element (but there are no restrictions on the length of the COMPS list).

(1.6) $$\begin{bmatrix} \textit{valence} \\ \text{SUBJ } \textit{list(synsem)} \\ \text{COMPS } \textit{list(synsem)} \end{bmatrix}$$

So far, we have not said much about how configurational dependencies are reflected in HPSG. As we said, the constituent structure is encoded in the value of the DTRS attribute, cf. (1.3), i.e., an object of the sort *head_struc*. The attributes HEAD-DTR (HEAD-DAUGHTER), SUBJ-DTR (SUBJECT-DAUGHTER) and COMP-DTRS (COMPLEMENT-DAUGHTERS) are appropriate to the sort *head_struc*. The objects in the denotation of subsorts of *head_struc* correspond to various kinds of phrases. The two subsorts introduced in Pollard and Sag (1994, p. 391) which will mostly concern us here are *head_subj_struc* (*head_subject_structure*) and *head_comp_struc* (*head_complement_structure*), cf. (1.7).

(1.7) $$\begin{array}{c} \begin{bmatrix} \textit{head\_struc} \\ \text{HEAD-DTR } \textit{sign} \\ \text{SUBJ-DTR } \textit{list(phrase)} \\ \text{COMP-DTRS } \textit{list(phrase)} \end{bmatrix} \\ \swarrow \qquad \searrow \\ \begin{bmatrix} \textit{head\_comp\_struc} \\ \text{SUBJ-DTR } \textit{e\_list} \end{bmatrix} \quad \begin{bmatrix} \textit{head\_subj\_struc} \\ \text{COMP-DTRS } \textit{e\_list} \end{bmatrix} \end{array}$$

All attributes introduced for a given sort are appropriate to all its subsorts. Therefore, HEAD-DTR and SUBJ-DTR / COMP-DTRS are appropriate to *head_subj_struc* and *head_comp_struc*.

Theory A theory is a set of descriptions, roughly expressions built from the symbols provided by the signature and logical operators, e.g., conjunction '$\wedge$', disjunction '$\vee$' or implication '$\rightarrow$'. An HPSG theory is a set of constraints (grammatical principles), usually given as implicational descriptions ($p \rightarrow q$), which must be simultaneously satisfied by all objects in a model of a grammar. Formally, a grammar is a pair consisting of a signature and a

[6]For simplicity, we omit here the third valence attribute SPR (SPECIFIER).

[7]Since ARG-ST is not appropriate to the sort *word* but to *category*, some modifications of the signature are necessary to ensure that the attribute appears only in lexical signs. Possible technical solutions can be found, e.g., in Przepiórkowski and Kupść (1997a).

theory. A model is a specific kind of interpretation of a grammar; namely an interpretation of the signature such that each object in the interpretation satisfies each description in the theory. In other words, each principle of the grammar describes every object in a model of the grammar. However, an arbitrary model of the grammar, in a sense, might not be big enough to characterize a natural language. For example, a model which contains only one sentence *Kim walks* is an unsatisfactory model of the grammar in Pollard and Sag (1994). In order to characterize a natural language adequately, we need the notion of an exhaustive model of the grammar. Informally, an exhaustive model contains instances of all potential objects that are licensed by the grammar. For a discussion of different kinds of adequate models see King (1999) and Pollard (1999).

At present, we turn to the presentation of grammatical principles (constraints) in Pollard and Sag (1994). An example of a constraint is the Head Feature Principle (HFP). This principle ensures that in a headed phrase morphosyntactic features (the HEAD value) of the mother are the same as the morphosyntactic features of the head daughter, cf. (1.8).

(1.8) $$\begin{bmatrix} \textit{phrase} \\ \text{DTRS } \textit{head\_struc} \end{bmatrix} \rightarrow \begin{bmatrix} \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY}|\text{HEAD } \boxed{1} \\ \text{DTRS}|\text{HEAD-DTR}|\text{SYNSEM}|\text{LOCAL}|\text{CATEGORY}|\text{HEAD } \boxed{1} \end{bmatrix}$$

The constraint in (1.8) should be understood as a logical implication with the antecedent specified on the left-hand side of '$\rightarrow$' and the consequent on its right-hand side: if an object satisfies a description in the antecedent, it must also satisfy the description in the consequent. More specifically, (1.8) says that if a linguistic object is of the sort *phrase* and its DTRS value is an object of the sort *head_struc*, then the SYNSEM|LOCAL|CATEGORY|HEAD value of this *phrase* object must be the same as its DTRS|HEAD-DTR|SYNSEM|LOCAL|CATEGORY|HEAD value. The tag '$\boxed{1}$' is a variable which indicates that the sequences of attributes (i.e., paths) lead to the same object, i.e., the values of these paths are token identical.

In HPSG, well-formed phrase structures are licensed by immediate dominance (ID) schemata. Only the immediate dominance of the components of a phrase is constrained but not their relative order; linear order is encoded independently and it will be discussed separately in Chapter 2. All ID schemata are defined as disjuncts in the consequent of the Immediate Dominance Principle: every phrase must satisfy one of the ID schemata, cf. (1.9).

(1.9) $\textit{phrase} \rightarrow (\text{SCHEMA1} \vee \text{SCHEMA2} \vee \ldots \vee \text{SCHEMAN})$

As before, every object which satisfies the description in the antecedent, i.e., if the object is of the sort *phrase*, must also satisfy the description in the consequent. In (1.9), the consequent is a disjunctive description and it is satisfied if at least one of its disjuncts is satisfied. Therefore, in order to be in a model of the grammar, any object of the sort *phrase* must satisfy the description provided by SCHEMA1 or SCHEMA2 or ... or SCHEMAN.

In this study we will be mostly concerned with two ID schemata: a schema which licenses a phrase in which the head daughter combines with the subject (SCHEMA1 in Pollard and Sag (1994))[8] and a schema which licenses phrases which consist of the lexical head and its complements (SCHEMA2 in Pollard and Sag (1994)). The corresponding definitions are given below.

- SCHEMA1 describes a saturated phrase (the values of its SUBJ and COMPS attributes are empty lists, i.e., $\langle\rangle$) with a DTRS value of sort *head_subj_struc* in which the HEAD-DTR is a phrasal sign and the SUBJ-DTR value is a list of length one:

[8] For convenience, we use SCHEMA1 as in Pollard and Sag (1994) and assume that the subject combines with a phrase. Nothing hinges on this assumption.

$$\textsc{Schema1} \equiv \left[\begin{array}{ll} \textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence} & \left[\begin{array}{l} \textit{valence} \\ \textsc{subj}\ \langle\rangle \\ \textsc{comps}\ \langle\rangle \end{array}\right] \\ \textsc{dtrs} & \left[\begin{array}{l} \textit{head\_subj\_struc} \\ \textsc{head-dtr}\ \textit{phrase} \\ \textsc{subj-dtr}\ \langle\ \_\ \rangle \end{array}\right] \end{array}\right]$$

- Schema2 describes a phrase with saturated complements (only the COMPS value is the empty list) with the DTRS value of sort *head_ comp_ struc* in which the HEAD-DTR value is a lexical sign:

$$\textsc{Schema2} \equiv \left[\begin{array}{ll} \multicolumn{2}{l}{\textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{comps}\ \langle\rangle} \\ \textsc{dtrs} & \left[\begin{array}{l} \textit{head\_comp\_struc} \\ \textsc{head-dtr}\ \textit{word} \end{array}\right] \end{array}\right]$$

These descriptions might look too general. However they are not the only source of restrictions on phrases admitted by the grammar. In addition, all objects admitted by the grammar must obey appropriateness conditions defined for sorts and attributes by the signature. For example, for objects that satisfy Schema1, the value of SUBJ-DTR is not just *any* list of length one but, due to the declaration of the sort *head_ struc* in (1.7), it must be a list of *phrase*s and its unique element must be an object of sort *phrase*. Furthermore, objects admitted by the grammar must satisfy *all* principles of the grammar simultaneously. Hence, they must also satisfy the HFP in (1.8) which forces the token identity of the corresponding HEAD values. Below we discuss another principle, the Valence Principle, which identifies elements of the SUBJ-DTR / COMP-DTRS lists as realised subcategorization requirements of the head daughter.

The Valence Principle ensures that the subcategorization requirements of a head daughter which are satisfied by non-head daughters are 'checked off' and 'removed' from the list of subcategorization requirements of the mother (similar to the cancellation of arguments in Categorial Grammar), cf. (1.10).

(1.10) a. For each headed phrase, the value of the SUBJ attribute of the head daughter is the concatenation of the list of SYNSEM values of its non-head daughters with the SUBJ value of the phrase:

$$\left[\begin{array}{l} \textit{phrase} \\ \textsc{dtrs}\ \textit{head\_struc} \end{array}\right] \rightarrow$$

$$\left(\left[\begin{array}{ll} \multicolumn{2}{l}{\textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{subj}\ \boxed{1}} \\ \textsc{dtrs} & \left[\begin{array}{l} \textsc{head-dtr}|\textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{subj}\ \boxed{2'} \oplus \boxed{1} \\ \textsc{subj-dtr}\ \boxed{2} \end{array}\right] \end{array}\right] \right.$$

$$\left. \wedge\ \texttt{phrases\_to\_synsems}(\boxed{2},\boxed{2'}) \right)$$

b. For each headed phrase, the value of the COMPS attribute of the head daughter is the concatenation of the list of SYNSEM values of its non-head daughters with the COMPS value of the phrase:

$$\left[\begin{array}{l} \textit{phrase} \\ \textsc{dtrs}\ \textit{head\_struc} \end{array}\right] \rightarrow$$

$$\left(\left[\begin{array}{ll} \multicolumn{2}{l}{\textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{comps}\ \boxed{1}} \\ \textsc{dtrs} & \left[\begin{array}{l} \textsc{head-dtr}|\textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{comps}\ \boxed{2'} \oplus \boxed{1} \\ \textsc{comp-dtrs}\ \boxed{2} \end{array}\right] \end{array}\right] \right.$$

$$\left. \wedge\ \texttt{phrases\_to\_synsems}(\boxed{2},\boxed{2'}) \right)$$

Let us have a closer look at this principle. The two constraints in (1.10a) and (1.10b) should be understood as two conjuncts of the Valence Principle. The principle specifies the values of SUBJ and COMPS lists of a headed phrase and regulates their relationship to non-head daughters. The concatenation of lists is obtained via the append relation indicated in (1.10b) in the functional notation using the '$\oplus$' symbol. Note that the Valence Principle specifies the order in which syntactic requirements are satisfied by non-head daughters: elements which appear first on valence lists ($\boxed{2'}$) must be identified with the corresponding elements of the SUBJ-DTR / COMP-DTRS lists first. We call an element of a valence list 'realised' if it is identified with an element of the SUBJ-DTR / COMP-DTRS lists. The Valence Principle does not specify the number of elements of valence lists which are (or can be) realised. This restriction is provided by ID schemata. For example, SCHEMA2 requires the COMPS value of the phrase to be empty. Therefore, all phrases which have a lexical head daughter will have to realise their complements (if any) simultaneously. Finally, observe that the tags in (1.10) denote objects of different sorts: values of SUBJ and COMPS are lists of *synsem* objects whereas the SUBJ-DTR / COMP-DTRS values are lists of *phrase*s, see (1.6) and (1.7), respectively. We use the **`phrases_to_synsems`** relation to map a list of *phrase*s to a corresponding list of *synsem*s (for a formalisation of an analogous relation see Richter (2000, p. 251)).

Lexicon In HPSG, the lexicon can be defined as a finite set of lexical descriptions (lexical entries L_1, L_2, ..., L_N) which denote the words admitted by the grammar, see Meurers (1994, 1999). Formally, then, the lexicon is part of the theory. The lexicon can be specified as a constraint which must be satisfied by every object of the sort *word*, see (1.11)

(1.11) $\quad \textit{word} \rightarrow (\text{L\_1} \vee \text{L\_2} \vee \ldots \vee \text{L\_N})$

Note that the constraint in (1.11), sometimes called the Word Principle, has the same logical status as principles discussed above: it specifies which objects of the sort *word* are in the models of the grammar. Observe also that lexical entries are descriptions of *word* objects. A single lexical entry may describe several non-isomorphic *word* objects.

Another principle which we assume for *word* objects is given in (1.12).

(1.12) $$\textit{word} \rightarrow \left[\text{SYNSEM|LOCAL|CATEGORY} \begin{bmatrix} \textit{category} \\ \text{VALENCE} \begin{bmatrix} \textit{valence} \\ \text{SUBJ } \boxed{1} \\ \text{COMPS } \boxed{2} \end{bmatrix} \\ \text{ARG-ST } \boxed{1} \oplus \boxed{2} \end{bmatrix}\right]$$

If this principle is combined with the Word Principle in (1.11), it will ensure that in every object of the sort *word* licensed by the Word Principle the relationship between the value of ARG-ST and valence attributes is as specified in (1.12).

An Example To show how the principles discussed above interact, let us consider the simple sentence *John likes Tom.*

The relevant parts of lexical entries for words *John, likes* and *Tom* are given in (1.13)–(1.15).

$$(1.13)\quad \left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \text{John} \rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \boxed{1}\textit{noun} \\ \text{VALENCE} & \left[\begin{array}{ll} \textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{array}\right] \end{array}\right] \end{array}\right]$$

$$(1.14)\quad \left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \text{likes} \rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \boxed{0}\textit{verb} \\ \text{VALENCE} & \left[\begin{array}{ll} \textit{valence} & \\ \text{SUBJ} & \langle\boxed{2}\text{NP}\rangle \\ \text{COMPS} & \langle\boxed{3}\text{NP}\rangle \end{array}\right] \end{array}\right] \end{array}\right]$$

$$(1.15)\quad \left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \text{Tom} \rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \boxed{4}\textit{noun} \\ \text{VALENCE} & \left[\begin{array}{ll} \textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{array}\right] \end{array}\right] \end{array}\right]$$

The 'NP' which appears in (1.14) abbreviates the description of the synsem component of a saturated noun phrase, see (1.20) below.

(1.16)–(1.18) describe phrases which are immediate head projections of words specified by the lexical entries in (1.13)–(1.15). These phrases satisfy the ID Principle (via SCHEMA2 since the value of HEAD-DTR is an object of sort *word*) as well as the HFP and the Valence Principle.

$$(1.16)\quad \left[\begin{array}{ll} \textit{phrase} & \\ \text{PHON} & \langle \text{John} \rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \boxed{1}\textit{noun} \\ \text{VALENCE} & \left[\begin{array}{ll} \textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{array}\right] \end{array}\right] \\ \text{DTRS} & \left[\begin{array}{ll} \textit{head\_comp\_struc} & \\ \text{HEAD-DTR} & \boxed{(1.13)} \\ \text{COMP-DTRS} & \langle\rangle \end{array}\right] \end{array}\right]$$

$$(1.17)\quad \left[\begin{array}{ll} \textit{phrase} & \\ \text{PHON} & \langle \text{likes Tom} \rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \left[\begin{array}{ll} \textit{category} & \\ \text{HEAD} & \boxed{0}\textit{verb} \\ \text{VALENCE} & \left[\begin{array}{ll} \textit{valence} & \\ \text{SUBJ} & \langle\boxed{2}\text{NP}\rangle \\ \text{COMPS} & \langle\rangle \end{array}\right] \end{array}\right] \\ \text{DTRS} & \left[\begin{array}{ll} \textit{head\_comp\_struc} & \\ \text{HEAD-DTR} & \boxed{(1.14)} \\ \text{COMP-DTRS} & \langle\boxed{(1.18)}\rangle \end{array}\right] \end{array}\right]$$

$$(1.18)\quad \begin{bmatrix} \textit{phrase} & \\ \text{PHON} & \langle\text{Tom}\rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \boxed{4}\textit{noun} \\ \text{VALENCE} & \begin{bmatrix}\textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle\end{bmatrix}\end{bmatrix} \\ \text{DTRS} & \begin{bmatrix}\textit{head\_comp\_struc} & \\ \text{HEAD-DTR} & \boxed{(1.15)} \\ \text{COMP-DTRS} & \langle\rangle\end{bmatrix} \end{bmatrix}$$

Finally, SCHEMA1 along with the HFP and the Valence Principle license the fully saturated phrase described in (1.19).

$$(1.19)\quad \begin{bmatrix} \textit{phrase} & \\ \text{PHON} & \langle\text{John likes Tom}\rangle \\ \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \boxed{0}\textit{verb} \\ \text{VALENCE} & \begin{bmatrix}\textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle\end{bmatrix}\end{bmatrix} \\ \text{DTRS} & \begin{bmatrix}\textit{head\_subj\_struc} & \\ \text{HEAD-DTR} & \boxed{(1.17)} \\ \text{SUBJ-DTR} & \langle\boxed{(1.16)}\rangle\end{bmatrix} \end{bmatrix}$$

A more detailed description of the phrase in (1.19) is given in fig. 1.2 on p. 21.

As such complex descriptions easily become unreadable, it is common in HPSG to picture the constituent structure in a more familiar tree structure notation, see fig. 1.3 on p. 22.

Abbreviations In order to make descriptions more compact, we introduce a number of conventional abbreviations as used in the HPSG literature.

(1.20) a saturated nominal phrase:

$$\text{NP} \equiv \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL}|\text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \textit{noun} \\ \text{VALENCE} & \begin{bmatrix}\textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle\end{bmatrix}\end{bmatrix} \end{bmatrix}$$

(1.21) a NP with a specified case value, e.g., NP[*nom*], or generally:

$$\text{NP}[\boxed{1}] \equiv \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL}|\text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \begin{bmatrix}\textit{noun} & \\ \text{CASE} & \boxed{1}\end{bmatrix} \\ \text{VALENCE} & \begin{bmatrix}\textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle\end{bmatrix}\end{bmatrix} \end{bmatrix}$$

(1.22) a prepositional phrase:

$$
\begin{bmatrix}
\textit{phrase} \\
\text{PHON} \ \langle \text{John likes Tom} \rangle \\
\text{SS}|\text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{0} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle\rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \\
\text{DTRS} \ \begin{bmatrix}
\textit{head\_subj\_struc} \\
\text{H-DTR} \ \begin{bmatrix}
\textit{phrase} \\
\text{PHON} \ \langle \text{likes Tom} \rangle \\
\text{SS}|\text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{0} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle \boxed{2}\text{NP} \rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \\
\text{DTRS} \ \begin{bmatrix}
\textit{head\_comp\_struc} \\
\text{H-DTR} \ \begin{bmatrix} \textit{word} \\ \text{PHON} \ \langle \text{likes} \rangle \\ \text{SS}|\text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{0}\textit{verb} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle \boxed{2}\text{NP} \rangle \\ \text{COMPS} \ \langle \boxed{3}\text{NP} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{C-DTRS} \ \left\langle \begin{bmatrix}
\textit{phrase} \\
\text{PHON} \ \langle \text{Tom} \rangle \\
\text{SS} \ \boxed{3} \begin{bmatrix} \text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{4} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle\rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{DTRS} \ \begin{bmatrix} \textit{head\_comp\_struc} \\ \text{H-DTR} \ \begin{bmatrix} \textit{word} \\ \text{PHON} \ \langle \text{Tom} \rangle \\ \text{SS}|\text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{4}\textit{noun} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle\rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{C-DTRS} \ \langle\rangle \end{bmatrix}
\end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix} \\
\text{S-DTR} \ \left\langle \begin{bmatrix}
\textit{phrase} \\
\text{PHON} \ \langle \text{John} \rangle \\
\text{SS} \ \boxed{2} \begin{bmatrix} \text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{1} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle\rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\
\text{DTRS} \ \begin{bmatrix} \textit{head\_comp\_struc} \\ \text{H-DTR} \ \begin{bmatrix} \textit{word} \\ \text{PHON} \ \langle \text{John} \rangle \\ \text{SS}|\text{LOC}|\text{CAT} \ \begin{bmatrix} \textit{category} \\ \text{HEAD} \ \boxed{1}\textit{noun} \\ \text{VAL} \ \begin{bmatrix} \textit{valence} \\ \text{SUBJ} \ \langle\rangle \\ \text{COMPS} \ \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{C-DTRS} \ \langle\rangle \end{bmatrix}
\end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

Figure 1.2: An analysis of the sentence *John likes Tom*

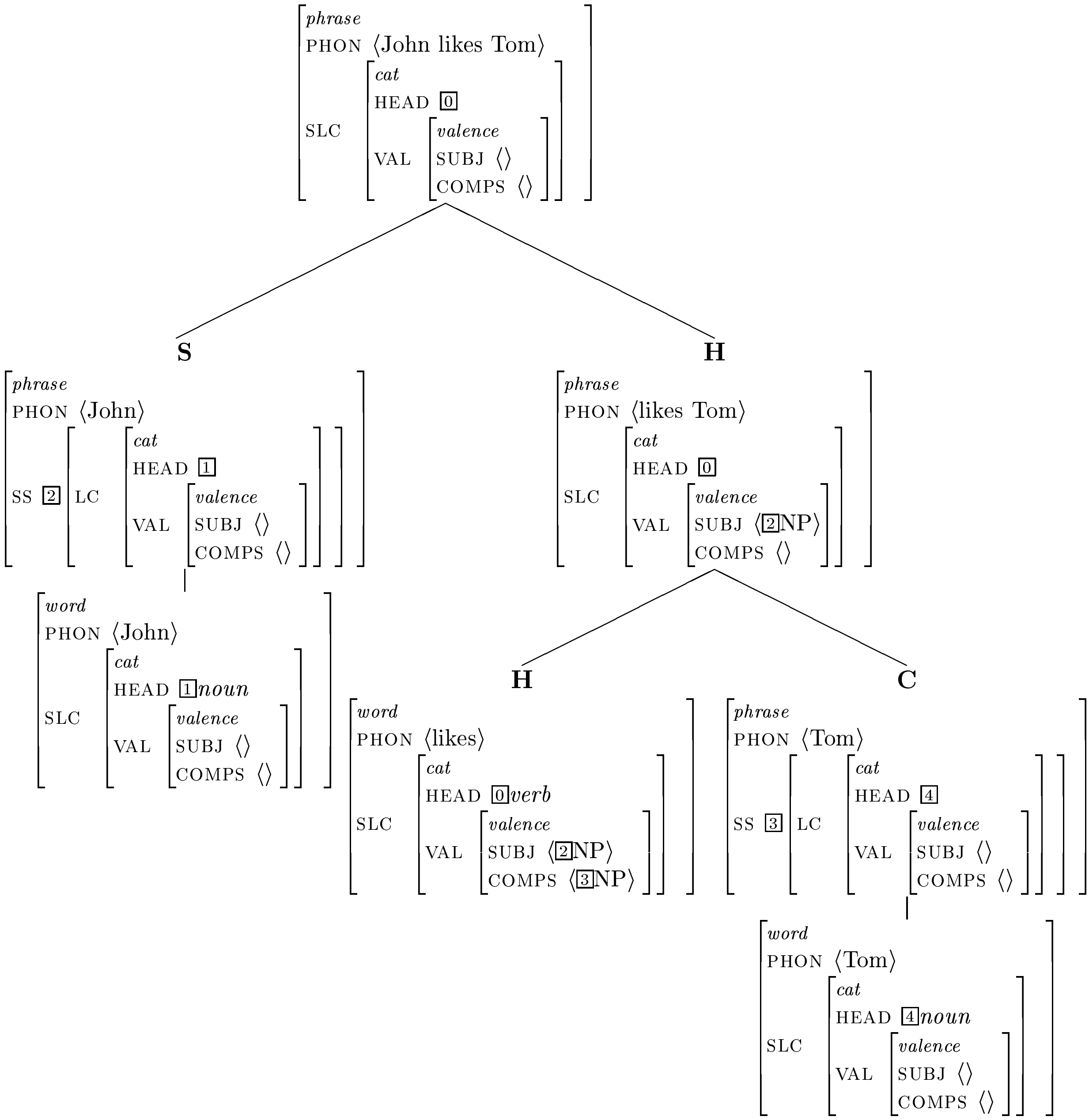


Figure 1.3: A tree structure for *John likes Tom*

$$\text{PP} \equiv \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL}|\text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \textit{preposition} \\ \text{VALENCE} & \begin{bmatrix} \textit{valence} & \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(1.23) a PP with a specified form of the preposition, e.g., PP[*at*], or generally:

$$\text{PP}[\boxed{1}] \equiv \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL}|\text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \begin{bmatrix} \textit{preposition} & \\ \text{PFORM} & \boxed{1} \end{bmatrix} \\ \text{VALENCE} & \begin{bmatrix} \textit{valence} & \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(1.24) a verb with the specified form, e.g., *verb*[*inf*], or generally:

$$\textit{verb}[\boxed{1}] \equiv \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \boxed{1} \end{bmatrix}$$

(1.25) the CONTENT value of the object:

$$\text{XP}{:}\boxed{2} \equiv \text{XP}\big[\text{CONTENT } \boxed{2}\big] \text{ e.g.,}$$

$$\text{NP}{:}\boxed{1} \equiv \begin{bmatrix} \textit{synsem} & \\ \text{LOCAL} & \begin{bmatrix} \textit{local} & \\ \text{CATEGORY} & \begin{bmatrix} \textit{category} & \\ \text{HEAD} & \textit{noun} \\ \text{VALENCE} & \begin{bmatrix} \textit{valence} & \\ \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONTENT} & \boxed{1} \end{bmatrix} \end{bmatrix}, \text{ etc.}$$

Other abbreviations of descriptions will be explained as they appear in the text.

We will also often use mnemonic abbreviations for sorts, e.g., *val* for *valence*, as in fig. 1.2, attributes, e.g., H-DTR for HEAD-DTR or SS for SYNSEM, as well as for paths, e.g., SLC for SYNSEM|LOCAL|CATEGORY in fig. 1.3.

1.3.2 Polish

In order to sketch the linguistic background for this study, we briefly present the main features of the Polish language.

Polish is a West Slavic language. Like many other Slavic languages, Polish is an inflectional pro-drop language with a relatively free word order; additionally there are no definite / indefinite articles in Polish.

Inflection Polish nouns have a grammatical gender. Traditionally, three genders are distinguished in Polish: masculine, feminine and neuter. In fact, a more fine grained classification is necessary, especially for 'masculine' nouns, e.g., Saloni and Świdziński (1985, 1998), Czuba (1997). For the purposes of this study, however, the three basic genders will be sufficient.

Polish nouns may appear either in the singular or plural form. Although there are reasons to consider the number and the gender categories as a single morphosyntactic category, cf. Saloni and Świdziński (1985, 1998), Czuba (1997), throughout this study we will retain the traditional distinction.

Polish has seven morphological cases: nominative, genitive, dative, accusative, instrumental, locative and vocative. For the purpose of the formal analysis, we adopt the cross-classification of Polish morphological cases proposed in Przepiórkowski (1996, 1999), namely a distinction into *lexical* (inherent) vs. *structural* (sensitive to the syntactic context) cases. Following Przepiórkowski (1999), we treat nominative and accusative as structural cases whereas dative, instrumental and locative are lexical cases in Polish. Genitive can be either lexical or structural case in Polish. For the analysis, we will assume the following maximally specific subsorts of *case*: *snom*, *sacc*, *ldat*, *lins*, *loc*, *sgen*, *lgen*.[9] (As in Przepiórkowski (1999), vocative is not considered in the formal analysis.)

Although pronominals are traditionally distinguished as a separate grammatical category, we will include personal pronominals in the class of nominals. Just like nouns, Polish personal pronouns are equipped with the gender, number, case and number information. Inflectional paradigms of Polish personal pronouns will be given in Chapter 2.

Some examples of inflectional paradigms for nouns are given in (1.26)–(1.28).

(1.26) *chłopak* 'a boy' (a masculine noun)

	SINGULAR	PLURAL
nom	chłopak	chłopaki
gen	chłopaka	chłopaków
dat	chłopakowi	chłopakom
acc	chłopaka	chłopaków
inst	chłopakiem	chłopakami
loc	chłopaku	chłopakach
voc	chłopaku!	chłopaki!

(1.27) *dziewczyna* 'a girl' (a feminine noun)

	SINGULAR	PLURAL
nom	dziewczyna	dziewczyny
gen	dziewczyny	dziewczyn
dat	dziewczynie	dziewczynom
acc	dziewczynę	dziewczyny
inst	dziewczyną	dziewczynami
loc	dziewczynie	dziewczynach
voc	dziewczyno!	dziewczyny!

(1.28) *dziecko* 'a child' (a neuter noun)

	SINGULAR	PLURAL
nom	dziecko	dzieci
gen	dziecka	dzieci
dat	dziecku	dzieciom
acc	dziecko	dzieci
inst	dzieckiem	dziećmi
loc	dziecku	dzieciach
voc	dziecko!	dzieci!

[9]One more maximally specific subsort of *case* will be introduced in Chapter 3.

Polish verbs are inflected for person, number, tense and mood but aspect is a derivational category.[10] Polish verbs may occur in the present, past and future tense forms. A part of an inflectional paradigm for the aspectual pair *kupować* 'to buy' (imperfective) vs. *kupić* (perfective) is illustrated in (1.29).

(1.29) a. *kupować* 'to buy' (an imperfective verb) PAST TENSE

	SINGULAR			PLURAL	
	MASC	NEUT	FEM	MASC	FEM & NEUT
1st	kupowałem	kupowałom	kupowałam	kupowaliśmy	kupowałyśmy
2nd	kupowałeś	kupowałoś	kupowałaś	kupowaliście	kupowałyście
3rd	kupował	kupowało	kupowała	kupowali	kupowały

b. *kupić* 'to buy' (a perfective verb) PAST TENSE

	SINGULAR			PLURAL	
	MASC	NEUT	FEM	MASC	FEM & NEUT
1st	kupiłem	kupiłom	kupiłam	kupiliśmy	kupiłyśmy
2nd	kupiłeś	kupiłoś	kupiłaś	kupiliście	kupiłyście
3rd	kupił	kupiło	kupiła	kupili	kupiły

Perfective verbs have no present tense, cf. (1.30a). Imperfective verbs have analytical forms in the future tense, cf. (1.30b).

(1.30) a. *kupić* 'to buy' (a perfective verb)

kupił / kupi
buy$_{3.sg.masc.past}$ buy$_{3.sg.fut}$
'he bought / will buy'

b. *kupować* 'to buy' (an imperfective verb)

kupował / kupuje / będzie kupować / kupował
buy$_{3.sg.masc.past}$ buy$_{3.sg.pres}$ AUX.FUT$_{3.sg}$ buy$_{inf}$ buy$_{3.sg.masc.l-part}$
'he was buying / is buying / will be buying'

As far as the mood is concerned, Polish has indicative (see (1.29) and (1.30)), conditional and imperative moods. The conditional forms, exemplified in (1.31), will be discussed in more detail in Chapter 2.

(1.31) *kupić* 'to buy' (a perfective verb) CONDITIONAL

	SINGULAR			PLURAL	
	MASC	NEUT	FEM	MASC	NEUT & FEM
1st	kupiłbym	kupiłobym	kupiłabym	kupilibyśmy	kupiłybyśmy
2nd	kupiłbyś	kupiłobyś	kupiłabyś	kupilibyście	kupiłybyście
3rd	kupiłby	kupiłoby	kupiłaby	kupiłiby	kupiłyby

In addition to these finite (personal) verb forms, there are a number of non-finite verbal forms such as an infinitive (e.g., *kupić* 'to buy') or various participles, e.g., present adverbial participle (for imperfective verbs, e.g., *kupując* 'buying'), past adverbial participle (for perfective verbs, e.g., *kupiwszy* 'having bought'), active adjectival participle (for imperfective verbs, e.g., *kupujący*), passive adjectival participle (perfective: *kupiony*, imperfective: *kupowany*).

[10] As we will see in Chapter 4, negation is also a morphological category of Polish verbs.

Occasionally, the so-called *l*-participle, the form used with an auxiliary in analytical verb forms is distinguished (e.g., in the future tense of imperfective verbs in (1.30b)),

Adjectives in Polish are inflected for case, number and gender. An example of an inflectional paradigm for the adjective *dobry* 'good' is given in (1.32).

(1.32) *dobry* 'good'

	SINGULAR			PLURAL	
	MASC	NEUT	FEM	MASC	NEUT & FEM
nom	dobry	dobre	dobra	dobrzy	dobre
gen	dobrego	dobrego	dobrej	dobrych	dobre
dat	dobremu	dobremu	dobrej	dobrym	dobrym
acc	dobrego	dobrego	dobrą	dobrych	dobre
inst	dobrym	dobrym	dobrą	dobrymi	dobrymi
loc	dobrym	dobrym	dobrej	dobrych	dobrych
voc	dobry!	dobre!	dobra!	dobrzy!	dobre!

Adjectives are also gradable, see (1.33).

(1.33) dobry lepszy najlepszy
good better best

Polish prepositions and adverbs are not inflected for person, number, gender or case but de-adjectival adverbs are gradable; (1.34) exemplifies all degrees of the adverb *dobrze* 'well'.

(1.34) dobrze lepiej najlepiej
well better best

Agreement Inflected verbs must agree in person, number and gender with the nominative subject. Since Polish is a pro-drop language, the subject may be dropped and its form can be reconstructed from the inflected verb form.

(1.35) a. (Maria) szła.
Mary$_{3.sg.fem}$ walk$_{past.3.sg.fem}$
'Mary walked.'

b. * Maria szli / szłam.
Mary$_{3.sg.fem}$ walk$_{past.3.pl.masc}$ walk$_{past.1.sg.fem}$

Also adjectives must agree in number, gender and case with the noun they modify, see (1.36).

(1.36) a. wesołego chłopaka
cheerful$_{masc.sg.gen/acc}$ boy$_{masc.sg.gen/acc}$
'a cheerful boy'

b. * wesołego chłopak / chłopakom
cheerful$_{masc.sg.gen/acc}$ boy$_{masc.sg.nom}$ boy$_{masc.pl.dat}$

See Czuba (1995b) and Czuba and Przepiórkowski (1995) for a detailed discussion and an HPSG analysis of agreement in Polish.

Word Order Polish is a language with a relatively free word order. Although the neutral word order is SVO (subject-verb-object), cf. (1.37a), all other permutations exemplified in (1.37b–f) are perfectly grammatical.

(1.37) a. Jan lubi Marysię.
John$_{nom}$ likes Mary$_{acc}$
'John likes Mary.'

b. Lubi Jan Marysię.

c. Lubi Marysię Jan.

d. Marysię lubi Jan.

e. Marysię Jan lubi.

f. Jan Marysię lubi.

A more detailed discussion of word order in Polish is postponed to Chapter 2.

Chapter 2

Pronominal Clitics

We begin our discussion of Polish clitics with the description of pronominal clitics. This chapter is concerned with the grammatical status of Polish pronominal clitics, syntactic properties of clitics in infinitival verbal complexes and tense auxiliary constructions (including the so-called clitic climbing (CC)) and also linear order of clitics.[1]

After a presentation of the terminology adopted in this study and a brief overview of general properties of Polish pronominal clitics, we proceed to a detailed discussion of their grammatical status, the issue which apparently has not been overtly addressed so far. Polish pronominal clitics, as we show, are syntactic items and we provide a syntactic account for them. The behaviour of pronominal clitics in verbal complexes and tense auxiliaries indicates that CC is a syntactic phenomenon in Polish: clitics are raised in syntax to a governing verb. The distribution of pronominal clitics distinguishes them from other NPs and we provide an analysis which uniformly accounts for their placement both in simple sentences and verbal complexes.

The HPSG analysis of pronominal clitics presented in this chapter provides a basis for the account of the reflexive clitic in Chapter 3.

2.1 Preliminaries

2.1.1 Terminology

In this subsection, we introduce the terminology which will be used throughout this dissertation for clitics.

The term *clitic* is usually associated with unstressed (prosodically deficient) elements which "exhibit behaviour intermediate between that of a word and that of an affix", see Trask (1993, p. 46). This definition is rather vague, which reflects the fact that elements called 'clitics' may have distinct properties both cross-linguistically as well as within the same language. In this dissertation, the term *clitic* will be used in this very general sense whereas several more specific terms are introduced below.

One of the common properties of clitics is that they cannot stand alone and require a phonological *host* to be attached to. For example, the Portuguese first person pronominal clitic *me* 'me' leans on the preceding verb in *via-me* 'he saw me', cf. Spencer (1991, p. 363). A

[1] This chapter originates from a series of my papers Kupść (1999e,c,b,f,a, 2000a).

clitic can be either a *proclitic* or an *enclitic*, depending whether the phonological host follows or precedes the clitic.

As the definition of a clitic is quite broad, there have been several attempts to work out a terminology for more specific subclasses and relate this pretheoretical term to standard grammatical concepts such as an affix or a word.

In his seminal paper, Zwicky (1977) distinguishes three classes of clitics: *simple clitics*, *special clitics* and *bound words*. Simple clitics are defined as phonologically reduced forms of corresponding full forms. The distribution of simple clitics is regulated by the same principles which are responsible for the distribution of full forms. A typical example of a simple clitic are reduced forms of English auxiliary verbs, e.g., *'ve* in 'We've met before' = 'We have met before'.

Special clitics usually have an opaque phonological form with respect to the corresponding full variant (if any). More importantly, rules which govern the distribution of special clitics differ from those which apply to non-clitics. Examples of special clitics given in Zwicky (1977, p. 3) include French personal pronouns, e.g., *le* 'him' (a clitic) vs. *lui* (a non-clitic), among others.

Bound words, although always bound and unaccented, show considerable freedom of phonological attachment. More specifically, they are "semantically associated with an entire constituent while being phonologically attached to one word of this constituent", cf. Zwicky (1977, p. 6). Additionally, bound words often do not have a corresponding non-clitic counterpart. The English possessive marker *'s* is given in Zwicky (1977, p. 6) as a representative of this class of clitics.

In the subsequent work, this tripartition seems to be silently reduced to a bipartition with bound words incorporated to the class of special clitics; for example the possessive *'s* is called "a paradigm example of one type of special clitic" in Zwicky (1987, p. 133).

In Anderson (1992), simple clitics are not treated as first-class citizens of the clitic class; "a simple clitic is merely a lexical item whose phonological form does not include assignment to a prosodic unit at the level of 'word"', cf. Anderson (1992, p. 201). The term 'clitic' is primarily reserved for special clitics. Building on the work of Klavans (1985), Anderson explores similarities between special clitics and affixes and calls the former *phrasal affixes* (PAs). The analogy is supported by the fact that a phrasal affix has a fixed position with respect to its *syntactic domain* (a constituent to which the clitic is attached) just like an affix is rigidly ordered with respect to its stem. Although the phonological host of a phrasal affix may vary, the clitic is always realized in the same kind of the syntactic domain. Moreover, the attachment of phrasal affixes is often associated with (morpho)phonological or prosodic changes which affect either a clitic, or its host, or both. Such phenomena are, of course, common in the usual morphological domain of affixes and stems. In fact, if such (morpho)phonological changes accompany a clitic which is always attached to the same kind of a phonological host, e.g., a verb, the clitic is often considered a usual morphological affix, see Miller (1992a), Miller and Sag (1997), Monachesi (1995, 1999) for analyses of Romance pronominal clitics as verbal affixes.

Miller (1992a) distinguishes two classes of clitics: *phrasal affixes*, essentially the same as those in Anderson (1992), and *postlexical clitics* (PLCs), which are postlexically attached syntactic items (essentially words which lack prosodic autonomy). In this dissertation, the main discussion of the status of Polish clitics will use Miller's terminology.

Before concluding this subsection, let us present the orthogonal classification proposed in Cardinaletti and Starke (1994). They provide an apparently universal typology (mostly) of pronouns into basic abstract classes of *strong* and *deficient* forms. The criteria for this division involve: a morphological form, a series of syntactic criteria (the possibility of being modified or conjoined, distribution) and referential properties. More specifically, deficient items usually have a reduced morphological form with respect to strong forms; a deficient item cannot be used as a conjunct in coordination, be modified or appear in 'peripheral' (isolated) positions; a strong pronoun may have only a human reference whereas a deficient one may refer either to human or non-human entities.

The class of deficient forms is further split into two abstract subclasses of *weak* and *clitic* forms. Weak forms behave like full phrases whereas clitic forms are heads. In contrast to weak forms, clitic forms can be doubled by a full NP or form a cluster. There are certain restrictions on combinations of clitic forms but not on weak forms. Finally, weak forms, unlike clitic forms, may have lexical stress.

In our study, this typology will be used to identify certain syntactic properties of clitics. Since our main objective, however, is to establish the grammatical status of clitics, the main stress will be put on the distinction between word-like and affix-like properties of Polish clitics.

2.1.2 Polish Personal Pronouns

In Polish, pronominal arguments of a verb may occur in two forms: orthotonic (stressed, strong) or atonic (unstressed, weak, clitic). First and second person pronouns are unspecified for gender whereas the gender of third person pronouns is morphologically indicated. All personal pronouns are inflected for case. The corresponding inflectional paradigms are given in table 2.1 on p. 32 (atonic forms are indicated in *italic*).[2]

As the table in fig. 2.1 and the examples in (2.1)–(2.2) indicate, the two forms of pronominals can be morphologically distinct, cf. (2.1), but most Polish atonic pronominals double as strong forms, e.g., *nas* 'us' or *jej* 'her' in (2.2), cf. Rappaport (1988, p. 309, fn. 13), Spencer (1991, p. 367).

(2.1) a. Jan go / jego zobaczył.
John him$_{cl}$ him$_{strong}$ saw
'John saw him.'

b. Chętnie mu / jemu pomagamy.
willingly him$_{cl}$ him$_{strong}$ help-we
'We help him willingly.'

(2.2) a. Zaprosili nas na kolację.
invited-they us$_{cl/strong}$ on dinner
'They invited us for dinner.'

b. Chętnie jej pomożemy.
willingly her$_{cl/strong}$ will.help-we
'We will help her willingly.'

In Polish, an NP complement of a verb cannot be accompanied by a corresponding atonic pronominal (i.e., a pronominal clitic cannot be doubled by a full NP), see (2.3) based on

[2] The paradigm in table 2.1 presents pronominal arguments of a verb and does not include possessive pronouns (e.g., *nasza* 'our$_{fem,nom}$', *twój* 'your$_{nom,masc}$', etc.) or so-called *n*-forms of third person pronouns which occur only as arguments of prepositions (e.g., *niego* 'him$_{gen/acc}$', *nim* 'them$_{dat}$', etc.). The anaphoric pronoun *siebie* and its weak form *się* are discussed in Chapter 3.

pronoun	nom	gen	dat	acc	inst	loc
‘I’	ja	mnie	mnie, *mi*	mnie, *mię*	mnie	mnie
‘you’ (sing.)	ty	ciebie	tobie, *ci*	ciebie, *cię*	tobą	tobie
‘we’	my	nas	nam	nas	nami	nas
‘you’ (plural)	wy	was	wam	was	wami	was
‘he’	on	jego, *go*	jemu, *mu*	jego, *go*	nim	nim
‘she’	ona	jej	jej	ją	nią	niej
‘it’	ono	jego, *go*	jemu, *mu*	je	nim	nim
‘they’ (masc)	oni	ich	im	ich	nimi	nich
‘they’ (fem & neut)	one	ich	im	je	nimi	nich

Figure 2.1: Inflectional paradigm of personal pronouns

Witkoś (1998, p. 155).[3] There are no inherent pronominal clitics in Polish and a full NP and a pronominal clitic are always interchangeable.[4]

(2.3) a. Kocham go *(,) drania.
love-I him$_{cl,acc}$ bastard$_{acc}$
‘I love him, the bastard.’

b. * Wczoraj Maria dała mu Piotrowi list.
yesterday Mary gave him$_{cl,dat}$ Peter$_{dat}$ letter
‘Yesterday Mary gave him a letter.’ (intended)

2.1.3 Deficient and Strong Pronominals

Polish weak personal pronouns are clitics, i.e., they are prosodically deficient items and cannot bear heavy stress (marked with UPPER case in (2.4)), cf. Dłuska (1974, p. 21), Rappaport (1988, p. 309), Dukiewicz and Sawicka (1995, p. 172).

(2.4) a. Wczoraj Maria spotkała JEGO / *GO w parku.
yesterday Mary met him$_{strong}$ him$_{cl}$ in park
‘Yesterday Mary met HIM in the park.’

b. Maria pomogła JEMU / *MU wczoraj.
Mary helped him$_{strong}$ him$_{cl}$ yesterday
‘Mary helped HIM yesterday.’

Prosody is not the only property which distinguishes atonic and orthotonic pronominal forms. Cardinaletti and Starke (1994) provide an apparently universal classification of pronominals into two **abstract** groups of strong and deficient forms (the latter is further split into weak and clitic forms, see §2.1.1). If the criteria proposed in Cardinaletti and Starke (1994) are applied to Polish pronominals, orthotonic pronouns are strong forms, whereas atonic pronominals are deficient forms (without a clear distinction whether Polish atonic pronominals are weak or clitic forms in the terminology of Cardinaletti and Starke (1994), see Witkoś (1998, ch.3.2.2)).

[3] The ‘*(,)’ in (2.3a) indicates that the sentence is grammatical only if the comma is present.

[4] The behaviour of the reflexive clitic *się* in this respect is different; see Chapter 3.

Below we present contrasts between Polish atonic and orthotonic personal pronouns which follow from general properties of deficient and strong forms discussed in Cardinaletti and Starke (1994) (similar data are also considered in Witkoś (1996b, 1998) and Kupść (1999e)). In what follows, the term 'weak pronominal' is used informally as a synonym of 'pronominal clitic' or 'atonic pronoun' and they are not related to the terminology introduced in Cardinaletti and Starke (1994).

Use in Isolation Only strong forms can be used in isolation whereas deficient forms cannot. As (2.5) shows, pronominal clitics behave on a par with deficient forms.

(2.5) a. A: Kogo widział wczoraj Jan w parku? B: Jego. / *Go.
who saw yesterday John in park him$_{strong}$ him$_{cl}$
A: 'Who did John see in the park yesterday?' B: 'Him.' Witkoś (1998, p. 150)

b. A: Komu Maria pomogła? B: Jemu. / *Mu.
whom Mary helped him$_{strong}$ him$_{cl}$
A: 'Whom did Mary help?' B: 'Him.'

Coordination A deficient form cannot occur as a conjunct in a coordinated structure and the strong form must be used instead. As (2.6) shows, pronominal clitics behave like deficient forms.

(2.6) a. Matka spotkała jego / *go i Marię.
mother met him$_{strong}$ him$_{cl}$ and Mary
'Mother met him and Mary.'

b. Maria pomogła jemu / *mu i matce.
Mary helped him$_{strong}$ him$_{cl}$ and mother
'Mary helped him and the mother.'

Modification According to Cardinaletti and Starke (1994), deficient pronominals do not accept any kind of modification whereas strong pronouns may be modified by certain adverbs. Polish pronominal clitics act on a par with deficient pronominals: (2.7) are ungrammatical if the adverb *nawet* 'even' is treated as a modifier of the pronominal clitic.[5]

(2.7) a. Wczoraj nawet jego / #go widziałem w teatrze.
yesterday even him$_{strong}$ him$_{cl}$ saw-I in theater
'Yesterday I saw even him in the theater.' (intended)

b. Maria nawet jemu / #mu pomogła.
Mary even him$_{strong}$ him$_{cl}$ helped
'Mary helped even him.' (intended)

As observed in Witkoś (1998, p. 152), another property which distinguishes deficient and strong forms is that only the latter allow for constituent negation. This difference is respected by Polish pronominal clitics and the corresponding strong forms:

[5] The symbol '#' indicates that a sentence is ungrammatical in the intended interpretation.

(2.8) a. Maria nie jego / *go spotkała, lecz jego brata.
Mary not him$_{strong}$ him$_{cl}$ met but his brother
'Mary met not him but his brother.'

b. Maria nie jemu / *mu pomogła, lecz jego bratu.
Mary not him$_{strong}$ him$_{cl}$ helped but his brother
'Mary helped not him but his brother.'

Distribution Deficient forms have a more restricted distribution than strong forms. As (2.9) illustrates, Polish pronominal clitics cannot occur sentence initially, unlike their prosodically full equivalents.

(2.9) a. Jego / *Go Maria widziała wczoraj w parku.
him$_{strong}$ him$_{cl}$ Mary saw yesterday in park
'Mary saw him yesterday in the park.'

b. Jemu / *Mu Maria pomogła.
him$_{strong}$ him$_{cl}$ Mary helped
'Mary helped him.'

Further restrictions on the distribution of pronominal clitics are discussed in §2.2.

Reference Cardinaletti and Starke (1994) observe that deficient and strong forms have distinct referential properties. Deficient forms may have either human or non-human reference whereas strong forms refer only to human individuals. This observation seems to be confirmed by the behaviour of Polish personal pronouns, see (2.10) based on Witkoś (1998, p. 153).

(2.10) a. To jest dobry film. Widzieliśmy go / *jego wczoraj.
this is good film saw-we him$_{cl}$ him$_{strong}$ yesterday
'This is a good film. We saw it yesterday.'

b. To jest mój brat. Wiem, że go / jego wczoraj widzieliście.
this is my brother know-I COMP him$_{cl}$ him$_{strong}$ yesterday saw-you
'This is my brother. I know you saw him yesterday.'

Conclusion The data discussed above show that the behaviour of Polish pronominal clitics differs from their strong counterparts. Pronominal clitics are prosodically deficient items and never occur alone. Also syntactic properties distinguish pronominal clitics from other NPs: pronominal clitics cannot be modified, they do not occur as a conjunct in coordination and have a more restricted distribution. Finally, the referential properties of Polish personal pronouns seem to be distinct from strong pronouns.

2.2 Grammatical Status of Pronominal Clitics

Although Polish pronominal clitics are atonic forms, this does not tell us whether they are syntactic or morphological items. The grammatical status of clitics is usually difficult to establish since they have both word-like and affix-like properties. Below we apply various tests developed recently in the literature, e.g., Zwicky (1985), Zwicky and Pullum (1983), Miller

(1992a), in order to determine whether Polish pronominal clitics are syntactic items (postlexical clitics, PLC) or (inflectional or phrasal) affixes, applying the terminology introduced in §2.1.1.[6]

2.2.1 Prosody

As discussed in §2.1.3, Polish weak pronominals are atonic (unstressed) items, cf. (2.4), and they always require a phonological host to attach to, see (2.5).

Polish pronominal clitics do not occur in sentence initial position, (2.9), e.g., Spencer (1991, p. 367), Dukiewicz and Sawicka (1995, p. 174), Witkoś (1998, p. 151), Kupść (1999e, p. 111), Derwojedowa (1998, p. 88). This fact should be attributed to their enclitic status since the sentences in (2.11) are fully grammatical.

(2.11) a. Piotr powiedział, że go Maria widziała wczoraj w parku.
Peter said COMP him$_{cl}$ Mary saw yesterday in park
'Peter said that Mary saw him yesterday in the park.'

b. Piotr poprosił, żeby mu / mi Maria pomogła.
Peter asked COMP him$_{cl}$ me$_{cl}$ Mary helped
'Peter asked Mary to help him / me.'

Polish pronominal clitics do not form a prosodic word with the phonological host, cf. Dłuska (1974, p. 21), Rappaport (1988, p. 310).[7]

(2.12) a. POdaj Piotrowi książkę!
give Peter book
'Give Peter the book!'

b. POdaj mu ją!
give him$_{cl}$ her$_{cl}$
'Give it to him!'

c. PoDAję Piotrowi książkę.
give-I Peter book
'I give Peter the book.'

If a sequence of clitics follows a word, (2.12b), the regular penultimate lexical stress of the word is not affected, similarly to (2.12a). However, if an inflectional affix is attached to a stem, the lexical stress can be shifted, see (2.12c).

Note that the verb followed by the sequence of clitics in (2.12b) cannot be treated as an irregularly stressed word. In Polish, certain polysyllabic words, e.g., *Osiemset* 'eight hundred', *okOlica* 'surroundings', are irregularly stressed on the antipenultimate syllable, see Rubach and Booij (1985, p. 285).[8] The sequence *podaj mu ją* in (2.12b) consists of four syllables and,

[6]A similar set of criteria applied to Romance pronominals led Miller (1992a) and Monachesi (1995, 1999) to the conclusion that, respectively, French and Italian pronominal clitics are verbal affixes.

[7]The stressed syllable is indicated in UPPER case.

[8]The regular penultimate and the irregular antipenultimate stress are the only productive lexical stress patterns in Polish, see Dukiewicz and Sawicka (1995). Some polysyllabic words of a foregin origin retain their primary stress, e.g., *attachÉ*, *DŻENtelmen* 'gentleman', *WAszyngton* 'Washington', etc., see Rubach and Booij (1985, p. 285).

if treated as a single word, the antipenultimate lexical stress should fall on the syllable *daj*, contrary to fact.

Although Polish pronominal clitics are usually enclitics, cf. (2.9) vs. (2.11), under certain prosodic conditions they can become phonological proclitics to a verb, see Dłuska (1974, p. 29–31) and Rappaport (1988, p. 321). Arguably, this is possible in (2.13a), due to poetic rhythmics, or in (2.13b), where a clitic is placed at the left margin of the intonational phrase, cf. Rappaport (1988, p. 321).

(2.13) a. ten tylko się dowie, kto cię stracił
this only RM will.know who you$_{cl}$ lost
'only this one will know who lost you'

b. Mój kolega, który tam był, go odwiedził.
my colleague who there was him$_{cl}$ visited
'My colleague who was there visited him.'

Even if special prosodic conditions make a pronominal clitic a proclitic, it must remain unstressed if attached to a monosyllabic verb, in contrast to the verbal proclitic *nie* 'not', which is obligatorily stressed in such a case, compare (2.14a) vs. (2.14b) (see Chapter 4 for the behaviour of *nie*).

(2.14) a. Maria, jak twierdzą, go ZNA / *GO zna.
Mary as claim-they him$_{cl}$ knows him$_{cl}$ knows
'Mary, as they say, knows him.'

b. Maria Piotra NIE zna / *nie ZNA.
Mary Peter NM knows NM knows
'Mary doesn't know Peter.'

We conclude that, unlike affixes, pronominal clitics do not form a prosodic word with the item they are phonologically attached to.

2.2.2 Phonology

As observed in Rappaport (1988, p. 310) and Dukiewicz and Sawicka (1995, p. 174), there is a phonological word boundary (an inter-word voicing, external sandhi) between a pronominal clitic and its host. This fact is directly observable in southern and western dialects of Polish. As discussed in Rappaport (1988, p. 310), the last consonant of the word *brat* in (2.15), in these dialects, is pronounced as [d], which is characteristic for inter-word voicing but not for the word-internal pronunciation (i.e., [t]).

(2.15) Brat mi się podoba.
brother me$_{cl}$ RM likes
'I like the brother.'

Such a behaviour clearly indicates that pronominal clitics do not form a phonological word with the phonological host. Therefore, as far as phonology is concerned, Polish pronominal clitics should be considered separate syntactic items (PLCs).

2.2.3 Ellipsis

Just like syntactic complements, (2.16a), Polish pronominals can be elided, cf. (2.16b).

(2.16) a. A: Przedstawisz Piotra Marii? B: Przedstawię.
will-introduce-you$_{sg}$ Peter to-Mary will-introduce-I
A: 'Will you introduce Peter to Mary?' B: 'I will.'

b. A: Przedstawisz mi go? B: Przedstawię.
will-introduce-you$_{sg}$ me$_{cl}$ him$_{cl}$ will-introduce-I
A: 'Will you introduce him to me?' B: 'I will.'

The possibility of ellipsis is characteristic of syntactic items but it is unusual for affixes, cf. Bresnan and Mchombo (1995).

We take the data in (2.16) to strongly support the syntactic status (PLC) of Polish weak pronominals.

2.2.4 Promiscuity

The promiscuous attachment of Polish pronominal clitics usually is not explicitly dealt with. It is mentioned in Dukiewicz and Sawicka (1995, p. 173) that enclitics (hence, pronominal clitics as well) may attach to different expressions in the sentence but no examples are given.

The phonological host for the clitic may be a verb, (2.12b) or (2.16b), a noun, (2.17a), an adjective, (2.17b), a complementizer, (2.17c), an adverb, (2.17d), or a conjunction, (2.17e).[9]

(2.17) a. Maria go wczoraj zaprosiła na obiad.
Mary him$_{cl}$ yesterday invited on lunch
'Mary invited him for lunch yesterday.'

b. Trudne go zadanie czekało.
difficult him$_{cl}$ task waited
'A difficult task waited for him.'

c. Jan powiedział, że go nigdy nie przeprosi.
John said COMP him$_{cl}$ never NM will.apologize-he
'John said he would not ever apologize to him.'

d. Często go teraz spotykam.
often him$_{cl}$ now meet-I
'I often meet him nowadays.'

e. Pracował tam wiele lat i go za tę pracę ceniono.
worked-he there many years and him$_{cl}$ for this work appreciated
'He worked there many years and he was very appreciated for his work.'

Two restrictions on the attachment of pronominal clitics are noticed in Spencer (1991, p. 367). Spencer (1991) notes that *nie* 'not' and prepositions never host clitics in Polish. The latter observation can be falsified by (2.18) if *dookoła* 'around' is treated as a complementless preposition rather than an adverb-like element.

[9] As said in §2.2.1, a pronominal clitic can become a proclitic only to a **verb**. As in none of (2.17) a clitic immediately precedes a verb, all clitics exemplified here are enclitics.

(2.18) Dookoła go starannie obszedł.
around him$_{cl}$ carefully walked
'He walked around it/him carefully.'

It is true, however, that a pronominal clitic cannot be used as a complement of a preposition, see (2.19) (for third person pronouns, the usual strong pronoun cannot be used either and a different pronominal form must appear, cf. (2.19a) and fn. 2.).

(2.19) a. Jan napisał do niego / *jego / *go list.
John wrote to him$_{gen}$ him$_{gen}$ him$_{gen,cl}$ letter
'John wrote him a letter.'

b. Maria liczyła na mnie / *mię / ciebie / *cię.
Mary counted on me$_{acc}$ me$_{acc,cl}$ you$_{acc}$ you$_{acc,cl}$
'Mary counted on me / you.'

Since similar restrictions are observable in other languages as well, see Cardinaletti and Starke (1994), we attribute the impossibility of pronominal clitics as complements of a preposition to general properties of deficient pronominals.

The sentential negative marker (NM) *nie* cannot host clitics, (2.20), see Spencer (1991, p. 367–368), Witkoś (1998, p. 156).

(2.20) * Nie go / cię / mię widzę.
NM him$_{cl}$ you$_{cl}$ me$_{cl}$ see-I
'I can't see him / you / me.'

We explain the ungrammaticality of (2.20) by the fact that, in contrast to Polish pronominal clitics, the sentential NM *nie* is a verbal prefix; for details, see Chapter 4 as well as Witkoś (1998, ch. 4) and Kupść and Przepiórkowski (1997).

As observed in Witkoś (1998, p. 152), also the homophonous constituent negation *nie* cannot host pronominal clitics, cf. (2.8) repeated below.

(2.8) a. Maria nie jego / *go spotkała, lecz jego brata.
Mary not him$_{strong}$ him$_{cl}$ met but his brother
'Mary met not him but his brother.'

b. Maria nie jemu / *mu pomogła, lecz jego bratu.
Mary not him$_{strong}$ him$_{cl}$ helped but his brother
'Mary helped not him but his brother.'

As stated in §2.1.3, we take such contrasts as yet another property which distinguishes deficient and strong pronominal forms.

Contrary to criterion A in Zwicky and Pullum (1983), the promiscuous phonological attachment is not a unique property of simple clitics (i.e., essentially PLCs), see Nevis (1985), Zwicky (1987), Miller (1992a, ch. 4.2). Phrasal affixes (PAs) are always realized within the same syntactic phrase but the phonological host may vary since the PA is not attached to the head of the phrase (unlike inflectional affixes). For example, the English possessive *'s* (a PA, according to Zwicky (1987), Miller (1992a)) is always realised at the right egde of an NP, (2.21), but its phonological host is any item which happens to be the final element of the NP: a noun, (2.21a), a preposition, (2.21b), a verb, (2.21c), etc., Miller (1992a, p. 110).

(2.21) a. the boy's book

b. the boy I spoke to's book

c. the boy I like's book

Therefore, the promiscuous phonological attachment of pronominal clitics does not *a priori* preclude them from being PAs or PLCs. The promiscuity of Polish pronominal clitics presents inconclusive evidence of their grammatical status.

2.2.5 Placement

Polish pronominal clitics are infamous for their relatively free distribution, see Spencer (1991, pp. 367ff), Rappaport (1988, p. 310), Derwojedowa (1998, p. 96), Franks (1998, ch. 5.2), Kupść (1999e, pp. 111ff), and especially Witkoś (1996b) and Witkoś (1998, ch. 3). As indicated in (2.22), the clitic may occur almost anywhere before the verb (as discussed above, sentence initial positions, e.g., (2.9), are excluded by prosodic rather than syntactic factors).

(2.22) a. Dziewczyna mu / mi cierpliwie wszystko wytłumaczyła.
girl him$_{cl}$ me$_{cl}$ patiently everything explained
'The girl patiently explained him / me everything.'

b. Dziewczyna cierpliwie mu / mi wszystko wytłumaczyła.

c. Dziewczyna cierpliwie wszystko mu / mi wytłumaczyła.

Note that clitics may appear in phrase internal positions as well, (2.23a) and (2.24a) (see §3.2.6 for similar examples including the reflexive marker *się*). In Polish, unlike in Serbian/Croatian or Czech, this is not a unique property of clitics. As observed by Borsley and Rivero (1994, p. 397), Polish NPs and PPs can be split by full prosodic words, see (2.23b) and (2.24b).

(2.23) a. [$_{NP}$ Żadnego mu / mi zadania] nie wytłumaczyłaś.
none him$_{cl}$ me$_{cl}$ exercise NM explained-you$_{sg}$
'You did not explain him any exercise.'

b. [$_{NP}$ Żadnego Piotrowi zadania] nie wytłumaczyłaś.
none Peter exercise NM explained-you$_{sg}$
'You did not explain any exercise to Peter.'

(2.24) a. [$_{PP}$ Z tym go mężczyzną] umówiła.
with this him$_{cl}$ man set-she a date
'She set him a date with this man.'

b. [$_{PP}$ Z tym umówiła mężczyzną] Piotra.
with this set-she a date man Peter
'She set Peter a date with this man.'

c. * [$_{PP}$ Z umówiła [$_{NP}$ tym mężczyzną]] Piotra.
with set-she a date this man Peter

(2.24c) illustrates a restriction on PPs split noticed in Borsley and Rivero (1994): a PP cannot be split right after a preposition. This fact can be partially explained by prosodic properties of most Polish prepositions. As shown in Rubach and Booij (1985), many prepositions in Polish are prosodically deficient and form a prosodic word with their complement. Note, however, that a similar ban on PP-split seems to hold also for PPs headed by prosodically full prepositions, e.g., *obok* 'next to', cf. (2.25).

(2.25) * Jan obok go nowego kolegi posadził.
John next him$_{cl}$ new colleague sat
'John gave him a place next to a new colleague.' (putatively)

As an intransitive preposition can be followed by a clitic, see (2.18), we will assume that in PPs headed by a transitive preposition, (2.24)–(2.25), the preposition always forms an unseparable unit with the first element of its NP complement.

The clause boundary constitutes a barrier for the leftward preverbal clitic placement: in contrast to (2.11) and (2.17c), the examples in (2.26) are ungrammatical.

(2.26) a. * Piotr powiedział go, że Maria widziała wczoraj w parku.
Peter said him$_{cl}$ COMP Mary saw yesterday in park
'Peter said that Mary saw him yesterday in the park.' (putatively)

b. * Piotr poprosił mu / mi, żeby Maria pomogła.
Peter asked him$_{cl}$ me$_{cl}$ COMP Mary helped
'Peter asked Mary to help him / me.' (putatively)

As (2.27) indicates, clitics may occur also postverbally.

(2.27) Dziewczyna cierpliwie wytłumaczyła mu / mi wszystko.
girl patiently explained him$_{cl}$ me$_{cl}$ everything
'The girl patiently explained him / me everything.'

Unlike other NPs, a postverbal pronominal clitic can be separated from the verb only by another clitic, (2.28a). Neither a strong pronoun, cf. (2.28b), nor a phrase, (2.28c), can be inserted between a postverbal clitic and the verb.

(2.28) a. Jan wytłumaczył mu go cierpliwie.
John explained him$_{dat,cl}$ him$_{acc,cl}$ patiently
'John patiently explained it to him.'

b. * Jan wytłumaczył jemu go cierpliwie.
John explained him$_{dat}$ him$_{acc,cl}$ patiently
'John patiently explained it to him.'

c. * Jan wytłumaczył [nowemu koledze]$_{NP}$ go cierpliwie.
John explained new colleague him$_{acc,cl}$ carefully
'John patiently explained it to a new colleague.'

Polish deverbal derivatives, i.e., gerunds (deverbal nouns ending on *-nie*, *-cie*) and participles, also can have pronominal clitic arguments, see (2.29a) and (2.30a).

(2.29) a. Jan zażądał przyniesienia mu / mi kawy.
John demanded bringing$_{gerund}$ him$_{cl}$ me$_{cl}$ coffee
'John demanded to bring him / me coffee.'

b. * Jan zażądał mu / mi przyniesienia kawy.
John demanded him$_{cl}$ me$_{cl}$ bringing$_{gerund}$ coffee
'John demanded to bring him / me coffee.' (putatively)

(2.30) a. Na głowie położywszy mu / mi dłonie, spojrzał w górę.
on head having-put him$_{cl}$ me$_{cl}$ hands looked-he up
'Having put him / me hands on the head, he looked up.'

b. ? Na głowie mu / mi położywszy dłonie, spojrzał w górę.
on head him$_{cl}$ me$_{cl}$ having-put hands looked-he up
'Having put him / me hands on the head, he looked up.'

The distribution of Polish pronominal clitics in gerund phrases is more rigid than in simple clauses or participle phrases: (2.30b) sounds much better than (2.29b) and would be perfectly acceptable in poetry. If a pronominal clitic is an argument of a gerund, the clitic must immediately follow the deverbal form (possibly preceded by other clitics), see (2.29), unlike in (2.22), (2.23) or (2.30), where a clitic may precede a verbal element as well. As in the case of simple verbs, in contrast to non-clitics, clitic arguments of a gerund or a participle must strictly succeed the deverbal form, see (2.31).

(2.31) a. Jan zażądał przyniesienia kawy Piotrowi / *mu / *mi.
John demanded bringing$_{gerund}$ coffee Peter him$_{cl}$ me$_{cl}$
'John demanded to bring Peter / him / me coffee.' (intended)

b. Na głowie położywszy dłonie Piotrowi / *mu / *mi, spojrzał w górę.
on head having-put hands Peter him$_{cl}$ me$_{cl}$ looked-he up
'Having put Peter / him / me hands on the head, he looked up.' (intended)

Note also that, unlike in many other languages, the relative order of Polish pronominal clitics is not fixed. Although Rappaport (1988) claims that pronominal clitics in Polish occur in a fixed order, i.e., *dat*<*się*<*acc*<*gen*, this order is just a preference but not a requirement.[10] In (2.32b), the order of clitics is reversed with respect to (2.32a) but both sentences are grammatical.

(2.32) a. Wczoraj mu go przedstawiła.
yesterday him$_{dat,cl}$ him$_{acc,cl}$ introduced-she
'She introduced him to him yesterday.'

b. Wczoraj go mu przedstawiła.

As observed in Rappaport (1988), if there are several pronominal clitics, either all clitics occur preverbally, cf. (2.33), or none, (2.28a).[11]

(2.33) a. Brat mi się podoba.
brother me$_{cl}$ RM$_{cl}$ likes
'I like the brother.'

b. * Brat mi podoba się.

Finally, unlike in many other languages, Polish pronominal clitics do not have to cluster, Spencer (1991, p. 369), Franks (1998, p. 48), Witkoś (1998, p. 158).

[10] Also Witkoś (1998, p. 158) observes a tendency of clitics to occur in *dat*<*acc* order.

[11] In fact, certain native speakers do not find (2.33b) fully ungrammatical, in contrast to our judgement. All informants, however, consider (2.33b) degraded with respect to (2.33a).

(2.34) Kiedy go wreszcie mu odebraliśmy...
when him$_{acc,cl}$ finally him$_{dat,cl}$ took-away-we
'When we have finally took it away from him...'

The examples above show that although the distribution of pronominal clitics is more restricted than the distribution of other NPs, cf. (2.11a) and (2.28b–c), clitic placement is nevertheless relatively free, see (2.22), (2.23), (2.27). Such a mobility is characteristic of syntactic items but not of (inflectional or phrasal) affixes.[12]

We conclude that the distribution of Polish pronominal clitics favours their syntactic treatment as PLCs.

2.2.6 Coordination

Miller (1992a,b) proposes a test based on coordination to distinguish postlexical clitics (PLCs) from phrasal affixes (PAs). The test consists of three general criteria listed below.

> Criterion 1: "An item which cannot have a wide scope over a coordination of hosts cannot be a postlexical clitic and must be an affix." Miller (1992a, p. 155))

As (2.35) shows, a pronominal clitic can have a wide scope over coordination of its hosts.

(2.35) a. Jan go umył i uczesał.
John him$_{cl}$ washed and brashed
'John washed him and brashed his hair.'

b. Jan mu nie pomógł, ale i nie przeszkodził.
John him$_{cl}$ NM helped but and NM disturbed
'John did not help him but he did not disturb him, either.'

Criterion 1 does not exclude the PLC analysis of pronominal clitics. On the other hand, the criterion does not force such an analysis, either. PAs may have wide scope over coordination of hosts as well, see Zwicky (1987) or Miller (1992a,b). Morphological affixes, however, do not have wide scope over coordination:

(2.36) *przed*wojenny i kolonialny ≠ *przed*wojenny i *przed*kolonialny
pre-war and colonial pre-war and pre-colonial

> Criterion 2: "A postlexical clitic can be repeated on each conjunct of its host if and only if it forms a constituent with its host in the syntax in the corresponding non-coordinate structure." Miller (1992a, p. 156)

As (2.37) illustrates, a pronominal clitic may be repeated in a coordinated structure.

(2.37) a. Jan go umył i go uczesał.
John him$_{cl}$ washed and him$_{cl}$ brashed
'John washed him and quickly brashed his hair.'

b. Jan mu nie pomógł, ale i mu nie przeszkodził.
John him$_{cl}$ NM helped but and him$_{cl}$ NM disturbed
'John did not help him but he did not disturb him, either.'

[12] A discussion of clitic placement in verbal complexes is postponed to §2.5 where we consider clitic climbing.

According to Criterion 2, Polish pronominal clitics can be considered PLCs. This criterion is based on the assumption that only constituents can be coordinated, e.g., Chomsky (1957). Since non-constituent coordination is possible in Polish as well, see (2.38) (more examples can be found in Kallas (1993)), the validity of this criterion for Polish might be questioned.

(2.38) Jan dał Marysi płytę, a Zosi książkę.
John gave Mary record and Sophia book

'John gave Mary a record and Sophia a book.'

Moreover, as discussed in Miller (1992a), PAs can be repeated in coordination as well. Therefore, we take Criterion 2 to be neutral with respect to the status of Polish pronominal clitics.

> Criterion 3: "In cases where repetition is obligatory in each conjunct, then the item is necessarily an affix and cannot be a postlexical clitic." Miller (1992a, p. 157)

As we have seen in (2.35) and (2.37), the clitic is repeated in coordination only optionally. Therefore, Criterion 3 does not imply that the clitic is an affix.

We conclude that coordination data are inconclusive for the PLC vs. PA status of Polish pronominal clitics.

2.2.7 Conclusion

The data discussed in this section show rather clearly that Polish pronominal clitics are syntactic items. This conclusion is strongly supported by prosody, phonology and ellipsis facts. The promiscuous attachment of clitics and coordination data are neutral with respect to this hypothesis. The distribution of pronominal clitics is more restricted than that of other NPs but clitic placement is not fixed, which is atypical for affixes.

We conclude that Polish pronominal clitics are syntactic items, i.e., postlexical clitics.

2.3 Representation of Clitics

2.3.1 Subsorts of *synsem*

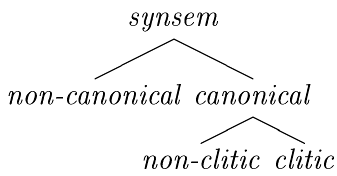
As argued in the previous section, Polish pronominal clitics are syntactic items. Apart from distinct prosodic properties, syntactic and semantic behaviour of clitics distinguishes them from other nominals, see §2.1.3. In HPSG, syntactic and semantic properties of a linguistic object are encoded in the *synsem* component of a *sign*. Thus, this is a natural place to specify the clitic vs. non-clitic distinction, see Miller and Sag (1997) and Monachesi (1999) for a similar solution proposed for French and Italian pronominal clitics.

We adopt the split of *synsem* into *canonical* and *non-canonical* sorts postulated in Sag (1997), Bouma *et al.* (2000), Miller and Sag (1997), Abeillé *et al.* (1998), among others. Objects of sort *canonical* appear in all *signs*, whereas *non-canonical* encodes local overtly unrealised arguments, e.g., *gaps* (as in Sag (1997)), or French pronominal affixes, cf. Miller and Sag (1997), Abeillé *et al.* (1998).[13]

[13] The partition of *synsem* in Monachesi (1999) is different.

Polish pronominal clitics are syntactic items and, thus, they are *signs*. We introduce the sorts *clitic* (for syntactic clitics) and *non-clitic* (for prosodically autonomous syntactic objects) as two subsorts of the sort *canonical*, see (2.39).[14]

(2.39)



Unlike Miller and Sag (1997), Abeillé *et al.* (1998), Monachesi (1999) or Kupść (1999a), we do not introduce separate subsorts corresponding to pronominal clitics. Since the clitic vs. non-clitic distinction is orthogonal to being a personal pronoun, these two pieces of information should be kept separate. In our setup, all pronominals have a uniform representation as NPs (the CONT value of nominal objects is discussed in §2.3.2) but their *synsem* components may specify objects of distinct sorts.[15]

Before discussing the representation of pronominal clitics in detail, let us briefly present the general HPSG encoding of nominal objects.

2.3.2 Nominal Objects in HPSG

We adopt the HPSG representation of NPs proposed in Pollard and Sag (1994). Every nominal object (a noun, a pronoun, a quantifier, etc.) has a CONTENT value of the sort *nominal object* (*nom-object*, a subsort of *content*, see §1.3.1). The attributes appropriate to the sort *nom-object* are given below, see Pollard and Sag (1994, p. 399).

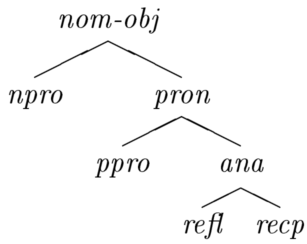
(2.40) $$\begin{bmatrix} \textit{nom-object} \\ \textsc{index}\ \textit{index} \\ \textsc{restr}\ \textit{set(psoa)} \end{bmatrix}$$

The sort *nom-object* is split into subsorts according to the properties of the NPs whose CONT values they describe. At the top of the hierarchy of *nom-object*s, nonpronominal (*npro*) and pronominal objects (*pron*) are distinguished; pronominal objects, in turn, are either personal pronouns (*ppro*) or anaphora (*ana*). Finally, there are two subsorts for two kinds of anaphors: reflexive (*refl*) and reciprocal (*recp*). The hierarchy of nominal objects, as used in HPSG, is presented in (2.41).

[14] Since subsorts of the sort *non-canonical* mentioned above, e.g., *gap*s in Sag (1997), are not the main topic of our interest, we do not specify them in (2.39).

[15] In order to prevent 'regular' third person pronouns, such as *jego* 'him$_{acc/gen}$' (2.19a) on p. 38, from occurring as complements of prepositions, a more detailed analysis is necessary. For example, one can introduce a boolean-valued attribute POST-PREP appropriate to the sort *noun*. These nominals which can appear only as complements of a preposition, e.g., *niego* 'him$_{acc/gen}$' in (2.19a), will be lexically specified as [POST-PREP +] whereas pronouns $jego_{strong}/go_{clitic}$ 'him$_{acc/gen}$', which cannot be complements of a preposition, will have the value of POST-PREP resolved to '−' in the lexicon. Nominals which can be either arguments of a preposition or a verb do not have the value of POST-PREP fixed in the lexicon. General principles will ensure that verbs must have ¬[POST-PREP +] arguments (hence, they can have also non-nominal arguments for which the POST-PREP attribute is not declared) whereas prepositions require [POST-PREP +] arguments.

(2.41)



The INDEX attribute in (2.40) is used as a (co-)reference marker, e.g., in order to resolve the (co-)reference of pronouns or anaphora (binding theory). The possible values of INDEX are objects the sort *index*, which can be either *non-referential* (e.g., the English expletive pronouns *it* and *there*) or *referential* (*ref*), see Pollard and Sag (1994, p. 397). The *ref* sort is used for nominals with a non-empty denotation, e.g., *John, herself, she, every person*, etc.

There are three attributes appropriate to the sort *index*, (2.42), see Pollard and Sag (1994, p. 399).[16]

(2.42) $$\begin{bmatrix} \textit{index} & \\ \text{PERSON} & \textit{person} \\ \text{NUMBER} & \textit{number} \\ \text{GENDER} & \textit{gender} \end{bmatrix}$$

If two referential NPs are co-indexed, their INDEX values are token-identical. Hence, not only must these NPs have the same denotation (or range over the same set of entities, as in the case of quantifiers) but, due to the internal structure of *index*, they must also agree in number, person and gender.

The denotation of a nominal object can be restricted via additional semantic specifications (anchoring conditions) encoded as the value of the RESTR (RESTRICTION) attribute, see (2.40). In this study, the impact of this attribute will be mostly ignored.

2.3.3 Lexical Entries

We represent pronominal clitics as syntactic nominal objects with *clitic* object as their SYNSEM value. This allows clitics to retain usual characteristics of NPs, such as referential properties, case, person, number, gender, etc. Since we do not attempt to incorporate prosodic or phonological information into our account, the main difference between lexical entries of a strong pronoun and its clitic counterpart is encoded in (the components of) their SYNSEM values.

The relevant parts of lexical entries for the clitic *go* 'it/him' and the strong pronoun *jego* 'him' are given in (2.43) and (2.44), respectively.

[16] As argued in Saloni and Świdziński (1985), and especially in Czuba and Przepiórkowski (1995) and Czuba (1997), the gender and number information in Polish can be represented by a single morphosyntactic GENDER attribute. For expository reasons, we retain the traditional specification.

$$(2.43)\quad \left[\begin{array}{l} \textit{word} \\ \text{PHON}\ \langle \text{go} \rangle \\ \text{SYNSEM} \left[\begin{array}{l} \textit{clitic} \\ \text{LOC} \left[\begin{array}{l} \textit{local} \\ \text{CAT} \left[\begin{array}{l} \textit{category} \\ \text{HEAD} \left[\begin{array}{l} \textit{noun} \\ \text{CASE}\ \textit{gen} \vee \textit{sacc} \end{array}\right] \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ}\ \langle\rangle \\ \text{COMPS}\ \langle\rangle \end{array}\right] \\ \text{ARG-ST}\ \langle\rangle \end{array}\right] \\ \text{CONT} \left[\begin{array}{l} \textit{ppro} \\ \text{INDEX} \left[\begin{array}{l} \textit{ref} \\ \text{NUMBER}\ \textit{sing} \\ \text{PERSON}\ \textit{third} \\ \text{GENDER}\ \textit{masc} \vee \textit{neut} \end{array}\right] \\ \text{RESTR}\ \{\} \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

$$(2.44)\quad \left[\begin{array}{l} \textit{word} \\ \text{PHON}\ \langle \text{jego} \rangle \\ \text{SYNSEM} \left[\begin{array}{l} \textit{non-clitic} \\ \text{LOC} \left[\begin{array}{l} \textit{local} \\ \text{CAT} \left[\begin{array}{l} \textit{category} \\ \text{HEAD} \left[\begin{array}{l} \textit{noun} \\ \text{CASE}\ \textit{gen} \vee \textit{sacc} \end{array}\right] \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ}\ \langle\rangle \\ \text{COMPS}\ \langle\rangle \end{array}\right] \\ \text{ARG-ST}\ \langle\rangle \end{array}\right] \\ \text{CONT} \left[\begin{array}{l} \textit{ppro} \\ \text{INDEX}\ \boxed{1} \left[\begin{array}{l} \textit{ref} \\ \text{NUMBER}\ \textit{sing} \\ \text{PERSON}\ \textit{third} \\ \text{GENDER}\ \textit{masc} \vee \textit{neut} \end{array}\right] \\ \text{RESTR}\ \left\{ \left[\begin{array}{l} \textit{psoa} \\ \text{QUANTS}\ \langle\rangle \\ \text{NUCLEUS} \left[\begin{array}{l} \textit{human-rel} \\ \text{INST}\ \boxed{1} \end{array}\right] \end{array}\right] \right\} \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

As (2.43) and (2.44) show, both *go* and *jego* are personal pronouns, [CONT *ppro*], assigned either genitive or accusative case (see the table on p. 32). As said in §1.3.2, in addition to a morphological specification of case values, we adopt the orthogonal classification into structural (context-sensitive) vs. lexical (inherent) case proposed for Polish in Przepiórkowski (1996, 1999). According to his analysis, genitive can be either a lexical or a structural case in Polish (the sort *gen* in (2.43) and (2.44) is the immediate supersort of *lgen* (lexical genitive) and *sgen* (structural genitive) but accusative is a structural case in Polish, hence *sacc* in (2.43) and (2.44).[17] From now on, if necessary, names of morphological cases are preceded by the appropriate prefix (i.e., *s* or *l*) to indicate whether the case is structural or lexical, e.g., *sacc*. We use NP[*str*] to indicate that the case of the NP is structural if the morphological case value can be successfully resolved via the Case Principle proposed for Polish in Przepiórkowski (1999, ch. 5).

[17]See, however, Przepiórkowski (1999, p. 104ff) for some evidence of a marginal presence of lexical accusative in Polish.

Note that the SYNSEM values in (2.43) and (2.44) are distinct: the clitic *go* has a *clitic* object as its SYNSEM value whereas *jego* has an object of sort *non-clitic* as its SYNSEM value. Pronominal clitics may have human or non-human reference, see §2.1.3, and the value of RESTR is empty in (2.43). On the other hand, strong personal pronouns refer only to human individuals, cf. §2.1.3; thus, the RESTR in (2.44) is not empty. In (2.44), the object of sort *psoa* (*parametrized-state-of-affairs*, a subsort of *content*, see Pollard and Sag (1994, p. 397)) appears in a complete form as specified in Pollard and Sag (1994, p. 399). Since quantification does not concern us here, from now on we will drop the attributes QUANTS and NUCLEUS in objects of sort *psoa*. The value of RESTR in (2.44) is simplified to $\{ \begin{bmatrix} \textit{human-rel} \\ \textsc{inst}\ \boxed{1} \end{bmatrix} \}$. We assume, as in Pollard and Sag (1994, p. 399), that semantic relations, such as *human-rel* in (2.44), are subsorts of *qfpsoa* (*quantifier-free-parametrized-state-of-affairs*), which is the direct subsort of *object*, see §1.3.1.

2.3.4 Clitics on COMPS

Polish pronominal clitics are syntactic items and they are interchangeable with other NP complements of a verb.[18] We represent pronominal clitics on a par with other syntactic complements, i.e., as elements of the verb's COMPS list. Let us consider a partial lexical entry of the verb *pomagać* 'to help' in (2.45).[19]

(2.45)
$$\begin{bmatrix} \textit{word} \\ \textsc{phon}\ \langle \text{pomagać} \rangle \\ \textsc{ss}|\textsc{loc} \begin{bmatrix} \textsc{cat} \begin{bmatrix} \textsc{head}\ \textit{verb} \\ \textsc{val} \begin{bmatrix} \textsc{subj}\ \langle \boxed{1} \rangle \\ \textsc{comps}\ \langle \boxed{2} \rangle \end{bmatrix} \\ \textsc{arg-st}\ \langle \boxed{1}\text{NP}_{\boxed{3}}[\textit{str}] \rangle \oplus \langle \boxed{2}\text{NP}_{\boxed{4}}[\textit{ldat}] \rangle \end{bmatrix} \\ \textsc{cont} \begin{bmatrix} \textit{help-rel} \\ \textsc{helper}\ \boxed{3} \\ \textsc{helped}\ \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The value of COMPS is a list of *synsem* objects, cf. §1.3.1. Thus, both *clitic* and *non-clitic* objects can be members of this list; unlike in Monachesi (1999), no special mechanisms are necessary to put clitics on the COMPS list. Since we assume that the ARG-ST list is the concatenation ('$\oplus$') of valence lists (cf. §1.3.1), clitics are present also on ARG-ST. Hence, they are subject to general principles encoded on this list. In particular, the case principle proposed for Polish in Przepiórkowski (1999) or the binding theory of Marciniak (1999) are valid for clitics as well. Note, however, that unlike in the analyses of French pronominal clitics, cf. Miller and Sag (1997), Abeillé *et al.* (1998), the ARG-ST list is not the only locus of Polish clitics since they are syntactic items.

In Pollard and Sag (1994), the value of the CONT attribute of a verb is an object of the sort *psoa*. As said above, we slightly simplify the feature geometry of *psoa* objects: in (2.45), an object specifying the semantic relation *help-rel* is used directly as the CONT value of the verb. As in Pollard and Sag (1994), semantic roles (specified by the attributes of *help-rel* in (2.45)) are assigned to indices of NP arguments of the verb ('$\text{NP}_{\boxed{1}}$' abbreviates the synsem

[18] As observed in §2.2.4, pronominal clitics cannot appear as complements of prepositions, which seems to be a general property of deficient forms, see Cardinaletti and Starke (1994).

[19] From now on, we use the term 'partial lexical entry' to indicate that a description does not contain all necessary lexical information; for example in (2.45), in the complete entry at least the VFORM or CONTEXT values (cf. 1.3.1) should be described as well.

component of a noun phrase with $\boxed{1}$ as the LOCAL|CONTENT|INDEX value; the meaning of the 'NP' abbreviation is given in §1.3.1). In (2.45), both NP arguments of the verb, i.e., $\boxed{1}$ and $\boxed{2}$, are assigned semantic roles: the values of HELPER and HELPED attributes are token identical with indices $\boxed{3}$ and $\boxed{4}$ of the NP arguments.[20] Therefore, if a complement is realised as a pronominal clitic, the clitic will also be a semantic argument of the verb.[21]

2.3.5 Realisation of Clitics

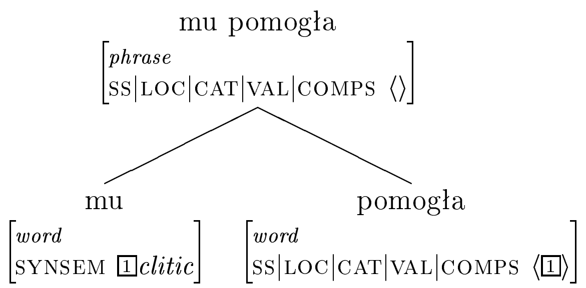
Being syntactic objects present on the COMPS list, Polish pronominal clitics are realised via general syntactic principles, i.e., the Valence Principle and ID schemata, see §1.3.1. No special mechanisms are necessary to ensure realisation of clitics, unlike in the HPSG analyses of Romance pronominal clitics in Miller and Sag (1997), Abeillé *et al.* (1998) (for French) and Monachesi (1995, 1999) (for Italian).

Assuming the lexical entry for the verb *pomagać* 'to help' as in (2.45), partial syntactic analyses of (2.46a–b) are given in (2.47) and (2.48).

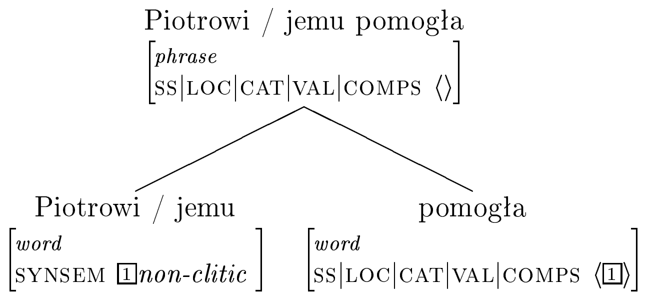
(2.46) a. Maria mu pomogła.
Mary him$_{cl}$ helped
'Mary helped him.'

b. Maria Piotrowi / jemu pomogła.
Mary Peter him helped
'Mary helped Peter/ him.'

(2.47)



(2.48)



The COMPS value of the verb *pomagać* in (2.45) is a list of *synsem*s and the complement of the verb can be realised either as a *clitic* or a *non-clitic* object. A pronominal clitic satisfies the syntactic requirement of the verb just like any other NP. Since *clitic* objects correspond to *sign*s, the verb combines with the clitic in the same way as with a 'regular' non-clitic

[20] In Chapter 3, the assignment of semantic roles is slightly different.

[21] As we will see in Chapter 3, our treatment of the reflexive clitic *się* differs in this respect.

complement. The appropriate immediate dominance schema (see SCHEMA2 in §1.3.1) licenses the verbal phrase with a clitic, (2.47), or a non-clitic complement, (2.47b). The Valence Principle, cf. (1.10), ensures that once the subcategorized element is realised, it is removed from the COMPS list of the phrase: the COMPS values of resulting phrases in (2.47) and (2.48) are empty.

2.4 Linearization

As we said in §1.3.1, HPSG schemata specify immediate dominance but the linear order is encoded separately. In this section, we adopt the idea of order domains of Reape (1992) and Kathol (1995) to sketch an account of the distribution of Polish pronominal clitics presented in §2.2.5.[22] Although the linearization approach advocated in this section might account for Polish word order in general, this task is far beyond the scope of this study.

2.4.1 Linear Order in HPSG

As we have seen in §2.2.5, word order in Polish is relatively free. In particular, syntactic constituents do not have to be linearly contiguous and can be split by external elements, as shown in (2.23) and (2.24) (repeated below for convenience).

(2.23) a. [$_{NP}$ Żadnego mu / mi zadania] nie wytłumaczyłaś.
none him$_{cl}$ me$_{cl}$ exercise NM explained-you$_{sg}$
'You did not explain him any exercise.'

b. [$_{NP}$ Żadnego Piotrowi zadania] nie wytłumaczyłaś.
none Peter exercise NM explained-you$_{sg}$
'You did not explain any exercise to Peter.'

(2.24) a. [$_{PP}$ Z tym go mężczyzną] umówiła.
with this him$_{cl}$ man set-she a date
'She set him a date with this man.'

b. [$_{PP}$ Z tym umówiła mężczyzną] Piotra.
with this set-she a date man Peter
'She set Peter a date with this man.'

In order to capture such data we need a mechanism which orders not only syntactic constituents but also allows them to be interleaved. For this purpose, we adopt order domains of Reape (1992) and Kathol (1995). More specifically, we follow Reape (1992) and Kathol (1995) and use the DOMAIN attribute (DOM, appropriate to *signs*), to indicate the order domain associated with a *sign* and encode linear order. The value of DOMAIN is a list of *dom_ obj* (*domain_ object*) objects which specify the elements present within a given order domain. We follow Kathol (1995) and assume the architecture of *dom_ obj* objects as in (2.49).[23]

(2.49) $\begin{bmatrix} \textit{dom\_ obj} \\ \text{PHON } \textit{list(phon\_ string)} \\ \text{SYNSEM } \textit{synsem} \end{bmatrix}$

[22] In §2.5.4, the current approach is extended to capture the order of clitics in verbal complexes.

[23] This setup crucially differs from Penn (1999) where *dom_ obj*s do not contain syntactico-semantic information.

In this setup, the order of elements in the order domain associated with a given *sign* (the order of *dom_obj*s on the DOM list) is parallel to the phonological sequence of these elements in the sign. Since the attachment of Polish pronominal clitics does not involve phonological modifications of the host or of the clitic (see §2.2.2), we assume that the PHON value of a *sign* is simply the concatenation of PHON values of its domain objects, as informally given in (2.50), cf. Reape (1992).[24]

(2.50) $$sign \rightarrow \begin{bmatrix} \text{PHON } \boxed{1} \oplus \ldots \oplus \boxed{n} \\ \text{DOM } \left\langle \begin{bmatrix} dom\_obj \\ \text{PHON } \boxed{1} \end{bmatrix}, \ldots, \begin{bmatrix} dom\_obj \\ \text{PHON } \boxed{n} \end{bmatrix} \right\rangle \end{bmatrix}$$

Crucially, ordering 'constituents' (i.e., *dom_obj*s) do not have to correspond to syntactic constituents. Such mismatches can be the result of two operations: sequence union and (partial or total) compaction.[25] Grammatical principles specify by these relations the relationship between domain objects representing the phrase and its daughters (for sequence union) or domain objects representing the phrase and signs corresponding to daughters (for compaction), e.g., Kathol (1995, p. 148). Sequence union (or shuffle), introduced in Reape (1992), allows one to interleave elements of distinct lists but the relative order between the original elements of the lists must be preserved. Thus, *dom_obj*s which are members of one DOM list can be inserted ('shuffled') between elements of a different DOM list in such a way that the relative order of elements on both DOM lists is preserved. To take a simple example, if A is ⟨a, b, c⟩ and B is ⟨d, e⟩, the list ⟨a, d, b, c, e⟩ is a correct result of sequence union but ⟨a, e, b, c, d⟩, where the order of B's elements is reversed, is not.

Roughly speaking, compaction is used to represent a sign as a single domain object in a higher domain, Kathol (1995). Technically, the SYNSEM values of the sign and the corresponding compacted domain object are identical. The phonology of the compacted domain object is the concatenation of PHON values of the sign's domain objects, cf. (2.51) for an illustration.

(2.51) $$\texttt{p-compaction}(\begin{bmatrix} sign \\ \text{SYNSEM } \boxed{0} \\ \text{DOM } \langle [\text{PHON } \boxed{1}], \ldots, [\text{PHON } \boxed{n}] \rangle \end{bmatrix}, \left\langle \begin{bmatrix} dom\_obj \\ \text{SYNSEM } \boxed{0} \\ \text{PHON } \boxed{1} \oplus \ldots \oplus \boxed{n} \end{bmatrix} \right\rangle, \langle\rangle)$$

Total compaction correlates a sign with a single domain object in a higher domain by 'merging' all domain objects of the sign (the third argument of the `p-compaction` relation is the empty list, as in (2.51)); partial compaction 'merges' only some of domain objects of the sign whereas the other *dom_obj*s remain individual entities in a higher domain (in this case, the third argument of the `p-compaction` relation is a non-empty list). A concrete example will be given in fig. 2.2.

Linear precedence (LP) constraints are encoded as general well-formedness conditions on the DOM list which must be satisfied by all *sign*s. The relation `order-constraints` in (2.52) specifies a set of precedence conditions which must be met by *dom_obj*s of every *sign*, cf. Kathol (1995, p. 130).

(2.52) $$sign \rightarrow [\text{DOM } \boxed{1}] \land \texttt{order-constraints}(\boxed{1})$$

Simplifying a little, a clause of the definition of Kathol's `order-constraints` relation can be written as follows in RSRL:

[24]See Penn (1999) for a different approach which permits to incorporate conditions on linear order imposed by phonology, prosody and discourse.

[25]These operations are presented here only informally; see Kathol (1995, ch. 4.4, ch. 5.4 and ch. 8.3.2) for descriptive details and Richter (2000, pp. 263ff) for a full formalisation.

(2.53) $\texttt{order-constraints}(\boxed{1}) \stackrel{\forall}{\Longleftarrow} \texttt{lp}_1(\boxed{1}) \wedge \ldots \wedge \texttt{lp}_n(\boxed{1})$

The relations $\texttt{lp}_1, \ldots, \texttt{lp}_n$ define language-specific linear precedence constraints which must be satisfied by all *sign*s. The symbol $\stackrel{\forall}{\Longleftarrow}$ defines clauses of relations; the universal quantifier, which is a part of this symbol, specifies that variables which occur in the head of the clause ($\boxed{1}$) are universally bound.

2.4.2 Assumptions about Polish

In order to apply the order domain approach to the distribution of Polish pronominal clitics, we have to make several general assumptions about the components of different order domains in Polish.

First, we assume that NPs and PPs are not fully compacted phrases since they can be interleaved with external elements, e.g., (2.23)–(2.24). The linear order in NPs or PPs, however, is not entirely free. For example, an LP constraint must be imposed to ensure that the preposition precedes its complement. Moreover, we assume that a preposition and the first element of its complement form a single domain object in a higher domain, cf. (2.24c), and a PP is partially compacted. A schematic analysis of the VP *mieszka w dużym domu* '(s)he lives in a big house' in fig. 2.2 illustrates partial compaction of a PP in a VP domain.[26] The symbol '$\bigcirc$' is the usual functional notation of the shuffle relation. The DOM values of *word*s are specified in the lexicon whereas those of *phrase*s are the result of principles that impose the desired compaction or shuffle relations. The correct order of domain objects in a PP ($\boxed{14}$) is obtained via the LP constraint mentioned above.

Second, various permutations of dependents of the verb are possible in the S domain. For example, all 24 permutations of elements of a VP in (2.54a) and 6 possible permutations with various linear positions of the subject in (2.54b) are acceptable:

(2.54) a. często Marysi daje kwiaty
often Mary_{dat} gives-he flowers_{acc}
'(he) often gives flowers to Mary'

b. Lubi Jan Marysię.
likes John_{nom} Mary_{acc}
'John likes Mary.'

We assume that the linear structure of a simple sentence is flat: a VP is not a compacted phrase and the subject is sequence unioned with elements of a VP domain. In particular, we represent pronominal clitics as individual domain objects. On the other hand, a subordinate clause is represented as a compacted domain object. As noted in §2.2.5, a subordinate clause boundary constitutes a barrier for the placement of clitics, e.g., (2.26), but also for other elements, see (2.55).[27]

[26] The symbols **H**, **C** and **A** indicate, respectively, the head, complement and adjunct daughters of a given mother node. We assume the standard head-adjunct schema of (Pollard and Sag, 1994, p. 56) which combines the adjunct daughter with the modified head phrase.

[27] *Wh*-extraction out of subordinate clauses of certain verbs is possible in Polish.

(i) Komu Piotr chciał, żeby Maria pomogła?
who Peter wanted COMP Mary helped
'Who did Peter want Mary to help?'

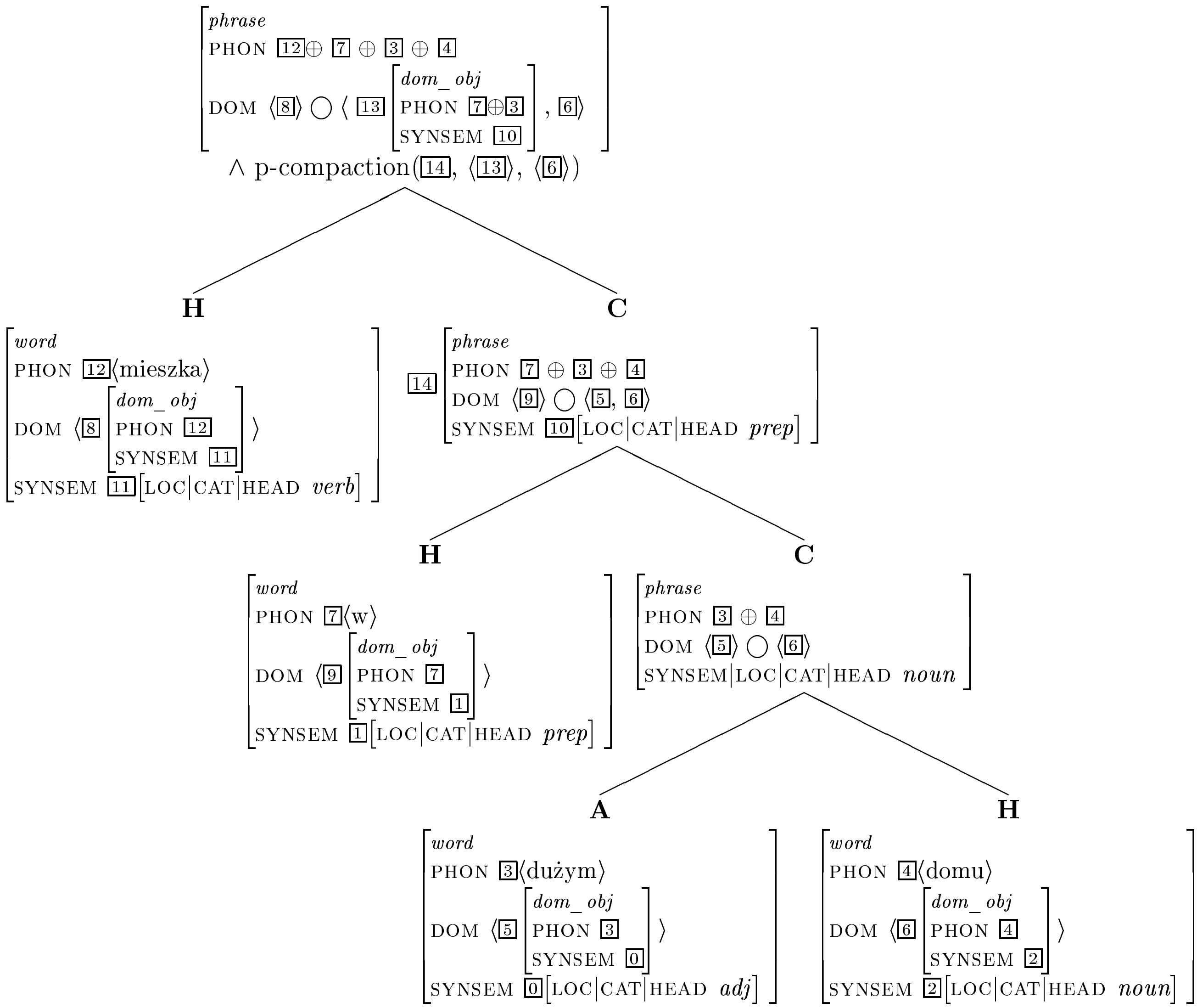


Figure 2.2: A schematic analysis of *mieszka w dużym domu*

(2.26) a. * Piotr powiedział go, że Maria widziała wczoraj w parku.
Peter said him$_{cl}$ COMP Mary saw yesterday in park
'Peter said that Mary saw him yesterday in the park.' (putatively)

b. * Piotr poprosił mu / mi, żeby Maria pomogła.
Peter asked him$_{cl}$ me$_{cl}$ COMP Mary helped
'Peter asked Mary to help him / me.' (putatively)

(2.55) * Piotr poprosił Janowi, żeby Maria pomogła.
Peter asked John COMP Mary helped
'Peter asked Mary to help Peter.' (putatively)

Third, as we have seen in §2.2.5, in phrases headed by gerunds a pronominal clitic can only immediately follow the head, cf. (2.29), but a similar restriction does not carry over to participle phrases, cf. (2.30) repeated below.

(2.29) a. Jan zażądał przyniesienia mu / mi kawy.
John demanded bringing$_{gerund}$ him$_{cl}$ me$_{cl}$ coffee
'John demanded to bring him / me coffee.'

b. * Jan zażądał mu / mi przyniesienia kawy.
John demanded him$_{cl}$ me$_{cl}$ bringing$_{gerund}$ coffee
'John demanded to bring him / me coffee.' (putatively)

(2.30) a. Na głowie położywszy mu / mi dłonie, spojrzał w górę.
on head having-put him$_{cl}$ me$_{cl}$ hands looked-he up
'Having put hands on his / my head, he looked up.'

b. ? Na głowie mu / mi położywszy dłonie, spojrzał w górę.
on head him$_{cl}$ me$_{cl}$ having-put hands looked-he up
'Having put hands on his / my head, he looked up.'

The distribution of clitics in (2.29) and (2.30) conforms to the general linear order in these phrases: a gerund cannot be preceded by a plain NP complement, (2.56a), but a participle can, (2.56b).

(2.56) a. * Jan zażądał Piotrowi przyniesienia kawy.
John demanded Peter bringing$_{gerund}$ coffee
'John demanded to bring Peter coffee.' (putatively)

b. ? Na głowie Janowi położywszy dłonie, spojrzał w górę.
on head John having-put hands looked-he up
'Having put hands on John's head, he looked up.'

In order to capture the distribution of pronominal clitics in gerund phrases, we assume that a gerund always precedes its complements. The fact that clitics, in contrast to non-clitic arguments, must immediately follow the head, see (2.31) on p. 41, is a result of the general restriction on postverbal clitic positions. The same linear constraint which captures the distribution of pronominal clitics in a simple sentence regulates the order of clitics within a participle phrase. As participle phrases cannot be interleaved with other phrases, (2.57b), we assume that participle phrases are compacted domain objects in a higher domain. It is not clear to us at the moment whether the same should be said about gerund phrases as (2.57a) seems to sound better. As the detailed investigation of linear order in Polish is far beyond the scope of this study, we leave the issue here.

(2.57) a. ?? Jan przyniesienia mu / Piortowi zażądał kawy.
John bringing$_{gerund}$ him$_{cl}$ Peter demanded coffee
'John demanded to bring him / Peter coffee.' (intended)

b. * Na głowie spojrzał położywszy mu / Janowi dłonie w górę.
on head looked-he having-put him$_{cl}$ John hands up
'Having put hands on John's / his head, he looked up.' (intended)

2.4.3 Linear Order of Pronominal Clitics

Simple Sentences On the basis of the facts presented in §2.2.5, a first approximation of possible linear positions of Polish pronominal clitics in a simple sentence can be formulated as follows: a clitic either occurs preverbally or, if postverbal, can be preceded only by other clitics.[28] (See Witkoś (1998, p. 159) for a similar conclusion.)

The following constraint on the DOM value of a *sign* uniformly accounts for preverbal and postverbal positions of clitics in a simple clause:

(2.58) a. If the DOM list $\boxed{4}$ contains *dom_ obj*s $\boxed{0}$, $\boxed{1}$ such that $\boxed{0}$ represents a verb and $\boxed{1}$ represents an argument clitic of this verb and $\boxed{0}$ precedes $\boxed{1}$ then
either $\boxed{0}$ immediately precedes $\boxed{1}$ or
there is a *dom_ obj* $\boxed{2}$, which represents an argument clitic of the same verb, such that $\boxed{2}$ immediately precedes $\boxed{1}$.

b.
$$\mathtt{lp}_1(\boxed{4}) \overset{\forall}{\Longleftarrow}$$
$$\forall \boxed{0}\ \forall \boxed{1}\ \forall \boxed{5}$$
$$\left(\begin{array}{l} \exists\boxed{3} \left(\begin{array}{l} \mathtt{member}(\boxed{0} \begin{bmatrix} \textit{dom\_ obj} \\ \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{VAL}|\text{COMPS } \boxed{5} \end{bmatrix} \end{bmatrix}, \boxed{4}) \wedge \mathtt{member}(\boxed{1} \begin{bmatrix} \textit{dom\_ obj} \\ \text{SS } \boxed{3}\textit{clitic} \end{bmatrix}, \boxed{4}) \wedge \\ \mathtt{member}(\boxed{3}, \boxed{5}) \wedge \mathtt{prec}(\boxed{0}, \boxed{1}, \boxed{4}) \end{array} \right) \\ \rightarrow \left(\begin{array}{l} \mathtt{imprec}(\boxed{0}, \boxed{1}, \boxed{4}) \vee \\ \exists\boxed{2}\ \exists\boxed{6}\ \mathtt{imprec}(\boxed{2} \begin{bmatrix} \textit{dom\_ obj} \\ \text{SS } \boxed{6}\textit{clitic} \end{bmatrix}, \boxed{1}, \boxed{4}) \wedge \mathtt{member}(\boxed{6}, \boxed{5}) \end{array} \right) \end{array} \right)$$

This condition should be understood as one of the conjuncts in the clause of the `order-constraints` relation in (2.53). According to (2.52), the DOM value of every *sign* must satisfy the condition in (2.58). The precedence on the DOM list is encoded using two relations: `prec` (precedence) and `imprec` (immediate precedence). The `prec` relation specifies the relative order between elements denoted by the first two arguments of the relation (the first argument precedes the second) on a given list (specified as the third argument of the relation) whereas `imprec` requires the element specified by the first argument of the relation to immediately precede the element specified by the second argument of the relation on a given list. For instance, `imprec`($\boxed{0}$, $\boxed{1}$, $\boxed{4}$) means that $\boxed{0}$ immediately precedes $\boxed{1}$ on the list $\boxed{4}$. The `member` relation is the usual list membership relation.

[28] We assume that sentence initial positions of clitics are excluded by prosody rather than syntax; examples such as (2.9) on p. 34 are left unaccounted for by the syntactically-driven linearization approach adopted here.

$$\begin{bmatrix} \textit{phrase} \\ \text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \langle\rangle \\ \text{COMPS } \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{DOM } \left\langle \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle\text{dziewczyna}\rangle \\ \text{SS } \textit{non-clitic} \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle\text{mu}\rangle \\ \text{SS } \boxed{1}\textit{clitic} \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle\text{wszystko}\rangle \\ \text{SS } \boxed{2}\textit{non-clitic} \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle\text{wytłumaczyła}\rangle \\ \text{SS} \begin{bmatrix} \textit{non-clitic} \\ \text{LC} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{VAL}|\text{COMPS } \langle\boxed{1},\boxed{2}\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 2.3: A domain representation of part of (2.22a)

The LP constraint in (2.58) ensures that the postverbal clitic can be separated from the verb only by another clitic. Therefore, (2.28a) is licensed but (2.28b–c) is not.

(2.28) a. Jan wytłumaczył mu go cierpliwie.
John explained $\text{him}_{dat,cl}$ $\text{him}_{acc,cl}$ patiently
'John patiently explained it to him.'

b. * Jan wytłumaczył jemu go cierpliwie.
John explained him_{dat} $\text{him}_{acc,cl}$ patiently
'John patiently explained it to him.'

c. * Jan wytłumaczył [nowemu koledze]$_{NP}$ go cierpliwie.
John explained new colleague $\text{him}_{acc,cl}$ carefully
'John patiently explained it to a new colleague.'

Note that the constraint in (2.58) does not restrict the DOM list to contain two elements. The DOM list can be longer or shorter but no requirements on the order of elements other than verbs and their clitic arguments are specified here. Since (2.58) is an implicational constraint, it remains valid if the antecedent is not satisfied (the logical implication 'p $\rightarrow$ q' is falsified only if the false consequent 'q' has a true antecedent 'p'). In particular, this constraint does not force clitics to be postverbal or to cluster. Thus, (2.34) is licensed as well.

(2.34) Kiedy go wreszcie mu odebraliśmy...
when $\text{him}_{acc,cl}$ finally $\text{him}_{dat,cl}$ took-away-we
'When we have finally took it away from him...'

Also if there are no clitic arguments, the implication remains valid.

Adopting the assumptions given in §2.4.2, (2.58) licenses the linear order in (2.22a), as indicated in fig. 2.3.

(2.22a) Dziewczyna mu / mi cierpliwie wszystko wytłumaczyła.
girl him_{cl} me_{cl} patiently everything explained
'The girl patiently explained him / me everything.'

An Additional Restriction As the contrast in (2.33) indicates, either all pronominal clitics should precede the verb or none, see §2.2.5.

(2.33) a. Brat mi się podoba.
brother me$_{cl}$ RM$_{cl}$ likes
'I like the brother.'

b. * Brat mi podoba się.

In order to exclude (2.33b),[29] we provide an additional constraint to that presented in (2.58) (the condition in (2.58) prevents postverbal clitics to be separated from the verb by non-clitics but does not require all clitics to occur either pre- or postverbally).

(2.59) a. If the DOM list $\boxed{4}$ contains *dom_obj*s $\boxed{0}$, $\boxed{1}$, such that $\boxed{0}$ represents a verb and $\boxed{1}$ represents an argument clitic of this verb and $\boxed{0}$ precedes $\boxed{1}$ then
if $\boxed{2}$ precedes $\boxed{0}$ then
$\boxed{2}$ must be a *dom_obj* representing a non-clitic.

b. $$\texttt{lp}_2(\boxed{4}) \overset{\forall}{\Longleftarrow} \forall \boxed{0}$$
$$\left(\begin{array}{l} \exists\boxed{3}\ \exists\boxed{5} \\ \left(\begin{array}{l} \texttt{member}\left(\boxed{0}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS|LOC|CAT} \begin{bmatrix}\text{HEAD } \textit{verb}\\ \text{VAL|COMPS } \boxed{5}\end{bmatrix}\end{bmatrix}, \boxed{4}\right) \wedge \texttt{member}\left(\boxed{1}\begin{bmatrix}\textit{dom\_obj}\\ \text{SS } \boxed{3}\textit{clitic}\end{bmatrix},\boxed{4}\right) \\ \wedge\texttt{member}(\boxed{3},\boxed{5}) \wedge \texttt{prec}(\boxed{0},\boxed{1},\boxed{4}) \end{array}\right) \\ \rightarrow \forall\boxed{2}\ \forall\boxed{6} \left(\texttt{prec}\left(\boxed{2}\begin{bmatrix}\textit{dom\_obj}\\ \text{SYNSEM } \boxed{6}\end{bmatrix},\boxed{0},\boxed{4}\right) \rightarrow \boxed{6}\textit{non-clitic}\right) \end{array}\right)$$

According to (2.59), if a verb is followed by a clitic, any element preceding the verb must be a non-clitic. This constraint ensures that if one of the arguments of the verb is a postverbal clitic, $\boxed{1}$, any other clitic must appear postverbally as well, which excludes (2.33b). Note that (2.59) by itself does not ensure that a postverbal clitic can be separated from the verb only by another clitic: according to (2.59), an argument clitic only has to succeed the verb (`prec` rather than `imprec` relation is used). The ungrammatical examples in (2.28b–c) are correctly excluded by the application of both (2.58) and (2.59). (Recall that, due to the definition in (2.53) and the constraint in (2.52), all LP constraints must be satisfied simultaneously by all *signs*).

Note that (2.59) does not require $\boxed{2}$ to represent a complement of the verb. Hence, (2.59) rules out also (2.60) where the preverbal clitic is accompanied by the adverb *codziennie* 'everyday' and the subject *Marysia*.

(2.60) * Codziennie mu Marysia pożycza go po pracy.
everyday to-him$_{cl}$ Mary lends him$_{cl}$ after work
'Mary lends it to him everyday after office hours.'

Since (2.59) is an implicational LP constraint, it does not require all *non-clitic*s to occur preverbally. (2.59) also remains valid if there are no preverbal *non-clitic*s, e.g., (2.16) on p. 37. Since the relative order of Polish pronominal clitics is not fixed, cf. (2.32) on p. 41, we do not restrict their relative order.

[29] See fn. 11 on p. 41.

$$
\begin{bmatrix}
\textit{phrase} \\
\text{SYNSEM}|\text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD}|\text{VFORM} & \textit{gerund} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \end{bmatrix} \\
\text{DOM} & \left\langle \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON} & \langle \text{przyniesienie} \rangle \\ \text{SS} & \begin{bmatrix} \textit{non-clitic} \\ \text{LC} & \begin{bmatrix} \text{HEAD}|\text{VFORM} & \textit{gerund} \\ \text{VAL}|\text{COMPS} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON} & \langle \text{mu} \rangle \\ \text{SS} & \boxed{1}\textit{clitic} \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON} & \langle \text{kawy} \rangle \\ \text{SS} & \boxed{2}\textit{non-clitic} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

Figure 2.4: A domain representation of part of (2.29a)

Gerund and Participle Phrases We assume that the HEAD values of gerunds and participles are also objects of the sort *verb*, see Przepiórkowski (1999, p. 420). Therefore, the constraints in (2.58) and (2.59) regulate the order of pronominal clitics in these phrases as well. We do not impose any specific restrictions on the order within a participle phrase (see (2.30) and (2.56b)) but we assume that a gerund must precede its complements, cf. (2.61).

(2.61) a. If the DOM list $\boxed{4}$ contains *dom_obj*s $\boxed{0}$ and $\boxed{3}$ such that $\boxed{0}$ represents a gerund and $\boxed{3}$ represents an complement of the gerund,

then $\boxed{0}$ must precede $\boxed{3}$.

b. $\texttt{lp}_3(\boxed{4}) \overset{\forall}{\Longleftarrow} \forall \boxed{0}\ \forall \boxed{3}$

$$
\left(\begin{array}{l}
\exists \boxed{1}\ \exists \boxed{2} \\
\left(\begin{array}{l} \texttt{member}\left(\boxed{0}\begin{bmatrix} \textit{dom\_obj} \\ \text{SLC} & \begin{bmatrix} \text{HEAD}|\text{VFORM} & \textit{gerund} \\ \text{VAL}|\text{COMPS} & \boxed{2} \end{bmatrix} \end{bmatrix}, \boxed{4}\right) \wedge \texttt{member}\left(\boxed{3}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS} & \boxed{1} \end{bmatrix}, \boxed{4}\right) \\ \wedge\ \texttt{member}(\boxed{1}, \boxed{2}) \end{array} \right) \\
\rightarrow\ \texttt{prec}(\boxed{0}, \boxed{3}, \boxed{4})
\end{array} \right)
$$

Since pronominal clitics are complements of a gerund, i.e., they are present on the COMPS list (cf. §2.3.4), (2.61) will force clitics to follow the gerund. Simultaneously, the constraint in (2.58) ensures that a clitic can be separated only by another clitic from the preceding gerund.

A schematic domain representation of the gerund phrase *przyniesienie mu kawy* 'bringing him coffee' in (2.29a) is given in fig. 2.4.

2.5 Clitic Climbing

The so-called clitic climbing (CC) is a cross-linguistic phenomenon present in many Romance but also Slavic languages. In certain verbal environments, a pronominal clitic can be realised as an argument of a governing verb rather than the verb the clitic semantically belongs to:

(2.62) Piotr go chce kupić w Austrii.
Peter him_{cl} wants buy_{inf} in Austria
'Peter wants to buy it in Austria.'

In (2.62), the clitic *go* 'it' originates as an argument of the infinitive *kupić* 'to buy' but it 'climbs' to the main verb *chce* 'wants' instead of being realised locally.

CC in Polish is a rather neglected phenomenon. Some attention has been devoted to CC in Dyła (1983), Dziwirek (1994), Rappaport (1998), but the phenomenon has been more extensively dealt with only recently in Witkoś (1996b, 1998), Kupść (1999a,b,c,d, 2000a). The main contribution of this section is the discussion of CC in all environments where the phenomenon occurs in Polish. More precisely, we consider CC in the complementizerless verbal complexes (verb clusters, VCs) such as in (2.62), and in tense auxiliary constructions. We discuss the behaviour of pronominal clitics in these environments and provide an HPSG account of the presented facts.[30]

2.5.1 Environments for CC in Polish

Verb Clusters Cross-linguistically, CC is possible out of certain infinitive complements. In Italian, CC occurs with 'restructuring verbs', cf. e.g., Rizzi (1982), while in French, causative verbs such as *faire* 'make', may attract clitic arguments of their infinitive complements, cf. e.g., Kayne (1975). In Polish, CC is possible out of bare infinitival complements which form a verbal complex with the main verb, so-called verb clusters (VCs). Pretheoretically, by verb clusters we mean chains of (non-auxiliary) verbs, not necessarily linearly contiguous, such that each (apart from the main verb) is subcategorized for by another,[31] and none (again, perhaps apart from the main verb) is introduced by a complementizer, e.g., verbs *chce* 'wants' and *kupić* 'to buy' form a VC in (2.62) but not in (2.63).

As (2.63) shows, CC does not occur just in any infinitive environment: the presence of the complementizer *żeby*, which introduces the infinitival subordinate clause in (2.63), blocks CC, see Witkoś (1998, p. 192) for similar data.[32]

(2.63) * Piotr chce go, żeby zaprosić na kolację.
Peter wants him_{cl} COMP invite_{inf} on dinner
'Peter wants to invite him for dinner.' (putatively)

Although the data are not uncontroversial, Dyła (1983) observes that not all VCs behave uniformly with respect to CC. If the subject control verb, e.g., *chcieć* 'want' or *przestać* 'stop' has no nominal complement itself, CC is possible, see (2.62). Certain subject control verbs may optionally occur with an NP complement, e.g., *obiecać*+NP[dat] or *przyrzec*+NP[dat] 'promise'. In this case, CC is allowed only if this nominal complement is not overtly realised, see the contrast in (2.64).

(2.64) a. Marek obiecał go zaprosić.
Mark promised him_{cl} invite_{inf}
'Mark promised to invite him.'

b. * Marek obiecał go Joli zaprosić.
Mark promised him_{cl} Jola_{dat} invite_{inf}
'Mark promised Jola to invite him.'

[30] The behaviour of the reflexive clitic *się* is discussed in Chapter 3.

[31] Hence, we do not consider (i) a VC since the infinitival modifier *kupić chleb* 'to buy bread' is not a complement subcategorized by the main verb.

(i) Jan poszedł do sklepu kupić chleb.
John went to shop buy_{inf} bread
'John went to the shop to buy some bread.'

[32] In (2.63)–(2.66), *go* 'him' is a semantic argument of the 'lowest' infinitive.

Some object control verbs, such as *uczyć* +NP[acc] 'teach' or *przeszkadzać* +NP[dat] 'prevent' do not allow CC, either, (2.65a). It might seem that it is the presence of a nominal complement that blocks CC in (2.64b) and (2.65a). Note, however, that 'causative' or 'prohibitive' (object control) verbs, such as *kazać* +NP[dat] 'order' or *zabronić* +NP[dat] 'forbid' in (2.65b), seem to allow CC in spite of having a nominal complement; see also Witkoś (1998, p. 191) for CC with object control verbs.

(2.65) a. * Janek przeszkodził go Piotrowi zaprosić.
John disturbed him_{cl} Peter_{dat} invite_{inf}
'John prevented Peter from inviting him.'

b. ? Janek kazał / zabronił go Piotrowi zaprosić.
John ordered forbade him_{cl} Peter_{dat} invite_{inf}
'John ordered/forbade Peter to invite him.'

Also the behaviour of subject raising verbs[33] does not conform to the view that the presence of a nominal complement prohibits CC out of a bare infinitival complement. The subject raising verb *wydaje się* 'seems' in (2.66) obligatorily occurs with the reflexive clitic *się*,[34] which nevertheless does not stop CC; similar data are given in Witkoś (1998, p. 191).

(2.66) Maria się go wydaje lubić.
Mary RM him_{cl} seems like_{inf}
'Mary seems to like him.'

As the contrast between the acceptable examples in (2.65b)–(2.66) and the ungrammatical in (2.64b)–(2.65a) indicates, the presence of a nominal complement is not an obstacle for CC in general. Moreover, verbs which allow CC do not form a uniform semantic class. For example, both *przeszkodzić* 'prevent', (2.65a), and *zabronić* 'forbid', (2.65b), have a dative nominal complement and are 'prohibitive' object control verbs but only *przeszkodzić* disallows CC.

Since no uniform syntactic or semantic characteristics of all verbs which (dis)allow CC is available, we conclude that verbs which accept CC are specified as such in the lexicon.

Tense Auxiliaries As in Romance, CC in Polish is possible in tense auxiliary constructions, see (2.67).

(2.67) Będę go jutro oglądał.
Aux_{fut}-I him_{cl} tomorrow watched
'I'll be watching it/him tomorrow.'

In (2.67), the clitic *go* 'it/him' originates as an argument of the verb *oglądał* but is realised on the auxiliary *będę* rather than locally.

Although CC to tense auxiliaries is well-known and has been a topic of discussion in the Romance literature since Kayne (1975) (for French) and Rizzi (1982) (for Italian), it has been noticed for Polish only recently (independently) by Witkoś (1996b, 1998) and Kupść (1999b,a). A detailed presentation of CC in tense auxiliary constructions is postponed to §2.5.5.

[33] According to Dziwirek (1994), Polish has no object raising verbs taking infinitival complements.
[34] See Chapter 3 for the grammatical status of *się*.

Other Environments Verbs which subcategorize for bare infinitive complements and tense auxiliaries exhaust the list of verbs which allow for CC in Polish. We present below some examples of potential environments of CC to show that CC is impossible in gerund or participle phrases and out of subordinate clauses in Polish.

In (2.63), we have shown that the presence of a complementizer blocks CC out of an infinitive complement. As (2.68) indicate, CC is not possible across a subordinate clause boundary in general.

(2.68) a. * Jan wie go / cię, czy Maria zaprosi.
John knows him_{cl} you_{cl} Q Mary will.invite
'John knows if Mary will invite him / you.'

b. * Jan chce go / cię, żeby Maria zaprosiła.
John wants him_{cl} you_{cl} COMP Mary invited
'John wanted that Mary invited him / you.'

In Polish, gerunds (verbal nominals) and participles may have clitic arguments as well, see §2.2. CC is impossible out of a gerund, (2.69a), or a participle phrase, (2.69b) (the former fact has been observed in Rappaport (1998, p. 467), Przepiórkowski (1999, p. 160)).[35]

(2.69) a. * Lekarz zalecił go / cię niezwłoczne odwiezienie do szpitala.
doctor ordered him_{cl} you_{cl} immediate sending_{gerund} to hospital
'The doctor ordered sending him / you immediately to the hospital.'

b. * Jan zemdleje go / cię ponownie widząc.
John will.faint him_{cl} you_{cl} again $\text{seeing}_{participle}$
'John will faint if he sees him / you again.'

We conclude that environments where CC is possible in Polish are lexically specified. The subsequent sections are dedicated to a discussion of CC in Polish.

2.5.2 CC in Verb Clusters

More Data In the previous subsection, we dealt with the data which showed restrictions on VCs which accept CC. At present, we turn to the behaviour of clitics themselves.

As (2.70) illustrates, CC in Polish VCs is optional. A clitic can be either realised locally, (2.70a), or climb to a governing verb, (2.70b–c), cf. Witkoś (1998, p. 192).

(2.70) a. Jan chciał obudzić go o szóstej.
John wanted wake_{inf} him_{cl} at six
'John wanted to wake him up at six.'

b. Jan chciał go obudzić o szóstej.

c. Jan go chciał obudzić o szóstej.

[35] Note that the ungrammaticality of (2.69) cannot be uniformly explained by a linear constraint which disallows pronominal clitics to precede a gerund or a participle. As discussed in §2.2.5, pronominal clitics can precede a participle but not a gerund; the equivalent of (2.69b) where a clitic appears within the participle phrase is grammatical, see (i).

(i) Jan zemdleje ponownie go / cię widząc.
John will.faint again him_{cl} you_{cl} $\text{seeing}_{participle}$
'John will faint if he sees him / you again.'

As (2.70b) shows, the clitic does not have to climb to the highest verb but can be realised on any intermediate verb as well.

Moreover, unlike in Romance, if an infinitival complement has several clitic arguments, they need not appear on the same verb, (2.71a–b) (*mu* 'him' and *go* 'him' are arguments of the verb *pokazać* 'to show'); for a comparison of CC in Polish and Romance see Kupść (1999c). As (2.71c) shows, this behaviour is optional and all clitics can be realised on the same verb as well, see Witkoś (1996b, 1998) for the same observation.

(2.71) a.

Jan	może	mu	chcieć	go	pokazać.
John	may	him$_{cl}$	want$_{inf}$	him$_{cl}$	show$_{inf}$

'John may want to show it to him.'

b. Jan może go chcieć mu pokazać.

c. Jan może mu go chcieć pokazać.

The examples in (2.71a–b) show also that clitics do not have to climb in any particular order.

Syntactic CC Since Polish is a relatively free word order language, CC might be seen as a result of linear order without syntactic climbing taking place. The placement of pronominal clitics in Polish is rather free. The distribution of clitics in the examples considered above conforms to the general picture presented in §2.2.5: clitics either precede the verb they originate from, e.g., (2.70b–c), (2.71), or immediately follow this verb, (2.70a), see Witkoś (1996b, 1998) and Kupść (1999e).

There is, however, a good reason not to consider CC a linear phenomenon. Data such as (2.72) provide evidence for the claim that CC takes place in syntax, see Rappaport (1998), Kupść (1999d), Przepiórkowski (1999, p. 162), Przepiórkowski (2000a).

(2.72) a.

Mogłem	zgubić	go	wczoraj.
could-I	loose$_{inf}$	him$_{cl}$	yesterday

'I could have lost it yesterday.'

b.

Zgubić	mogłem	go	wczoraj.
loose$_{inf}$	could-I	him$_{cl}$	yesterday

'I could have lost it yesterday.'

c. *

Mogłem	zgubić	w parku	go	wczoraj.
could-I	loose$_{inf}$	in park	him$_{cl}$	yesterday

'I could have lost it yesterday in the park.'

If the clitic *go* in (2.72b) (as in (2.72a)) is realised locally as an argument of its semantic governor *zgubić* 'to loose', a clitic follows this verb and they are separated by a non-clitic element, which is impossible in general, §2.2.5. As the ungrammaticality of (2.72c) shows, such an exceptional behaviour is not a property of VCs. The puzzling contrast between (2.72b) and (2.72c) disappears if the clitic in (2.72b) is raised to and realised on the main verb *mogłem* '(I) could' rather than on the infinitive, as in (2.72a). If this perspective is adopted, the clitic in (2.72b) immediately follows the verb it is realised from, as in (2.72a).

In order to avoid spurious ambiguity, we will assume that CC in VCs is always a syntactic rather than a linear phenomenon: as in (2.72b), the clitic *go* 'him' in (2.70b–c) is raised to and realised as an argument of the main verb rather than the argument of its infinitival complement.

CC and Negation On the basis of the data in (2.73), Witkoś (1998, p. 193) claims that in Polish, CC out of a negated bare infinitive complement is possible.

(2.73) Kapitan go mógł nie bić.
captain him_{cl} could NM beat_{inf}
'The captain might not beat him.'

Przepiórkowski (1999, p. 162ff) argues, however, that examples such as (2.73) are misleading and negation stops CC. Let us consider (2.74) (simialar data are given by Przepiórkowski (1999, p. 162)).

(2.74) a. Mogłem nie bić go wczoraj.
could-I NM beat_{inf} him_{cl} yesterday
'I could not beat him yesterday.'

b. ?* Nie bić mogłem go wczoraj.
NM beat_{inf} could-I him_{cl} yesterday
'I could not beat him yesterday.'

c. ? Nie bić go mogłem wczoraj.
NM beat_{inf} him_{cl} could-I yesterday
'I could not beat him yesterday.'

(2.74b) shows that if a verb is negated, it cannot be easily fronted without a clitic, in contrast to (2.72b), where a non-negated verb can be fronted without a clitic. The example in (2.74c) indicates that if the clitic is fronted with the negated verb, the grammaticality of the sentence is clearly improved. We conclude, after Przepiórkowski (1999), that negation stops syntactic CC.

Once we exclude CC across negation, (2.73) has to be accounted for on different grounds. We follow Przepiórkowski (1999, p. 161) and assume that *go* is realised locally as the argument of the negated verb and they form a VP. Since syntactic phrases in Polish do not have to be linearly contiguous (cf. §2.2.5), we assume that the clitic and the negated verb form such a discontinuous constituent in (2.73). Note that our treatment of CC in non-negated VCs, e.g., (2.70), is different in this respect: clitic arguments of a non-negated verb are raised in syntax rather than locally realised.

Properties of Polish VCs Before proceeding to the analysis of CC in VCs, let us concentrate on properties of the verbal complex itself.

From what we have seen so far, it is clear that CC in VCs is a syntactic phenomenon, i.e., a clitic argument of a verb is raised in syntax to a governing verb. Moreover, this process is stopped by intervening negation: clitic arguments of a negated verb in VCs cannot be further raised. Nevertheless, several questions remain: do other arguments of a non-negated verb raise with the clitic? does negation force local realisation of other complements as well? what is the syntactic structure of VCs in general? etc. Providing full answers to these questions requires a much careful investigation and we address them only briefly here.

Dziwirek (1994, 1998) observes that a number of phenomena (extraction/scrambling, CC, genitive of negation (GoN), Negative Concord (NC), binding) show 'clause union' effects and suggest a monoclausal analysis of VCs. Simultaneously, the optionality of these phenomena seems to argue in favour of bi-clausal characteristics of VCs. As discussed in Przepiórkowski

(1999, pp. 158ff), the locality constraints on NC, binding and extraction/scrambling in Polish are more relaxed than those on CC and GoN.[36] Therefore, a single mechanism cannot be responsible for locality constraints on all these phenomena. For this reason, we concentrate only on CC and GoN and disregard here NC, binding and extraction/scrambling data.

Przepiórkowski (1999, pp. 145ff), Przepiórkowski (2000a) observes the optionality of the so-called genitive of negation (GoN): in Polish (as well as in many other languages), an accusative complement of a non-negated verb, (2.75a), occurs in genitive if the verb is negated, (2.75b).

(2.75) a. Janek lubi Marysię / *Marysi.
John likes Mary_{acc} Mary_{gen}
'John likes Mary.'

b. Janek nie lubi Marysi / *Marysię.
John NM likes Mary_{gen} Mary_{acc}
'John doesn't like Mary.'

Although GoN is a clausebound phenomenon in Polish, the same case alteration as in (2.75) may occur in VCs, see (2.76).

(2.76) a. Jan lubi jeść ser.
John likes eat_{inf} cheese_{acc}
'John likes to eat cheese.'

b. Jan nie lubi jeść sera.
John NM likes eat_{inf} cheese_{gen}
'John does not like to eat cheese.'

GoN is considered obligatory in most of the literature, see Dziwirek (1994, 1998), Witkoś (1996a, 1998), Przepiórkowski and Kupść (1997a,c), Przepiórkowski and Świdziński (1997), but examples such as (2.77), cited in Przepiórkowski (1999, p. 146), show optionality of GoN in VCs.

(2.77) a. Czy nie można by sklepy / sklepów zaopatrzyć w artykuły chemiczne?
Q NM may AUX.COND shops_{acc} shops_{gen} supply_{inf} with articles chemical
'Couldn't one supply shops with chemical articles?'

b. Nie sposób sprawdzić im bilety / biletów.
NM possible check_{inf} them tickets_{acc} tickets_{gen}
'It is impossible to check their tickets.'

In order to account for these facts, Przepiórkowski (1999) assumes that 'clause union' occurs in VCs only optionally. If an accusative argument of the non-negated infinitive is raised to the negated verb, the NP is assigned genitive case, as in (2.76b). On the other hand, if no raising takes place, the NP is marked as accusative by the non-negated verb, similarly to (2.76a).

Moreover, in order to exclude the ungrammatical accusative case marking in (2.78), Przepiórkowski (1999, p. 155) assumes that negation is an obstacle for further argument raising: arguments can be raised only to a negated verb in a VC (i.e., the main verb in (2.77) and the infinitive in (2.78)), see also Przepiórkowski and Kupść (1997c).

[36] Przepiórkowski (1999, p. 158) mentions also haplology of the reflexive clitic *się* as yet another argument for 'clause union' analysis of VCs. We postpone the discussion of haplology to Chapter 3.

(2.78) Mogłem nie robić tego / *to dalej.
could-I NM do$_{inf}$ it$_{gen}$ it$_{acc}$ anymore
'I could have refrained from doing that.'

If the argument *tego* 'this' of the negated infinitive cannot raise to the main (non-negated) verb, only genitive is possible in (2.78). Unlike in (2.77), the NP is assigned case by the negated verb and accusative is correctly ruled out.

Taking into account the above data, we conclude that in VCs any nominal argument (and not just clitics) can be optionally raised to the governing non-negated verb, cf. Przepiórkowski (1999, 2000a), Kupść (2000a). Since clitics can be raised independently of each other, (2.71a–b), we assume that raising of one argument does not necessarily force raising of its co-arguments. On the other hand, we assume that all arguments (and not only clitics) of negated verbs cannot raise any further.[37]

2.5.3 An HPSG Analysis of CC in VCs

As noted in §2.5.1, CC in VCs is allowed only with certain verbs, cf. (2.64)–(2.66). Verbs which take a bare infinitival complement are enumerated in the lexicon and we encode the (im)possibility of CC in their lexical entries.

Argument Composition In HPSG, CC has been often modelled via different variants of the argument composition technique incorporated to HPSG in Hinrichs and Nakazawa (1989),[38], e.g., Abeillé and Godard (1994), Miller and Sag (1997), Abeillé *et al.* (1998) (for French), Monachesi (1993, 1998a, 1999) (for Italian), Monachesi (1998b) (for Romanian). This mechanism allows one to treat unrealised arguments of a lower verb as additional arguments of the higher verb. Raising of arguments is encoded lexically by sharing unrealised arguments of the unsaturated complement with the governing verb, as schematically indicated in (2.79).

(2.79)
$$\begin{bmatrix} \textit{word} \\ \text{SYNSEM|LOC|CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{SUBCAT } \langle \begin{bmatrix} \text{LOC|CAT} \begin{bmatrix} \text{HEAD|VFORM } \textit{inf} \\ \text{SUBCAT } \boxed{1} \end{bmatrix} \end{bmatrix} \rangle \oplus \boxed{1} \end{bmatrix} \end{bmatrix}$$

Note that (2.79) does not specify whether the verb subcategorizes for a phrasal or a lexical infinitive complement; the SUBCAT value of the complement is not further restricted. Such a specification will be used below to account for the optionality of argument raising in Polish.

Argument Raising in Polish In (2.79), argument composition has been encoded in terms of the SUBCAT list. As discussed in §1.3.1, the function of the SUBCAT attribute has been distributed in most recent HPSG between ARG-ST (ARGUMENT-STRUCTURE) present only in *words* and valence attributes (SUBJ and COMPS lists) present in all *signs*. In this study, we assume that the value of ARG-ST is the concatenation of the SUBJ and COMPS lists (in this order), see §1.3.1.

[37]As discussed in Przepiórkowski and Kupść (1997c,a), traditional constituency tests (preposing, coordination, pronominalization) are indecisive for the syntactic structure of Polish VCs with or without negation. The same criteria applied to (non-negated) VCs with or without CC are not conclusive either.

[38]The idea of argument composition stems from categorial grammar, Geach (1970), and has been adopted for GPSG in Johnson (1986) Meurers (see 1999, p. 26).

As argued in §2.5.2 for GoN, argument raising (as well as CC) in Polish VCs is optional, i.e., a (non-negated) verb can subcategorize either for a phrasal or a lexical infinitival complement. Since the combinatoric potential of a *sign* is encoded in the valence attributes, we assume that raised arguments originate from the COMPS list (rather than ARG-ST which appears only in *words*). On the other hand, we want raised arguments to be assigned case or obey binding principles. As these principles have ARG-ST as their locus (cf. §2.3.4), we assume that raised arguments land on ARG-ST, see (2.80) for an illustration; for an implementation of argument composition as raising from COMPS to the ARG-ST list see Abeillé *et al.* (1998) (for French) and Przepiórkowski (1999, 2000a) (for Polish).

(2.80)
$$\begin{bmatrix} \textit{word} \\ \text{SS|LOC|CAT} \begin{bmatrix} \textit{category} \\ \text{HEAD } \textit{verb} \\ \text{ARG-ST } \textit{ne\_list} \oplus \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LOC|CAT} \begin{bmatrix} \text{HEAD|VFORM } \textit{inf} \\ \text{VAL|COMPS } \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \oplus \boxed{1} \end{bmatrix} \end{bmatrix}$$

As in (2.79), the value of $\boxed{1}$ in (2.80) is not further restricted and it can be any list of *synsems*. In particular, the list can be empty if all complements of the infinitive (including clitics, if any) are realised, i.e., if the main verb combines with an infinitival phrase. On the other hand, $\boxed{1}$ may be non-empty, which we exploit for clitic climbing below.

Lexical Entries As illustrated in (2.80), all verbs which subcategorize for a bare infinitive may inherit unrealised arguments of their infinitival complement. As discussed in §2.5.1, only some of these verbs allow for CC. Due to the split of the *synsem* sort, see (2.39) on p. 44, and the lexical encoding of CC, cf. (2.80), this fact is not problematic for our analysis.

In particular, the object control verb *kazać* 'order' in (2.65b) can be described as in (2.81).[39]

(2.81)
$$\begin{bmatrix} \textit{word} \\ \text{PHON } \langle \text{kazać} \rangle \\ \text{SLC} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{ARG-ST } \langle \text{NP}[\textit{str}], \text{NP}_{\boxed{2}}[\textit{ldat}] \rangle \oplus \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LC} \begin{bmatrix} \text{HEAD|VFORM } \textit{inf} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \langle \text{NP}_{\boxed{2}}[\textit{str}] \rangle \\ \text{COMPS } \boxed{1}\textit{list(synsem)} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \oplus \boxed{1} \end{bmatrix} \end{bmatrix}$$

The fact that *kazać* 'to order' is an object control verb is indicated via identity of the INDEX value of the dative object ($\boxed{2}$) with the INDEX value of (unrealised) subject of the infinitival complement, cf. Pollard and Sag (1994, p. 135). According to (2.81), the verb *kazać* inherits all unrealised complements of its infinitive complement, including clitics (if any). The infinitive's COMPS is any list of *synsems* and it may contain both *clitics* or *non-clitics*. Since $\boxed{1}$ remains uninstantiated, the lexical entry in (2.81) licenses CC, as in (2.65b), but also the local realisation of arguments.

The object control verb *przeszkodzić* 'prevent' requires an infinitive complement as well, i.e., it is compatible with (2.80), but CC is not possible, see (2.65a). We account for this fact in the lexical entry of the verb: *przeszkodzić* may inherit only *non-clitic* complements rather

[39]Technically, there is no singleton **list** which contains the infinitival complement in (2.81) if $\boxed{1}$ is non-empty. In a formal RSRL account, one should use a chain instead of a list, cf. Richter (2000, ch. 4.3). For simplicity, we will use the familiar list notation here. The same remark applies to (2.79) and (2.80). We will continue to ignore the technical distinction between lists and chains.

than just any unrealised complement as in (2.81). The relevant part of the lexical entry for *przeszkodzić* is given in (2.82).

(2.82)
$$\left[\begin{array}{l} \textit{word} \\ \textsc{phon}\ \langle \text{przeszkodzić} \rangle \\ \textsc{slc}\ \left[\begin{array}{l} \textsc{head}\ \textit{verb} \\ \textsc{arg-st}\ \langle \text{NP}[\textit{str}], \text{NP}_{\boxed{2}}[\textit{ldat}] \rangle \oplus \langle \left[\begin{array}{l} \textit{synsem} \\ \textsc{lc}\ \left[\begin{array}{l} \textsc{head}|\textsc{vform}\ \textit{inf} \\ \textsc{val}\ \left[\begin{array}{l} \textsc{subj}\ \langle \text{NP}_{\boxed{2}}[\textit{str}] \rangle \\ \textsc{comps}\ \boxed{1}\textit{list(non-clitic)} \end{array}\right] \end{array}\right] \end{array}\right] \rangle \oplus \boxed{1} \end{array}\right] \end{array}\right]$$

Note that (2.82) does not prevent the infinitival complement from having clitic arguments. What (2.82) says is that the list of raised arguments, $\boxed{1}$, does not contain clitics. This means that if the infinitival complement has any clitic arguments, they must be realised locally and the main verb subcategorizes for a phrasal complement (an infinitive with clitic complements).[40]

As observed in §2.5.1, also subject control verbs do not behave uniformly with respect to CC: apparently the overt realisation of a nominal complement (if any) blocks CC, see *obiecał* '(he) promised' in (2.64). Such verbs satisfy the description in (2.80), i.e., in principle raising of any complement is possible. The partial lexical entry for *obiecał* '(he) promised', which optionally appears with a dative NP complement, is given in (2.83).

(2.83)
$$\left[\begin{array}{l} \textit{word} \\ \textsc{phon}\ \langle \text{obiecał} \rangle \\ \textsc{slc}\ \left[\begin{array}{l} \textsc{head}\ \textit{verb} \\ \textsc{arg-st}\ \langle \text{NP}_{\boxed{2}}[\textit{str}] \rangle \oplus \boxed{3} \oplus \langle \left[\begin{array}{l} \textit{synsem} \\ \textsc{lc}\ \left[\begin{array}{l} \textsc{head}|\textsc{vform}\ \textit{inf} \\ \textsc{val}\ \left[\begin{array}{l} \textsc{subj}\ \langle \text{NP}_{\boxed{2}}[\textit{str}] \rangle \\ \textsc{comps}\ \boxed{1}\textit{list(synsem)} \end{array}\right] \end{array}\right] \end{array}\right] \rangle \oplus \boxed{1} \end{array}\right] \end{array}\right]$$
$$\wedge\ \boxed{3}\ (\langle\rangle \vee \langle \text{NP}[\textit{ldat}] \rangle)$$

The fact that *obiecał* '(he) promised' is a subject control verb is encoded via identity of subjects' INDEX values ($\boxed{2}$), cf. Pollard and Sag (1994, p. 135). The optionality of NP[dat] object is indicated via $\boxed{3}$: it is either the empty list or it contains a singleton NP[dat] element.

The ungrammaticality of (2.64b) is predicted by the lexical constraint in (2.84).

(2.84)
$$\forall \boxed{1}\ (\left[\begin{array}{l} \textit{word} \\ \textsc{slc}\ \left[\begin{array}{l} \textsc{head}\ \textit{verb} \\ \textsc{arg-st}\ \langle \text{NP}_{\boxed{2}}[\textit{str}], \text{NP}[\textit{ldat}] \rangle \oplus \langle \left[\begin{array}{l} \textit{synsem} \\ \textsc{lc}\ \left[\begin{array}{l} \textsc{head}|\textsc{vform}\ \textit{inf} \\ \textsc{val}\ \left[\begin{array}{l} \textsc{subj}\ \langle \text{NP}_{\boxed{2}}[\textit{str}] \rangle \\ \textsc{comps}\ \boxed{1}\textit{list(synsem)} \end{array}\right] \end{array}\right] \end{array}\right] \rangle \oplus \boxed{1} \end{array}\right] \end{array}\right]$$
$$\rightarrow \boxed{1}\textit{list(non-clitic)})$$

If a subject control verb has an overt complement NP[*ldat*], the constraint in (2.84) restricts inherited complements ($\boxed{1}$) to be non-clitics.

Finally, the subject raising verb *wydawać się* 'to seem' in (2.66) does not block CC. The relevant part of its lexical entry is given in (2.85).[41]

[40] As we will see shortly, we assume a version of the Valence Principle which allows us to realize arguments in any order, i.e., all clitics can be realized before non-clitics are, independently of their order on the COMPS list.

[41] As we will see in Chapter 3, the reflexive clitic *się* is in fact a subcategorized for syntactic element rather than a part of the verb, as suggested by the PHON value in (2.85).

(2.85) $$\begin{bmatrix} \textit{word} \\ \text{PHON} \ \langle \text{wydawać się} \rangle \\ \text{SLC} \begin{bmatrix} \text{HEAD} \ \textit{verb} \\ \text{ARG-ST} \ \langle \boxed{2}\text{NP}[\textit{str}]\rangle \oplus \langle \begin{bmatrix} \textit{synsem} \\ \text{LC} \begin{bmatrix} \text{HEAD}|\text{VFORM} \ \textit{inf} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \ \langle\boxed{2}\rangle \\ \text{COMPS} \ \boxed{1}\textit{list(synsem)} \end{bmatrix} \end{bmatrix} \end{bmatrix} \rangle \oplus \boxed{1} \end{bmatrix} \end{bmatrix}$$

As in Pollard and Sag (1994, p. 135), the unrealised subject of the infinitival complement is token-identical with the subject of the subject raising verb. Note that subjects share the entire *synsem*, not just an index as in (2.83). Since subject raising verbs allow for CC, $\boxed{1}$ is any list of *synsem*s, similarly to (2.81).

Syntactic Realisation Polish pronominal clitics are realised on a par with other syntactic arguments via general syntactic principles, see §2.3. We have to slightly modify standard HPSG syntactic principles responsible for the realisation of arguments in order to account for the complexity of Polish CC.

First, we assume a version of the Valence Principle which allows us to realise elements of the COMPS list in any order and quantity.[42]

Clitics do not necessarily climb all at a time, (2.71a), and do not have to be realised in any specific order, (2.71a–b).

(2.71) a. Jan może mu chcieć go pokazać.
John may him$_{cl}$ want$_{inf}$ him$_{cl}$ show$_{inf}$
'John may want to show it to him.'

b. Jan może go chcieć mu pokazać.

c. Jan może mu go chcieć pokazać.

Such a flexibility cannot be captured by the standard HPSG Valence Principle; the conjunct of the Valence Principle responsible for the realisation of complements is repeated below.

(1.10b) $$\begin{bmatrix} \textit{phrase} \\ \text{DTRS} \ \textit{head\_struc} \end{bmatrix} \rightarrow$$

$$\left(\begin{bmatrix} \text{SYNSEM}|\text{LOCAL}|\text{CATEGORY}|\text{VALENCE}|\text{COMPS} \ \boxed{1} \\ \text{DTRS} \begin{bmatrix} \text{HEAD-DTR}|\text{SYNSEM}|\text{LOCAL}|\text{CATEGORY}|\text{VALENCE}|\text{COMPS} \ \boxed{2'} \oplus \boxed{1} \\ \text{COMP-DTRS} \ \boxed{2} \end{bmatrix} \end{bmatrix} \right.$$

$$\left. \wedge \ \texttt{phrases\_to\_synsems}(\boxed{2},\boxed{2'}) \right)$$

The constraint in (1.10b) ensures that elements of the COMPS list of the lexical head are realized in any quantity (the list of remaining arguments, $\boxed{1}$, might not be empty) but in a specific order: from left to right in the order in which they appear on the COMPS list of the head daughter: elements of $\boxed{2'}$ are realized before elements of $\boxed{1}$.

Note that in (2.71a), the clitic *go* is realised before the clitic *mu* but the reversed order of realisation in (2.71b) is possible as well (as argued in §2.5.2, CC is a syntactic phenomenon and the positions of clitics cannot be reduced to their relative order). The standard HPSG Valence Principle is not able to account for both examples. The relevant modification of (1.10b) is given in (2.86).

[42]The need for such a reformulation of the Valence Principle for Polish was previously noted in Kupść (1999f) (on the basis of CC) and Przepiórkowski (2000a) (on the basis of optional GoN).

(2.86) $$\begin{bmatrix} \textit{phrase} \\ \textsc{dtrs}\ \textit{head\_struc} \end{bmatrix} \rightarrow$$

$$\left(\begin{bmatrix} \textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{comps}\ \boxed{1} & \\ \textsc{dtrs} & \begin{bmatrix} \textsc{head-dtr}|\textsc{synsem}|\textsc{local}|\textsc{category}|\textsc{valence}|\textsc{comps}\ \boxed{2} \\ \textsc{comp-dtrs}\ \boxed{3} \end{bmatrix} \end{bmatrix} \right.$$

$$\left. \wedge\ \texttt{shuffle}(\boxed{3'},\boxed{1},\boxed{2})\ \wedge\ \texttt{phrase\_to\_synsems}(\boxed{3'},\boxed{3}) \right)$$

In (2.86), we use the `shuffle` relation to specify any ordered subset $\boxed{3'}$ of elements of the COMPS list $\boxed{2}$; $\boxed{1}$ is a list of remaining elements. The `phrase_to_synsems` relation maps a list of *synsem*s (elements of the COMPS list) to the corresponding list of *sign*s (elements of the COMP-DTRS list).[43]

Second, since not all arguments have to be raised simultaneously, we have to license phrases with a number of unrealised complements. In (2.87), we replace SCHEMA2 of Pollard and Sag (1994) (cf. §1.3.1) with the SCHEMA2' in (2.87) which combines the lexical head with an arbitrary number of its complements.

(2.87) $$\textsc{Schema2'} \equiv \begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{val}|\textsc{comps}\ \textit{list(synsem)} & \\ \textsc{dtrs} & \begin{bmatrix} \textit{head-comp-struc} \\ \textsc{head-dtr}\ \textit{word} \\ \textsc{comp-dtrs}\ \textit{list(phrase)} \end{bmatrix} \end{bmatrix}$$

Unlike in SCHEMA2 proposed in Pollard and Sag (1994, p. 38), the COMPS value of the resulting phrase in SCHEMA2' is any list of *synsem* objects. Thus, a phrase with a lexical head (and only such a phrase) may have a number of complements unrealised. As the head daughter in (2.87) is an object of the sort *word*, unrealised complements can be realised only present on a different COMPS list (raised). Since *clitic* is a subsort of *synsem*, the unrealised complements can be clitics. The COMPS value in (2.87) can be empty as well, which accounts for the local realisation of clitics, e.g., (2.70a).

Finally, we assume that only a phrase with a fully saturated COMPS list can combine with the subject. Therefore, we slightly modify SCHEMA1 presented in §1.3.1.

(2.88) $$\textsc{Schema1'} \equiv \begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{val}|\textsc{subj}\ \langle\rangle & \\ \textsc{dtrs} & \begin{bmatrix} \textit{head-subj-struc} \\ \textsc{head-dtr}|\textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{val}|\textsc{comps}\ \langle\rangle \\ \textsc{subj-dtr}\ \langle \textit{phrase} \rangle \end{bmatrix} \end{bmatrix}$$

SCHEMA1' in (2.88) requires that phrases with unrealised complements are not combined with the subject, i.e., it does not license (2.89) in non-eliptic reading (indicated via '#' here).

(2.89) # Jan chciał mu kupić.
John wanted him$_{cl}$ buy$_{inf}$
'John wanted to buy him.' (eliptic)

An Example Let us illustrate the analysis of CC taking (2.71a) as an example:

(2.71a) Jan może mu chcieć go pokazać.
John may him$_{cl}$ want$_{inf}$ him$_{cl}$ show$_{inf}$
'John may want to show it to him.'

[43] Definitions of the ALE predicates, as used in our implementation of CC, corresponding to relations in (2.86), are given in Chapter 5.

A partial lexical entry for the verb *pokazać* ‘to show’ is given in (2.90).

$$(2.90)\quad \begin{bmatrix} \textit{word} \\ \text{PHON} \ \langle \text{pokazać} \rangle \\ \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \ \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \ \langle \boxed{1}\text{NP}[\textit{str}] \rangle \\ \text{COMPS} \ \langle \boxed{2}\text{NP}[\textit{str}], \boxed{3}\text{NP}[\textit{ldat}] \rangle \end{bmatrix} \\ \text{ARG-ST} \ \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The conjunct of the Valence Principle in (2.86) and the schema in (2.87) allow any number of complements to be realised in any order. Hence, the direct object, as in (2.91), or the indirect object, (2.92), can be realised first, or both complements can be realised simultaneously, (2.93).

$$(2.91)\quad \begin{bmatrix} \textit{phrase} \\ \text{PHON} \ \langle \text{pokazać go} \rangle \\ \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \ \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \ \langle \boxed{1}\text{NP}[\textit{str}] \rangle \\ \text{COMPS} \ \langle \boxed{3}\text{NP}[\textit{ldat}] \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$(2.92)\quad \begin{bmatrix} \textit{phrase} \\ \text{PHON} \ \langle \text{pokazać mu} \rangle \\ \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \ \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \ \langle \boxed{1}\text{NP}[\textit{str}] \rangle \\ \text{COMPS} \ \langle \boxed{2}\text{NP}[\textit{str}] \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

$$(2.93)\quad \begin{bmatrix} \textit{phrase} \\ \text{PHON} \ \langle \text{pokazać mu go} \rangle \\ \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD} \ \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \ \langle \boxed{1}\text{NP}[\textit{str}] \rangle \\ \text{COMPS} \ \langle \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The phrases in (2.91)–(2.93) are licensed by SCHEMA2’ in (2.87). As the complements of the verb in (2.90) are syntactically realised, they are removed from the phrase’s COMPS by the Valence Principle. The ARG-ST list is absent in (2.91)–(2.93) since, as mentioned previously, it is present only in *word*s.

The verbs *może* ‘may’ and *chcieć* ‘to want’ in (2.71a) are subject control verbs and require an infinitival complement but do not have a nominal complement. Therefore, CC to these verbs is permitted, see (2.83). The partial lexical entry for *chcieć* ‘to want’ is given in (2.94) (the lexical entry for *może* is analogous).

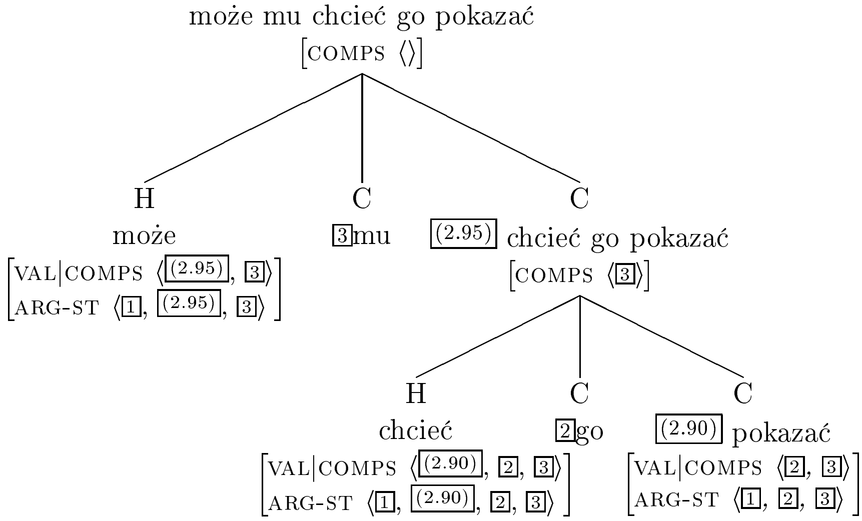
$$(2.94)\quad \begin{bmatrix} \textit{word} \\ \text{PHON} \ \langle \text{chcieć} \rangle \\ \text{SS|LOC|CAT} \begin{bmatrix} \text{HEAD} \ \textit{verb} \\ \text{ARG-ST} \ \langle \text{NP}[\textit{str}] \rangle \oplus \left\langle \begin{bmatrix} \textit{synsem} \\ \text{LC} \begin{bmatrix} \text{HEAD|VFORM} \ \textit{inf} \\ \text{VAL|COMPS} \ \boxed{0}\textit{list(synsem)} \end{bmatrix} \end{bmatrix} \right\rangle \oplus \boxed{0} \end{bmatrix} \end{bmatrix}$$

Since the COMPS value of the infinitival complement in (2.94) is not further restricted, the verb *chcieć* ‘to want’ can combine either with a phrasal or a lexical infinitival complement, i.e., either any of the phrases in (2.91)–(2.93) or the lexical verb in (2.90) can be used. A description of the phrase *chcieć go pokazać* ‘to want to show it’ , with the raised argument *go* ‘it/him’ of the verb *pokazać* ‘to show’ in (2.90), is given in (2.95).

(2.95)
$$\begin{bmatrix} \textit{phrase} \\ \text{PHON } \langle \text{chcieć go pokazać} \rangle \\ \text{SS|LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \langle \boxed{1}\text{NP}[\textit{str}] \rangle \\ \text{COMPS } \langle \boxed{3}\text{NP}[\textit{ldat}] \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Assuming that the main verb *może* 'may' combines with the phrase in (2.95), and the inherited clitic *mu* 'him' ($\boxed{3}$), a syntactic analysis of (2.71a) is sketched in (2.96).[44]

(2.96)



No Raising Across Negation As we have seen in §2.5.2, negation is an obstacle for argument raising in VCs. We account for this fact adopting the principle which prevents raising across negation proposed in Przepiórkowski (1999), given informally in (2.97).[45]

(2.97)
$$\begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{verb} \\ \text{NEG } + \end{bmatrix} \\ \text{ARG-ST } \langle \boxed{0} \rangle \oplus \boxed{1} \end{bmatrix} \rightarrow \boxed{1} \text{ list}(\text{XP}^{-})$$

According to (2.97), no non-subject arguments of a negated verb (marked as [NEG +])[46] can be raised higher up (list(XP^{-}) abbreviates a list of XP elements of ARG-ST marked as [RAISED −], i.e., a list of non-raised arguments, see Przepiórkowski (1999, pp. 93ff)).[47] Formally, we have to use a list of objects of the sort *argument* rather than simply a list of *synsem*s as the ARG-ST value (and also valence attributes), see (2.98).

(2.98)
$$\begin{bmatrix} \textit{argument} \\ \text{ARGUMENT } \textit{synsem} \\ \text{RAISED } \textit{boolean} \end{bmatrix}$$

In addition to the ARGUMENT attribute, *argument* has the RAISED attribute which specifies whether the argument is raised, [RAISED+], or not, [RAISED−]. Informally, an element of ARG-ST is [RAISED+] if and only if it appears on the immediately higher ARG-ST; see Przepiórkowski (1999, pp. 93ff) for details.

[44]In fact, only SYNSEM values of (2.94) and (2.95) should appear in (2.96).

[45]For a full formalisation see Przepiórkowski (1999, p. 428).

[46]The NEG attribute is appropriate to the sort *verb*. The value of NEG is an object of the sort *boolean*, indicated here as either '+' or '−'. Verbal negation is discussed in detail in Chapter 4.

[47]Note that XP^{-} in (2.97) is not a sort. Hence, list(XP^{-}) has a different status than, e.g., *list(synsem)* in (2.94), which we indicate by using different fonts.

Therefore, if the infinitival complement of the main verb is negated, as in (2.74), the principle in (2.97) prevents its arguments from raising to the main verb.

(2.74) a. Mogłem nie bić go wczoraj.
could-I NM beat$_{inf}$ him$_{cl}$ yesterday
'I could not beat him yesterday.'

b. ?* Nie bić mogłem go wczoraj.
NM beat$_{inf}$ could-I him$_{cl}$ yesterday
'I could not beat him yesterday.'

c. ? Nie bić go mogłem wczoraj.
NM beat$_{inf}$ him$_{cl}$ could-I yesterday
'I could not beat him yesterday.'

Since raising is blocked and all complements of the infinitival complement are shared with the main verb (the lexical entry of the verb *mogłem* '(I) could' in (2.74) is similar to (2.94)), this can happen only if all complements of the negated infinitive are realised locally, i.e., the list of inherited complements is empty. Therefore, the ungrammaticality of (2.74b) is correctly predicted and only (2.74a,c) may result.

Summary of CC in VCs We presented a lexical analysis of CC in VCs. CC is encoded as raising from COMPS to ARG-ST, which makes the subcategorization frame of the main verb compatible with either a lexical or a phrasal infinitival complement. In order to account for the optionality of CC and clitic split, we slightly liberate the Valence Principle and the ID schemata. CC out of negated phrases is blocked by a general principle which prevents raising across negation.

2.5.4 Linear Order of Clitics in VCs

In order to account for the linear order of clitics in VCs, we have to make some assumptions about the components of the VC domains. As we have seen in the previous subsection, linear order is independent of constituent structure in VCs, e.g., (2.73) and various permutations of the main verb and its complements are possible in VCs, e.g., (2.72a) vs. (2.72b). Thus, we represent elements of a VC domain as in other VP domains in §2.4.2: the main verb, infinitives as well as their dependents are represented as individual domain objects within a VC domain.

Adopting this assumption, the constraint which orders clitics in a simple sentence, (2.58), can directly account for the order of clitics in (2.70) or (2.71).

(2.70b) Jan chciał go obudzić o szóstej.
John wanted him$_{cl}$ wake$_{inf}$ at six
'John wanted to wake him up at six.'

The position of the clitic *go* 'him' in (2.70b) is licensed both by its semantic governor *obudzić* 'to wake up' (the clitic precedes the verb) and the verb *chciał* 'wanted', the clitic climbed to (*go* immediately follows this verb): according to the analysis of CC presented in §2.5.3, the clitic is present on the COMPS lists of both verbs.

However, if we also adopt the constraint in (2.59) which requires all clitics to either follow or precede the verb, examples such as (2.71a) become problematic.

(2.71a) Jan może mu chcieć go pokazać.
John may him_{cl} want_{inf} him_{cl} show_{inf}
'John may want to show it to him.'

The linear positions of clitics are licensed by the verb *pokazać* 'to show' from which both clitics originate but not by the intermediate verb *chcieć* 'to want' since one of the raised clitics precedes and the other follows this verb, contrary to the constraint in (2.59).

Similarly, the position of *go* 'him' in (2.72b) with respect to the verb *zgubić* 'to loose' (the semantic governor of the clitic) remains unaccounted for.

(2.72b) Zgubić mogłem go wczoraj.
loose_{inf} could-I him_{cl} yesterday
'I could have lost it yesterday.'

According to our analysis of CC, see §2.5.3, the clitic ($\boxed{1}$ in (2.99)) is present on the COMPS lists of both *zgubić* 'to loose' and *mogłem* '(I) could'. Following our assumptions about linear structure of Polish VCs, both verbs are represented as individual *dom_obj*s on the DOM list of a VC, as illustrated in (2.99):

(2.99)
$$\begin{bmatrix} \text{SS VC} \\ \text{DOM} \left\langle \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle \text{zgubić} \rangle \\ \text{SS|LOC|CAT|VAL|COMPS } \langle \boxed{1} \rangle \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle \text{mogłem} \rangle \\ \text{SS|LOC|CAT|VAL|COMPS } \langle \boxed{1} \rangle \end{bmatrix}, \begin{bmatrix} \textit{dom\_obj} \\ \text{PHON } \langle \text{go} \rangle \\ \text{SS } \boxed{1}\textit{clitic} \end{bmatrix}, \ldots \right\rangle \end{bmatrix}$$

The DOM value in (2.99) does not satisfy the constraint (2.58): the postverbal clitic *go* is separated from the verb *zgubić* by a non-clitic element (the fact that (2.99) satisfies the constraint (2.59) is not enough since every object admitted by the grammar must satisfy all constraints simultaneously). Hence, the position of *go* is predicted to be ungrammatical, contrary to fact.

In order to account for examples such as (2.72b), we require the clitic to be ordered with respect to the verb from which it is not raised any further and replace (2.58) and (2.59) with the constraints (2.100) and (2.101), respectively. (As in (2.97), we use here lists of *argument*s rather than lists of *synsem*s as values of valence attributes.)

(2.100) a. If the DOM list $\boxed{4}$ contains *dom_obj*s $\boxed{0}$, $\boxed{1}$, such that $\boxed{0}$ represents a verb and $\boxed{1}$ represents a non-raised argument clitic of this verb and $\boxed{0}$ precedes $\boxed{1}$ then
either $\boxed{0}$ immediately precedes $\boxed{1}$ or
there is a *dom_obj* $\boxed{2}$, which represents a non-raised argument clitic of the same verb, such that $\boxed{2}$ immediately precedes $\boxed{1}$.

b. $\texttt{lp}_4(\boxed{4}) \overset{\forall}{\Longleftarrow} \forall \boxed{0}\ \forall \boxed{1}\ \forall \boxed{5}$

$$\left(\begin{array}{l} \exists \boxed{3}\ \exists \boxed{7} \\ \left(\begin{array}{l} \texttt{member}\left(\boxed{0}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS|LOC|CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{VAL|COMPS } \boxed{5} \end{bmatrix} \end{bmatrix}, \boxed{4}\right) \wedge \texttt{member}\left(\boxed{1}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS } \boxed{3} \end{bmatrix}, \boxed{4}\right) \\ \wedge\ \texttt{member}\left(\boxed{7}\begin{bmatrix} \textit{arg} \\ \text{ARG } \boxed{3}\textit{clitic} \\ \text{RAISE } - \end{bmatrix}, \boxed{5}\right) \wedge \texttt{prec}(\boxed{0}, \boxed{1}, \boxed{4}) \end{array} \right) \\ \rightarrow \left(\begin{array}{l} \texttt{imprec}(\boxed{0}, \boxed{1}, \boxed{4})\ \vee \\ \exists \boxed{2}\ \exists \boxed{6}\ \exists \boxed{8}\ \texttt{imprec}\left(\boxed{2}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS } \boxed{6} \end{bmatrix}, \boxed{1}, \boxed{4}\right) \wedge \texttt{member}\left(\boxed{8}\begin{bmatrix} \textit{arg} \\ \text{ARG } \boxed{6}\textit{clitic} \\ \text{RAISE } - \end{bmatrix}, \boxed{5}\right) \end{array} \right) \end{array} \right)$$

(2.101) a. If the DOM list $\boxed{4}$ contains *dom_obj*s $\boxed{0}$, $\boxed{1}$, such that $\boxed{0}$ represents a verb and $\boxed{1}$ represents a non-raised argument clitic of this verb and $\boxed{0}$ precedes $\boxed{1}$ then
for all $\boxed{2}$ such that $\boxed{2}$ represents a non-raised complement of this verb
if $\boxed{2}$ precedes $\boxed{0}$ then
$\boxed{2}$ is a non-clitic.

b. $$\texttt{lp}_5(\boxed{4}) \overset{\forall}{\Longleftarrow} \forall \boxed{0}\ \forall \boxed{5}$$
$$\left(\begin{array}{l} \exists\boxed{3}\ \exists\boxed{7} \\ \left(\begin{array}{l} \texttt{member}(\boxed{0}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS|LOC|CAT} \begin{bmatrix}\text{HEAD } \textit{verb}\\ \text{VAL|COMPS } \boxed{5}\end{bmatrix}\end{bmatrix}, \boxed{4}) \wedge \texttt{member}(\boxed{1}\begin{bmatrix}\textit{dom\_obj}\\ \text{SS } \boxed{3}\textit{clitic}\end{bmatrix},\boxed{4}) \\ \wedge\ \texttt{member}(\boxed{7}\begin{bmatrix}\textit{arg}\\ \text{ARG } \boxed{3}\\ \text{RAISE } -\end{bmatrix},\boxed{4})\ \wedge\ \texttt{prec}(\boxed{0},\boxed{1},\boxed{4})\end{array}\right) \\ \to \forall\boxed{2}\ \exists\boxed{8}\ \forall\boxed{6} \left(\begin{array}{l} \left(\begin{array}{l}\texttt{prec}(\boxed{2}\begin{bmatrix}\textit{dom\_obj}\\ \text{SS } \boxed{6}\end{bmatrix},\boxed{0},\boxed{4}) \wedge \\ \texttt{member}(\boxed{8}\begin{bmatrix}\textit{arg}\\ \text{ARG }\boxed{6}\\ \text{RAISE } -\end{bmatrix},\boxed{5})\end{array}\right) \\ \to\ \boxed{6}\textit{non-clitic}\end{array}\right)\end{array}\right)$$

(2.100) differs from (2.58) only minimally: the postverbal clitic arguments cannot be further raised, [RAISE−], to a higher verb, if any. Note that this minor modification does not affect the analysis in §2.4.3: in a simple sentence no raising occurs and all arguments of a verb are uniformly marked [RAISE−].

The modification in (2.101) requires that if a verb is followed by its non-raised clitic argument, all non-raised complements preceding the verb must be non-clitics. Preverbal elements can be clitics as long as they are raised arguments of the verb or they are not complements of the verb. The linear positions of the clitics *mu* 'him$_{dat}$' and *go* 'him$_{acc}$' in (2.71a) are now licensed by the verb *chcieć* 'to want' since the first clitic (*mu*) is raised ([RAISE+]) from the verb to the main verb *może* 'may', unlike the clitic *go* ([RAISE−]).

These modifications allow us also to account for (2.72b). The DOM list with indicated RAISE values is given in (2.102).

(2.102) $$\begin{bmatrix} \text{SS VC} \\ \text{DOM } \left\langle \begin{bmatrix}\textit{dom\_obj}\\ \text{PHON } \langle\text{zgubić}\rangle \\ \text{SLCV|COMPS } \left\langle \begin{bmatrix}\textit{arg}\\ \text{ARG } \boxed{1}\\ \text{RAISE } +\end{bmatrix}\right\rangle\end{bmatrix}, \begin{bmatrix}\textit{dom\_obj}\\ \text{PHON } \langle\text{mogłem}\rangle \\ \text{SLCV|COMPS } \left\langle \begin{bmatrix}\textit{arg}\\ \text{ARG } \boxed{1}\\ \text{RAISE } -\end{bmatrix}\right\rangle\end{bmatrix}, \begin{bmatrix}\textit{dom\_obj}\\ \text{PHON } \langle\text{go}\rangle\\ \text{SS } \boxed{1}\textit{clitic}\end{bmatrix}, \ldots \right\rangle \end{bmatrix}$$

Since the clitic *go* 'him' immediately follows the verb *mogłem* '(I) could', for which the clitic is [RAISE−], the linear position of the clitic is licensed by (2.100) (for the verb *zgubić* 'to loose' the clitic is [RAISE+] and *go* does not have to immediately succeed this verb).

2.5.5 CC in Tense Auxiliary Constructions

Tense Auxiliaries Unlike Romance or Germanic languages, Polish employs only one tense auxiliary verb *być* 'be'. The tense auxiliary *być* may occur in two forms, weak and strong, each used in a different analytic verb form.

The future tense of imperfective verbs is formed with the strong form of the auxiliary:

(2.103) Ty będziesz widział ten film.
you AUX.FUT.2sg seen this film
'You will see this film.'

As observed in Borsley and Rivero (1994), verb forms usually called past tense, (2.104a), and conditionals, (2.104b), consist of a weak auxiliary (indicated in *italic*) and the participle (verb forms such as *widział* 'saw' in (2.103) and (2.104), are also called *l*-participle):

(2.104) a. Ty widział*eś* ten film.
you seen AUX.PAST.2sg this film
'You saw this film.'

b. Ty widział*byś* ten film.
you seen AUX.COND.2sg this film
'You would see this film.'

Although the orthography of verb forms in (2.104) suggests that weak auxiliaries are verbal inflections, data are more complex than that.

Weak auxiliaries can be detached from the participle and occur on almost any other preceding word, (2.105) (see Booij and Rubach (1987, p. 34) or Spencer (1991, p. 372) for more examples).[48,49]

(2.105) a. Ty*ś* widział ten film.
you AUX.PAST.2sg seen this film
'You saw this film.'

b. Ty *byś* widział ten film.
you AUX.COND.2sg seen this film
'You would see this film.'

The behaviour of Polish 'floating inflections', illustrated in (2.104) and (2.105), has been often discussed in the literature, e.g., Mikoś and Moravcsik (1986), Booij and Rubach (1987), Rappaport (1988), Spencer (1991), Borsley and Rivero (1994), Witkoś (1998), Borsley (1999), Franks and Bański (1999). We follow here Borsley and Rivero (1994) and Borsley (1999) and treat all analytic verb forms in (2.103)–(2.105) as consisting of an auxiliary (weak or strong) and a participle.

Unlike the future auxiliary, (2.106a), the past or conditional auxiliaries cannot occur in the sentence initial position, (2.106b).

(2.106) a. Będziesz pomagał Piotrowi.
AUX.FUT.2sg helped Peter
'You will be helping Peter.'

b. * *Ś* / *Byś* Piotrowi pomagał.
AUX.PAST.2sg AUX.COND.2sg Peter helped

[48] Transitive prepositions and the verbal negative marker *nie* 'not' are the exceptions discussed in Borsley and Rivero (1994). Franks and Bański (1999, pp. 127ff) consider also certain phrase-level phonological restrictions on hosts of weak auxiliaries.

[49] Past tense auxiliaries are morphologically realised only in the first and second person verb forms. The past auxiliary is always written together with the element it is attached to, whether a participle, (2.104a), or not, (2.105a). The orthography of conditional auxiliaries distinguishes them as separate elements, (2.105b), unless the auxiliary follows the participle, (2.104b).

‘You were helping / would be helping Peter.’ (putatively)

c. Maria pytała czy*ś* / *byś* Piotrowi pomagał.
Mary asked Q AUX.PAST.2sg AUX.COND.2sg Peter helped
‘Mary asked whether you had been / would be helping Peter.’

The contrast between (2.106b) and (2.106c) indicates that past tense and conditional auxiliaries are phonological enclitics, i.e., they are weak (prosodically deficient), unlike the future auxiliary, see (2.106a).

As discussed in Franks (1998), Franks and Bański (1999) and Borsley (1999), among others, weak auxiliaries act differently if they are attached directly to the participle, as in (2.104), and if attached to any other element, as in (2.105) or (2.106c). Evidence confirming such a behaviour comes from prosody, phonology, word order and ellipsis facts. At least for some speakers, postverbal auxiliaries may form a phonological and a prosodic word with the phonological host, see (2.107a), but this is never the case if the auxiliary is detached from the participle, (2.107b), see Franks and Bański (1999).[50]

(2.107) a. POszli*śmy* / ?poszLI*śmy*
went AUX.PAST.1pl went AUX.PAST.1pl
‘(we) went’

b. dlaCZEgo*śmy* / *dlaczeGO*śmy*
why AUX.PAST.1pl why AUX.PAST.1pl
‘why have we’

Then, the weak auxiliary may be omitted in coordination if detached from the participle, (2.108a), but such an omission is ungrammatical if the auxiliary follows the verbal host, (2.108b), Franks and Bański (1999).

(2.108) a. Zmęczony*m* i głodny.
tired AUX.PAST.1sg and hungry
‘I’m tired and hungry.’

b. * Czytał*em* i pisał.
read AUX.PAST.1sg and wrote
‘I read and wrote.’

Finally, word order facts show that only the auxiliary preceding the participle can be separated from it by other elements, cf. (2.105) or (2.106c), Borsley (1999).

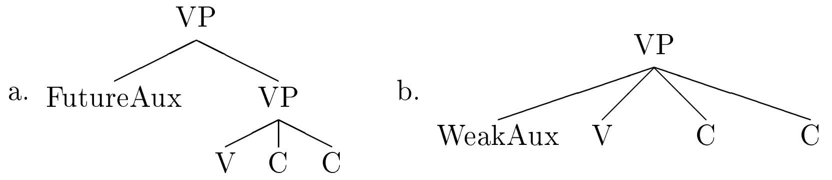
In order to account for the dual nature of weak auxiliaries, Borsley (1999) treats them as syntactic clitics (postlexical clitics, PLC) but assumes that they are not always combined with the participle in the same way. If the weak auxiliary is attached directly to the participle, as in (2.104), they form a morphological compound (i.e., they are combined in morphology). On the other hand, if the weak auxiliary is detached from the participle, as in (2.105), they form a syntactic construction (similarly to the future tense auxiliary constructions, see (2.103)). Although we largely agree with such an analysis, our view on the syntactic structure of tense auxiliary constructions is different from that in Borsley (1999).[51]

[50] CAPITAL letters mark lexical stress.

[51] Borsley (1999) treats past tense auxiliaries on a par with conditional ones. As he admits, the behaviour of both auxiliaries differs in certain respects, e.g., prosody and phonology, Borsley (1999, p. 32), as well as postverbal positions, see Borsley (1999, p. 59, fn. 12). Since weak auxiliaries are not the main topic of our interest, we content ourselves with Borsley’s analysis.

Borsley (1999) represents weak auxiliaries as subject raising verbs which subcategorize for a participle, similarly to the future tense auxiliary. He argues, however, that the strong and weak auxiliaries should have different complementations. The former arguably take a VP complement and a traditional hierarchical structure results, (2.109a). On the other hand, if a weak auxiliary precedes the participle (i.e., a syntactic construction is formed), the auxiliary subcategorizes for a lexical verbal complement (the participle) and its complements, which results in a flat syntactic structure, (2.109b) (such a flat syntactic structure has been previously proposed for French and Italian tense auxiliary constructions, cf. Abeillé and Godard (1994) and Monachesi (1999), respectively).

(2.109)



Such an analysis, however, makes incorrect predictions with respect to CC in Polish.

CC in Tense Auxiliaries In Polish, unlike in Romance, CC to tense auxiliaries is optional. As illustrated in (2.110), the argument clitic is either raised to the future auxiliary, (2.110a), or remains 'downstairs' on the participle, (2.110b).

(2.110) a. Będę go oglądał jutro.
AUX.FUT.1sg him$_{cl}$ watched tomorrow
'I'll be watching him tomorrow.'

b. Będę oglądał go jutro.

The behaviour of Polish pronominal clitics in past tense and conditional auxiliary constructions is analogous: clitics are facultatively realised on the auxiliary, cf. (2.111)–(2.112).

(2.111) a. Często*ś* go widywał przedtem.
often AUX.PAST.2sg him$_{cl}$ seen before
'You saw him often before.'

b. Często*ś* widywał go przedtem.

(2.112) a. Chętnie *bym* go obejrzał jutro.
willingly AUX.COND.1sg him$_{cl}$ watched tomorrow
'I would like to watch it/him tomorrow.'

b. Chętnie *bym* obejrzał go jutro.

Also if there are several clitics, they occur on the auxiliary only optionally, cf. (2.113a–b). It seems that, unlike in VCs (cf. §2.5.2), in auxiliary constructions all clitics must be realised on the same verb, cf. (2.113c–d).

(2.113) a. Dokładnie*ś* / *byś* pokazał mu go wczoraj.
carefully AUX.PAST.2sg AUX.COND.2sg showed him$_{cl}$ him$_{cl}$ yesterday
'Yesterday you showed it to him carefully.'

b. Dokładnie*ś* / *byś* mu go pokazał wczoraj.

c. ?? Dokładnie*ś* / *byś* mu pokazał go wczoraj.

d. ?? Dokładnie*ś* / *byś* go pokazał mu wczoraj.

Properties of CC in the future tense auxiliary constructions are analogous.

(2.114) a. Będzie pokazywał mu go jutro.
AUX.FUT.3sg show him$_{cl}$ him$_{cl}$ tomorrow
'He will be showing it to him tomorrow.'

b. Będzie mu go pokazywał jutro.

c. ? Będzie mu pokazywał go jutro.

d. ? Będzie go pokazywał mu jutro.

Polish pronominal clitics are syntactic items, see §2.2. Therefore, if the two subcategorization frames in (2.109) are adopted, CC to the future tense auxiliary shouldn't be possible: complements of the participle cannot be raised to the auxiliary according to (2.109a). On the other hand, all clitics should be raised to the weak auxiliary, according to (2.109b), rather than occur on the participle. As (2.110)–(2.114) show, these expectations are not born out: all types of auxiliaries optionally trigger CC, i.e., clitics appear either on the participle or on the auxiliary.

Against Two Complementations The main argument put forward in Borsley (1999) for two distinct structures given in (2.109), is the contrast in (2.115).

(2.115) a. * Widział książkę*ś* / *byś*.
seen book AUX.PAST.2sg AUX.COND.2sg
'You saw/would see the book.'

b. Widział książkę będziesz.
seen book AUX.FUT.2sg
'You will see the book.'

This contrast apparently shows that in (2.115a), unlike in (2.115b), there is no constituent to be preposed. Since Polish is a relatively free word order language and syntactic constituents do not have to be linearly contiguous (see §2.2), it is not clear whether the participle forms a constituent with the complement *książkę* 'a book' in (2.115b).

Other traditional constituency tests, e.g., coordination, (2.116), or pronominalization, (2.117), do not confirm the difference in subcategorization frames of weak or strong auxiliaries.

(2.116) a. Ty będziesz czytał książki i pisał wiersze.
you AUX.FUT.2sg read books and write poems
'You will be reading books and writing poems.'

b. Ty*ś* / *byś* czytał książki i pisał wiersze.
you AUX.FUT.2sg AUX.COND.2sg read books and write poems
'You were / would be reading books and writing poems.'

(2.117) * Ty*ś* / *byś* / będziesz tego.
you AUX.PAST.2sg AUX.COND.2sg AUX.FUT.2sg this

In (2.116), coordination of the auxiliary's complements is possible irrespectively of whether the complement forms an alleged constituent, (2.116a), or not, (2.116b). Similarly, pronominalization of the auxiliary's complements is impossible regardless of their syntactic structure postulated in (2.109).

We conclude that the contrast indicated in (2.115) is not a strong argument in favour of two distinct syntactic structures in (2.109). We attribute the ungrammaticality of (2.115a) to general restrictions on the linear order of Polish clitics rather than to the syntactic structure of tense auxiliaries: the distribution of weak auxiliary clitics is analogous to that of pronominal clitics (see §2.2.5).

Syntactic CC CC in tense auxiliaries (as in VCs, cf. §2.5.2) cannot be explained by the word order alone. In (2.118a), if no syntactic raising took place, *go* would be an argument of the verb *oglądał* 'watched'. In this case, the clitic *go* would be separated by non-clitics from the verb it is realised on (i.e., *oglądał*), which is not possible in (2.118b) or in general, see §2.2.5.

(2.118) a. Oglądał jutro będę go w TV.
watched tomorrow AUX.FUT.1sg him_{cl} in TV
'I'll be watching it/him tomorrow on TV.'

b. * Będę oglądał jutro go w TV.
AUX.FUT.1sg watched tomorrow him_{cl} in TV

The unexpected grammaticality of (2.118a) is easily explained once we assume that the clitic is syntactically raised to the auxiliary in (2.118a). The linear position of the clitic in (2.118a) is licensed by the familiar constraint which orders pronominal clitics with respect to the verb from which they are not raised any further (see §2.5.4).

Since the weak auxiliary cannot follow the participle, (2.115a), a similar argumentation cannot be applied to weak auxiliary constructions. The equivalent of (2.118) with a weak auxiliary is ungrammatical for independent reasons:

(2.119) * Oglądał często*m* / *bym* go w TV.
watched often AUX.PAST.1sg AUX.COND.1sg him_{cl} in TV
'I was/would be watching it/him often on TV.'

There seems to be no explicit evidence for syntactic CC in weak tense auxiliary constructions. Thus, one may argue that in weak auxiliary constructions, unlike in future tense auxiliaries, no syntactic CC occurs (clitics are always realised as arguments of the participle). Such an account would require a flat syntactic structure for (2.118a) (due to the syntactic CC) and a hierarchical structure for, e.g., (2.111a) and (2.112a) (no syntactic CC). As we have seen in (2.116)–(2.117), however, the existence of two distinct syntactic structures for auxiliary constructions is not supported by constituency tests.

We conclude that in all tense auxiliary constructions, CC happens in syntax.

CC and Negation As has been observed, e.g., in Borsley and Rivero (1994), Dornisch (1997), Witkoś (1998), Borsley (1999), the sentential negative marker (NM) *nie* cannot precede weak auxiliaries, (2.120b) and (2.121b), but has to be attached to the morphological complex, cf. (2.120a) and (2.121a), or to the participle, (2.120c) and (2.121c).[52]

[52]In Chapter 4 we show that, contrary to the orthography, the NM is a verbal prefix, which explains the fact that some verbs do not have negated forms.

(2.120) a. Nie widział*em* wczoraj Jana.
NM saw AUX.PAST.1sg John
'I did not see John yesterday.'

b. * Nie*m* widział wczoraj Jana.

c. Wcale*m* go wczoraj nie widział.
at-all AUX.PAST.1sg him$_{cl}$ yesterday NM saw
'Yesterday I haven't seen him at all.'

(2.121) a. Nie widział*bym* wtedy Jana.
NM saw AUX.COND.1sg then John
'I would not see John at that time.'

b. * Nie *bym* widział wtedy Jana.

c. Wcale *bym* go wtedy nie widział.
at-all AUX.COND.1sg him$_{cl}$ then NM saw
'I wouldn't see him at all at that time.'

As we have seen in §2.5.2, raising across negation is impossible in Polish. Therefore, we assume that the clitic *go* in (2.120c) and (2.121c) forms a linearly discontinuous syntactic phrase with the negated participle: the clitic is realised locally as an argument of the participle which it linearly precedes.

Weak auxiliaries contrast with the strong auxiliary with respect to negation. As (2.122) shows, the future tense auxiliary *będzie* 'will be' can be overtly negated, (2.122a), but its participial complement cannot, (2.122b), see Dornisch (1997), Witkoś (1998, p. 232).[53]

(2.122) a. Piotr nie będzie jutro śpiewał.
Peter NM AUX.FUT.3sg tomorrow sing
'Peter will not sing tomorrow.'

b. # Piotr będzie jutro nie śpiewał.

c. Piotr go nie będzie jutro widział.
Peter him$_{cl}$ NM AUX.FUT.3sg tomorrow see
'Peter will not see him tomorrow.'

Unlike in (2.120c) and (2.121c), the clitic *go* in (2.122c) is allowed to climb to the auxiliary since the participle is a non-negated verb.

Summary of CC in Tense Auxiliaries In order to account for CC in tense auxiliary constructions, we adopt the analysis of tense auxiliaries proposed in Borsley (1999). We assume that weak auxiliaries either form a syntactic construction with the following participle or, if the participle immediately precedes a weak auxiliary, they form a morphological compound. Contrary to Borsley (1999), however, we do not see a motivation for distinct subcategorization frames of weak and strong auxiliaries. We explain different properties of constructions with weak and strong auxiliaries by restrictions on the linear order rather than by their distinct syntactic structures.

[53](2.122b) is grammatical only in the contrastive meaning, i.e., if *nie* is not the sentential negative marker:

(i) Piotr będzie jutro nie śpiewał, lecz tańczył.
Peter AUX.FUT.3sg tomorrow not sing but dance
'Peter will not sing but dance tomorrow.'

2.5.6 An HPSG Analysis of CC in Auxiliary Constructions

Complementation of Auxiliaries Since there are no strong arguments supporting the two complementations proposed in (2.109) by Borsley (1999), we assume the following general subcategorization frame for all auxiliaries:[54]

(2.123)
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AUX} & + \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \left\langle \boxed{0} \begin{bmatrix} \textit{synsem} & \\ \text{LC} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{l-part} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \oplus \boxed{2} \\ \text{ARG-ST} & \boxed{1} \oplus \langle \boxed{0} \rangle \oplus \boxed{2} \end{bmatrix} \end{bmatrix}$$

As in Borsley (1999), auxiliaries are subject raising verbs, indicated as the token identity of the SUBJ values ($\boxed{1}$), which ensures agreement between the participle (*l-part*) and the subject in all auxiliary constructions.

Unlike in Borsley (1999), the value of the participle's COMPS list, i.e., $\boxed{2}$, is not fully specified. This makes the subcategorization frame of the auxiliary compatible with either a lexical or a phrasal participle complement. Therefore, $\boxed{2}$ may be empty if no complements are raised, or non-empty if some (possibly all) complements of the participle are raised. Since clitics are present on the COMPS list, they are raised optionally as well.

Lexical Entries of Auxiliaries As in the case of pronominal clitics, cf. §2.3, we do not introduce an additional subsort of *clitic* to represent verbal clitics. We represent weak auxiliaries on a par with other verbs but their SYNSEM value is an object of the sort *clitic* (rather than *non-clitic* as in the case of the future tense auxiliary).

As we have seen in (2.120)–(2.122), weak and strong auxiliaries behave differently with respect to negation. As in Borsley (1999, p. 42), these differences are captured in the lexical entries of auxiliaries, cf. (2.124) and (2.125).

(2.124)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{ś} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \textit{clitic} & \\ \text{LOC}|\text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AUX} & + \\ \text{NEG} & - \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \left\langle \boxed{0} \begin{bmatrix} \text{LC} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{l-part} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \oplus \boxed{2} \end{bmatrix} \\ \text{ARG-ST} & \boxed{1} \oplus \langle \boxed{0} \rangle \oplus \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

[54]The attribute AUX is appropriate to the *verb* sort and its value is an object of the sort *boolean*. The sort *l-part* is a subsort of *vform*.

(2.125)
$$\begin{bmatrix} \textit{word} \\ \text{PHON } \langle \text{będziesz} \rangle \\ \text{SYNSEM} \begin{bmatrix} \textit{non-clitic} \\ \text{LOC|CAT} \begin{bmatrix} \text{HEAD} \begin{bmatrix}\textit{verb}\\ \text{AUX } + \end{bmatrix} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \boxed{1} \\ \text{COMPS } \left\langle \boxed{0} \begin{bmatrix} \text{LC} \begin{bmatrix} \text{HEAD} \begin{bmatrix} \textit{verb} \\ \text{VFORM } \textit{l-part} \\ \text{NEG } - \end{bmatrix} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \boxed{1} \\ \text{COMPS } \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix} \right\rangle \oplus \boxed{2} \end{bmatrix} \\ \text{ARG-ST } \boxed{1} \oplus \langle \boxed{0} \rangle \oplus \boxed{2} \end{bmatrix}\end{bmatrix}\end{bmatrix}$$

The weak auxiliary in (2.124) and the participle in (2.125) are marked as [NEG −], which ensures that they are not negated. On the other hand, the NEG value of the participle in (2.124) and the auxiliary in (2.125) is left unspecified, which means that they may be either negated or not, see Chapter 4 for details.

We treat weak and strong auxiliaries as lexical items, i.e., they are entities of the sort *word*. Unlike in Borsley (1999), no subsorts of *word* are introduced. Instead, we employ the split of *synsem*s introduced in §2.3 and use the standard AUX attribute to mark auxiliaries, [AUX +]. This allows us to avoid complex lexical hierarchies adopted in Borsley (1999) from Riehemann (1998).

Morphological Compounds After Borsley (1999), we assume that a weak auxiliary forms a morphological compound with the immediately preceding participle.

The lexical description of past tense and conditional analytic morphological compounds in (2.126) incorporates our modifications (the split of *synsem* and elimination of lexical subsorts) and is equivalent to that proposed in Borsley (1999, p. 47).[55]

(2.126)
$$\begin{bmatrix} \textit{word} \\ \text{PHON } f(\boxed{1}, \boxed{2}) \\ \text{MORPH-B } \left\langle \begin{bmatrix} \textit{word} \\ \text{PHON } \boxed{1} \\ \text{SS } \boxed{3} \begin{bmatrix} \textit{non-clitic} \\ \text{LC} \begin{bmatrix} \text{HEAD|VFORM } \textit{l-part} \\ \text{VAL|COMPS } \boxed{5}\end{bmatrix}\end{bmatrix}\end{bmatrix}, \begin{bmatrix} \textit{word} \\ \text{PHON } \boxed{2} \\ \text{SS} \begin{bmatrix} \textit{clitic} \\ \text{LC} \begin{bmatrix} \text{HEAD|AUX } + \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \boxed{4} \\ \text{COMPS } \langle \boxed{3} \rangle \oplus \boxed{5} \end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix} \right\rangle \\ \text{SYNSEM} \begin{bmatrix} \textit{non-clitic} \\ \text{LOC|CAT} \begin{bmatrix} \text{HEAD|VFORM } \textit{personal} \\ \text{VAL} \begin{bmatrix} \text{SUBJ } \boxed{4} \\ \text{COMPS } \boxed{5}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Borsley (1999) uses the MORPH-B attribute (after Riehemann (1998)) to encode morphological components of compound words. The phonology of the compound is specified as a function of PHON values of these components. Phonologically, the participle, $\boxed{1}$, appears first, followed by the weak auxiliary, $\boxed{2}$. We assume, after Borsley (1999), that the function f, which

[55] Formally, this description should probably be added as a disjunct to the consequent of the Word Principle given in (1.11) in §1.3.1. Note that (2.126) is impoverished with respect to lexical entries, e.g., its PHON value is not described as PHON values of specific words. Hence, its status is different from other disjuncts in the Word Principle.

combines both phonological strings in (2.126), is more sophisticated than a simple concatenation. The function has to account for prosodic and phonological processes which appear when the participle is suffixed, e.g., the vowel *-e-* is inserted between the participle and the second person masculine singular past tense auxiliary *ś*, see (2.104a). We do not attempt to specify this function here.

The second element of MORPH-B satisfies the lexical description of a weak auxiliary, e.g., as in (2.124). Therefore, negation of weak auxiliaries is blocked also if they appear as part of the morphological compound. The NEG value of the resulting compound is left unspecified. Hence, the entire complex can be either a negated, cf. (2.120a), (2.121a), or a non-negated personal verb form, cf. (2.104).

Note that the subcategorization frame of the resulting morphological complex differs from that of the auxiliary. The first element of the auxiliary's COMPS list, $\boxed{3}$, is satisfied by the participle in morphology. The COMPS list of the compound contains only unrealised arguments of the participle, $\boxed{5}$, which remain to be realised in syntax.

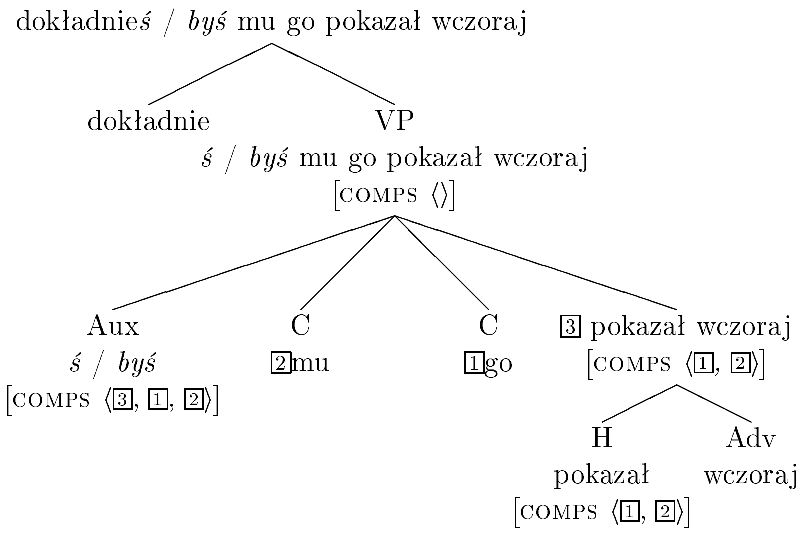
Syntactic Realisation Both pronominal clitics and weak auxiliaries are syntactic items and, in syntactic constructions, they are realised in syntax. We use the same syntactic principles and schemata as in §2.5.2 to account for CC in tense auxiliary constructions: a version of the Valence Principle which permits to realise arguments of a lexical head in any number and order, and modified SCHEMA2' which licenses phrases with unrealised arguments, see (2.86) and (2.87), respectively.

Assuming that the participle combines first with the adjunct *wczoraj* 'yesterday', a schematic analysis of CC in (2.113b) is given in (2.127).

(2.113b)	Dokładnieś	/ *byś*	mu	go	pokazał	wczoraj.
	carefully AUX.PAST.2sg	AUX.COND.2sg	him_{cl}	him_{cl}	showed	yesterday

'Yesterday you showed it to him carefully.'

(2.127)



In (2.127), the two pronominal clitic arguments are present on the COMPS list, see §2.3.4. Since the complements of the participle *pokazał* '(he) showed' are not realised locally, they have to be raised to the auxiliary, see (2.124). SCHEMA2' combines the lexical head (the auxiliary) with all its complements and the COMPS list of the resulting phrase is empty.

As noted above, the contrast between (2.113c–d) and (2.113b) indicates that, unlike in VCs, in auxiliary constructions all clitics should be realised on the same verb (an auxiliary or a participle).

(2.113)	c. ??	Dokładnieś	/ *byś*	mu	pokazał	go	wczoraj.
		carefully AUX.PAST.2sg	AUX.COND.2sg	him$_{cl}$	showed	him$_{cl}$	yesterday

'Yesterday you carefully showed it to him.'

d. ?? Dokładnieś / *byś* go pokazał mu wczoraj.

The syntactic mechanisms described above license examples (2.113c–d) since the Valence Principle and SCHEMA2' do not force realisation of all arguments at a time. Since this degree of flexibility is necessary for CC in VCs, we do not restrict general principles but exclude (2.113c–d) by the lexical constraint in (2.128).[56]

(2.128)

$$
\forall \boxed{1}\ \exists \boxed{2}
$$

$$
\left(\begin{bmatrix} \textit{word} \\ \textsc{ss}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} \begin{bmatrix} \textit{verb} \\ \textsc{aux}\ + \end{bmatrix} \vee \begin{bmatrix} \textit{verb} \\ \textsc{vform}\ \textit{l-part} \end{bmatrix} \\ \textsc{val}|\textsc{comps}\ \text{list}(\begin{bmatrix} \textit{arg} \\ \textsc{arg}\ \textit{non-clitic} \end{bmatrix}) \bigcirc \boxed{1}\text{list}(\boxed{2} \begin{bmatrix} \textit{arg} \\ \textsc{arg} \begin{bmatrix} \textit{clitic} \\ \textsc{loc}|\textsc{cont}\ \textit{pron} \end{bmatrix} \end{bmatrix}) \end{bmatrix} \end{bmatrix} \right.
$$

$$
\wedge\ \texttt{member}(\begin{bmatrix} \textit{arg} \\ \textsc{arg} \begin{bmatrix} \textit{clitic} \\ \textsc{loc}|\textsc{cont}\ \textit{pron} \end{bmatrix} \\ \textsc{raise}\ - \end{bmatrix}, \boxed{1})
$$

$$
\left. \rightarrow \boxed{1}\text{list}(\begin{bmatrix} \textit{arg} \\ \textsc{arg} \begin{bmatrix} \textit{clitic} \\ \textsc{loc}|\textsc{cont}\ \textit{pron} \end{bmatrix} \\ \textsc{raise}\ - \end{bmatrix}) \right)
$$

As in §2.5.2 and §2.5.4, we use lists of *argument*s as values of valence attributes. The lexical constraint in (2.128) should be read as a logical implication which applies to *word*s which are either auxiliaries or their participial complements (*l-part*). More specifically, the antecedent of (2.128) says that if such a lexical verb has a pronominal clitic which is not further raised ($\begin{bmatrix} \textit{arg} \\ \textsc{arg} \begin{bmatrix} \textit{clitic} \\ \textsc{loc}|\textsc{cont}\ \textit{pron} \end{bmatrix} \\ \textsc{raise}\ - \end{bmatrix}$ is a member of the COMPS list of the verb), the consequent requires that none of its clitic co-arguments can be raised either (the list of clitics, $\boxed{1}$, contains only elements marked as [RAISE−]). Since the auxiliary allows complements of the participle to be raised (see (2.123)), and (2.128) forbids clitics to be raised if at least one clitic is not raised, these two conditions are not contradictory only if the list of raised elements does not contain clitics and all clitics are realised on the same verb.

2.5.7 Linear Order of Clitics in Auxiliary Constructions

Auxiliary Clitics The distribution of weak auxiliaries is relatively free and they can appear in almost any position preceding the participle, see p. 74, but positions such as in (2.115a) are ungrammatical.

[56] Technically, lists such as $\boxed{1}$ in (2.128) are functionally written relations, see fn. 47 on p. 70.

(2.115a)* Widział książkęś / *byś*.
seen book AUX.PAST.2sg AUX.COND.2sg
'You saw/would see the book.'

The LP constraint in (2.129) licenses only pre-participial syntactic positions of weak auxiliaries and excludes incorrect sentences such as (2.115a).

(2.129) a. If the DOM list $\boxed{4}$ contains *dom_obj*s $\boxed{0}$, $\boxed{3}$, such that $\boxed{0}$ represents an auxiliary clitic and $\boxed{3}$ represents a participle complement of the auxiliary then
$\boxed{0}$ must precede $\boxed{3}$.

b. $$\texttt{lp}_6(\boxed{4}) \stackrel{\forall}{\Longleftarrow} \forall \boxed{0} \forall \boxed{3}$$
$$\left(\begin{array}{l} \exists\boxed{2}\ \exists\boxed{1} \\ \left(\begin{array}{l} \texttt{member}\left(\boxed{0}\begin{bmatrix} \textit{dom\_obj} \\ \text{SS}\ \begin{bmatrix}\textit{clitic}\\ \text{LC}\ \begin{bmatrix}\text{HEAD}\ \begin{bmatrix}\textit{verb}\\ \text{AUX}\ +\end{bmatrix}\\ \text{COMPS}\ \boxed{1}\end{bmatrix}\end{bmatrix}\end{bmatrix}, \boxed{4}\right) \\ \wedge\ \texttt{member}\left(\boxed{3}\begin{bmatrix}\textit{dom\_obj}\\ \text{SS}\ \boxed{2}\begin{bmatrix}\textit{non-clitic}\\ \text{LC|HEAD|VFORM}\ \textit{l-part}\end{bmatrix}\end{bmatrix}, \boxed{4}\right)\ \wedge\ \texttt{member}(\boxed{2},\boxed{1}) \end{array}\right) \\ \rightarrow\ \texttt{prec}(\boxed{0},\boxed{3},\boxed{4}) \end{array}\right)$$

This constraint differs from that proposed in Borsley (1999). Instead of ordering a lexical participle with respect to *any* auxiliary, (2.129) refers only to weak auxiliaries (*clitic*s) and does not require the participle (*l-part*) to be lexical. Note that (2.129) is formulated in terms of *dom_obj*s and does not refer to syntactic constituents directly. Due to such a representation of linear order, the syntactic structure of auxiliary constructions is in principle inessential for (2.129).

Since LP constraints are encoded as general well-formedness conditions, they apply to all *sign*s. By (2.129), whenever an auxiliary clitic is separated from the participle, it must precede the latter. Hence, only those linear positions of weak auxiliaries are licensed where they precede the participle, e.g., (2.105), (2.111) (sentence initial positions, as in (2.106b), are excluded by prosody and remain unaccounted for in the present syntactically-driven approach).

Pronominal Clitics and Auxiliary Clitics As we said in §2.4.3, we assume that the relative order of Polish pronominal clitics is not fixed and we do not restrict their relative positions.

As noted in Rappaport (1988, p. 320), Witkoś (1998, p. 185), among others, auxiliary and pronominal clitics do not have to cluster (although there is such a possibility if all clitics precede the participle as witnessed, e.g., by (2.111a), (2.112a) or (2.113b)):

(2.130) a. Jan *by* często go wtedy pożyczał Piotrowi.
John AUX.COND.3sg often him$_{cl}$ then borrowed to-Peter
'John would often borrow it to Peter at that time.'

b. Jan będzie często go wtedy pożyczał Piotrowi.
John AUX.FUT.3sg often him$_{cl}$ then borrowed to-Peter
'John will often borrow it to Peter at this time.'

If (2.130) exemplify CC, the constraint (2.100) licenses the linear position of *go* 'him' with respect to the participle but not the auxiliary: the raised pronominal clitic is separated by the non-clitic *często* 'often' from the preceding verb to which it is raised. As syntactic constituents in Polish are not necessarily linearly contiguous, we assume that (2.130) does not exemplify CC but only the linear precedence of the non-raised clitic *go* with respect to the participle *pożyczał* 'borrowed'. In such a case, the pronominal clitic is [RAISE−] on the COMPS list of the participle and the linear position of *go* is fully licensed by (2.100).

The relative order of pronominal and auxiliary clitics in Polish is often considered to be fixed. As noted in Rappaport (1988, p. 310), among others, auxiliary clitics precede pronominal clitics.

(2.131) a. Jan *by* go często pożyczał Piotrowi.
John AUX.COND.3sg him$_{cl}$ often borrowed to-Peter
'John would often borrow it to Peter.'

b. * Jan go *by* często pożyczał Piotrowi.

The LP constraint in (2.132) captures this fact.

(2.132) a. If the DOM list $\boxed{4}$ contains *dom_obj*s $\boxed{1}$, $\boxed{2}$, such that $\boxed{1}$ represents an auxiliary clitic and $\boxed{2}$ represents a pronominal clitic then

$\boxed{1}$ precedes $\boxed{2}$

b. $\texttt{lp}_7(\boxed{4}) \stackrel{\forall}{\Longleftarrow}$

$$\forall \boxed{1} \forall \boxed{2} \left(\left(\begin{array}{l} \texttt{member}\left(\boxed{1} \begin{bmatrix} \textit{dom\_obj} \\ \text{SS} \begin{bmatrix} \textit{clitic} \\ \text{LOC}|\text{CAT}|\text{HEAD} \begin{bmatrix} \textit{verb} \\ \text{AUX} \ + \end{bmatrix} \end{bmatrix} \end{bmatrix}, \boxed{4}\right) \wedge \\ \texttt{member}\left(\boxed{2} \begin{bmatrix} \textit{dom\_obj} \\ \text{SS} \begin{bmatrix} \textit{clitic} \\ \text{LOC}|\text{CONT} \ \textit{pron} \end{bmatrix} \end{bmatrix}, \boxed{4}\right) \end{array} \right) \rightarrow \texttt{prec}(\boxed{1}, \boxed{2}, \boxed{4}) \right)$$

The linear position of the pronominal clitic in (2.133) is not a counterexample to this restriction since we treat the verb *pożyczałby* 'would borrow' as a morphological complex (see (2.126)). The auxiliary *by* is invisible for the syntactic ordering here.

(2.133) Jan go często pożyczał*by* Piotrowi.
John him$_{cl}$ often borrowed AUX.COND.3sg to-Peter
'John would often borrow it to Peter.'

Franks (1998, p. 85), however, quotes the data in (2.134) (after Witkoś (1993), who considers them marginal) where a pronominal clitic serves as a phonological host for the auxiliary enclitic and precedes the auxiliary.

(2.134) a.?# Ty go*ś* widział często w parku.
you him$_{cl}$ AUX.PAST.2sg seen often in park
'You've seen him often in the park.'

b. Bardzo mu*ś* pomógł.
very him$_{cl}$ AUX.PAST.2sg helped
'You've helped him a lot.'

We admit that these examples sound better than (2.131b). Note, however, that the examples in (2.135) are clearly degraded with respect to (2.134).

(2.135) a. ?* Wy go*ście* / *byście* widzieli w parku.
you him$_{cl}$ AUX.PAST.2pl AUX.COND.2pl seen in park
'You saw / would have seen him in the park.'

b. ?* Bardzo mu*ście* / *byście* pomogli.
very him$_{cl}$ AUX.PAST.2pl AUX.COND.2sg helped
'You helped / would have helped him a lot.'

The contrast between (2.134) and (2.135) suggests that the relative position of auxiliary and pronominal clitics cannot be accounted for in purely syntactic terms. We attribute the relative grammaticality of (2.134) to non-syntactic factors (the relative prosodical weight of past and conditional auxiliaries) which are not expressed in the syntactically oriented approach presented here.

2.5.8 Summary of CC

In this section we presented a syntactic analysis of CC in Polish. As in many other languages, CC in Polish is possible only in certain environments. In Polish, CC occurs with verbs which subcategorize for a bare infinitival complement and in tense auxiliary constructions. Since CC is acceptable only with certain verbs which occur in these environments, we encode CC lexically: each verb which (dis)allows CC is specified as such in the lexicon.

CC in Polish is always optional and in VCs clitics do not have to climb simultaneously. Due to the flexibility of CC, we re-address the issue of complementation of Polish tense auxiliaries discussed recently in Borsley (1999). Since the behaviour of pronominal clitics turns out to be incompatible with his analysis, we assume a much more relaxed syntactic structure of auxiliary constructions, similarly as for VCs.

In order to account for CC in Polish, we relax some of HPSG syntactic principles. These modifications allow us to capture the optionality of CC and the split of clitics in VCs. The same syntactic mechanisms are employed for all instances of CC. A slightly more restricted behaviour of CC in auxiliary constructions is captured by a lexical constraint whereas general syntactic principles remain unchanged.

Finally, the account of linear positions of clitics proposed for simple sentences in §2.4 is extended to capture order of clitics in VCs but also in tense auxiliary constructions. We provide an LP constraint, which differs from that in Borsley (1999), to account for the order of weak auxiliaries with respect to their participial complements. Also the relative order of pronominal and auxiliary clitics is licensed by a separate LP constraint.

Chapter 3

Reflexive Marker

In this chapter, we deal with the behaviour of the Polish reflexive marker (RM) *się* 'self'.[1] We discuss the RM separately from other nominal clitics since its relationship to the orthotonic (stressed) form *siebie* 'self' is more complex than in the case of weak pronominal forms discussed so far. Unlike weak pronominals, the RM is multifunctional and can occur with a verb in various constructions. Another feature which distinguishes the RM from the clitics discussed in the previous chapter is the so-called haplology of the RM. Roughly speaking, if several RMs are required in a clause, the meaning of the sentence is preserved if just one RM is present and other RMs are dropped (in (3.1), both *boję się* '(I) fear' and *roześmiać się* 'to laugh' obligatorily require the RM).

(3.1) Boję się głośno roześmiać.
fear-I RM aloud laugh$_{inf}$
'I'm afraid to laugh aloud.'

Cross-linguistically, haplology is often morphologically conditioned and attributed to the fact that affixes cannot be multiplied. The careful examination of properties of the RM in various constructions shows, however, that the RM is a syntactic item rather than an affix. Such an account is strongly supported by prosody, phonology, syntactic ellipsis, the relatively free distribution and the promiscuity of the RM. Once the RM is treated as a syntactic item, we have to provide a syntactic account of haplology. We examine haplology in all environments where the RM can occur, i.e., simple clauses, verb clusters (VC), participial and gerundial phrases, as well as in coordinated structures. We provide an HPSG account of various constructions with the RM and propose a syntactic treatment of haplology.

3.1 Introduction

3.1.1 'Reflexive' Pronouns

The so-called reflexive pronoun (pol. *zaimek zwrotny*) *siebie* 'self' is unspecified for gender, person and number, but it is inflected for case. The pronoun may be used with respect to singular as well as plural NPs, appearing in any person or gender, see (3.2).

(3.2) a. Myślisz tylko o sobie.
think-you$_{sg}$ only about self$_{loc}$

[1] This chapter is by and large based on my two papers Kupść (1999d, 2000b).

pronoun	nom	gen	dat	acc	inst	loc	voc
'self'	—	siebie / *się*	sobie	siebie / *się*	sobą	o sobie	—

Figure 3.1: Inflectional paradigm of the anaphoric pronoun *siebie*

'You think only about yourself.'

b. Kupiliśmy sobie prezenty.
bought-we self$_{dat}$ presents
'We bought presents for ourselves / each other.'

c. Jan narysował siebie.
John drew self$_{acc}$
'John drew himself.'

The pronoun *siebie* may be used as a reflexive or a reciprocal pronoun: the examples in (3.2a,c) have only the reflexive meaning but (3.2b) is ambiguous: either a reflexive or a reciprocal reading can be obtained. In order to reflect this ambiguity, we call *siebie* an anaphoric pronoun.

In the genitive and accusative form, the pronoun can be usually substituted by the weak form *się*, also unspecified for person, gender and number, see (3.3).

(3.3) a. Jan się narysował.
John RM$_{acc}$ drew
'John drew himself.'

b. Nie umyłeś się wczoraj.
NM washed-you$_{sg}$ RM$_{gen}$ yesterday
'You did not wash yesterday.'

As we will see in §3.1.2, this apparent correspondence is often misleading since *się* is multifunctional and cannot be always substituted by *siebie*. For this reason, we call *się* the reflexive marker (henceforth the RM), see the glosses in (3.3).

The pronoun *siebie* does not occur in nominative and vocative cases; its inflectional paradigm is given in fig. 3.1 (the weak form *się* is indicated in *italic*).

3.1.2 The Multifunctional RM

Not unlike its counterparts in Romance or other Slavic languages, the Polish RM *się* is multifunctional and it can occur in a variety of constructions. There seems to be no consensus among linguists with respect to the classification of these constructions in Polish. Criteria of classification and terminology vary from one author to another. For example, Wilczewska (1966) distinguishes 17 different types of constructions with *się*, Saloni (1976) reduces this number to 8, while in Kański (1986) and Kubiński (1987) only 6 constructions are distinguished.

We will have little to say about such a classification since this would lead us too far away from the main topic of this study. Below, we illustrate these functions of the RM (informally specified) which will concern us in this study. Throughout this chapter, whenever the function of the RM is not important, we call the verb with an RM a 'reflexivized verb', irrespectively of its semantic properties.

In its most prominent function, *się* is often treated as a weak form of the *anaphoric* (both reflexive and reciprocal)[2] pronoun *siebie* 'self', see (3.4).[3]

(3.4) a. Janek zobaczył się / siebie w lustrze.
John saw RM$_{refl}$ self$_{refl}$ in mirror
'John saw himself in the mirror.'

b. Dzieci się / siebie straszą opowieściami o duchach.
children RM$_{recp}$ self$_{recp}$ frighten stories about ghosts
'Children frighten each other with stories about ghosts.'

In (3.4), the pronoun *siebie* and the anaphoric RM are in principle interchangeable. As we will see in §3.1.3 and §3.2.7, this is not always the case.

In its other functions, the RM cannot be replaced by *siebie* and the sentence does not have a reflexive meaning (see glosses in (3.5), (3.6), (3.8)). One of the most productive Polish constructions with the RM is what we call the *impersonal construction*, (3.5) (the impersonal reflexive form in Wilczewska (1966), pro-nominative constructions in Kański (1986, 1992)).

(3.5) a. Czyta się / *siebie tę książkę z przyjemnością.
reads$_{3.sg.neut}$ RM$_{imp}$ self this book with pleasure
'One reads this book with pleasure.'

b. Czytało mi się / *siebie tę książkę z przyjemnością.
read$_{3.sg.neut}$ me$_{dat}$ RM$_{imp}$ self this book with pleasure
'I enjoyed reading this book.'

In these constructions, extensively discussed in Dziwirek (1994), the agent either remains unspecified, (3.5a), or it is (optionally) expressed as a dative NP, (3.5b). The verb always occurs in the third person neuter singular form (no agreement between the verb and NP[dat] holds, see (3.5b)), the RM must be present and usually a verbal modifier is required to render the sentence fully grammatical, cf. Kański (1986, 1992).

Another quite productive construction, illustrated in (3.6), is called here the *middle construction* (characterizing verbs in Wilczewska (1966), passive constructions with *się* in Saloni (1976), anti-causative and passive reflexives in Rozwadowska (1992)).

(3.6) a. Książka się / *siebie dobrze sprzedawała.
book$_{sg.fem.nom}$ RM$_{mid}$ self well sold$_{3.sg.fem}$
'The book was selling well.'

b. Szklanka się / *siebie zbiła.
glass$_{sg.fem.nom}$ RM$_{mid}$ self broke$_{3.sg.fem}$
'The glass broke.'

In (3.6), the verb agrees with the nominative subject NP. These constructions resemble passive sentences since the subject NP has patient-like rather than agent-like properties if compared with the corresponding active sentence, see (3.7a). Unlike in passive constructions, the subject of middles is perceived as not fully passive, cf. Kański (1986), but the real agent cannot be overtly expressed, cf. the passive in (3.7b) vs. the middle in (3.7c).

[2]Many linguists discuss separately these two functions of the RM, e.g., Wilczewska (1966), Saloni (1976), Kański (1986), Kubiński (1987).

[3]Functions of the RM are indicated by mnemonic subscrtipts.

(3.7) a. Janek sprzedał książkę.
John sold book
'John sold the book.'

b. Książka została sprzedana przez Janka.
book was sold by John
'The book was sold by John.'

c. * Książka się dobrze sprzedawała przez Janka.
book RM_{mid} well sold by John

Thus, these constructions are neither truly active nor passive and we call them middle (after Kardela (1986), Kański (1986), Kubiński (1987), among others). Since we focus mostly on syntactic properties of (3.6), the semantic difference between inchoatives, (3.6b), which involve change of state, and true middles, (3.6a), will be ignored.

Finally, we consider a large irregular group of verbs which have neither syntactic, (3.8a), nor semantic, (3.8b), equivalents without *się*. The verb *bać* 'fear' must occur with the RM whereas the verb *zadawać* alone means 'to give' and is semantically unrelated to *zadawać się* 'to deal with' in (3.8b). Apart from that, the examples in (3.8) do not differ from usual active sentences.

(3.8) a. Jan bał się / *siebie Piotra.
$\text{John}_{masc.3sg.nom}$ $\text{feared}_{masc.3sg}$ RM_{inh} self Peter
'John was afraid of Peter.'

b. Nie zadawaj się / *siebie z nim.
not deal_{2sg} RM_{inh} self with him
'Stay away from him.'

In the Polish linguistic tradition, such verbs are often called reflexiva tantum. We use the term *inherent reflexives* to be able to refer to the RM in these verbs as well (the inherent RM).

The above presentation of functions of the RM is anything but exhaustive; for the purpose of our discussion, however, such a typology is sufficient. A more detailed characteristics of the above constructions will be provided in §3.3 and §3.4.

3.1.3 The RM as a Deficient Form

As we said above, the anaphoric RM is treated as a weak counterpart of *siebie*, cf. (3.4). Unlike *siebie*, the RM is a prosodically deficient (atonic) item and cannot be stressed, see (3.9).[4]

(3.9) Jan zobaczył SIEBIE / *SIĘ, a nie Piotra.
John saw self_{refl} RM_{refl} and not Peter
'John saw himself and not Peter.'

This is true also of impersonal, middle, and inherent RMs: none of them can bear heavy stress, (3.10).

[4] Heavy stress is marked with UPPER case letters.

(3.10) a. * Dobrze SIĘ czyta tę książkę.
well RM_{imp} reads this book
'One reads this book easily.'

b. * Lód SIĘ stopił.
ice RM_{mid} melt
'The ice melt.'

c. * Jan SIĘ boi.
John RM_{inh} fears
'John is afraid.'

Below, we discuss other properties which distinguish deficient and strong forms and show that the RM in all its functions is a deficient form in the sense of Cardinaletti and Starke (1994), see §2.1.1. Similar data, but for the anaphoric RM only, have been discussed in Kubiński (1982), Lubowicz (2000) and Kupść (2000b). (See §2.1.3 for the analogous discussion of Polish pronominal clitics.)

Peripheral Positions Deficient forms, in contrast to strong forms, cannot occur in peripheral positions or isolation:

(3.11) Kogo widzisz na tym zdjęciu? Siebie. / *Się.
who see-you on this picture self_{strong} $RM_{deficient}$
'Whom do you see in this picture? Myself.'

(3.11) shows that although *siebie* can be substituted by the anaphoric RM in (3.4), only the former can be used alone. As (3.12) show, in its other functions, the RM cannot be used in isolation either.[5]

(3.12) a. Kto pójdzie na dworzec po Kasię? *Się.
who will.go on station for Kathy RM_{imp}
'Who will pick up Kathy from the station? One.' (intended)

b. Co lód stopił? *Się.
what ice melt RM_{mid}
'What has ice melt? Itself.' (putatively)

As only referential NPs can be used in isolation, the ungrammaticality of the answers in (3.12) can be blamed on the fact that the RM in these constructions is not referential. As the examples in (3.13) show, the RM cannot be used in cleft sentences which are considered peripheral positions in Cardinaletti and Starke (1994).

(3.13) a. # To się pójdzie na dworzec po Kasię.
it RM_{imp} will.go on station for Kathy
'It is one who will pick up Kathy from the station.' (intended)

b. # To się lód stopił.
it RM_{mid} ice melt
'It is oneself that ice melt.' (intended)

[5] The relevant examples with the inherent RM are missing here since the inherent RM cannot be substituted by another NP (see (3.8)). Questions which ask for this RM and sentences with a coordinated or modified inherent RM are uninterpretable.

Coordination A deficient pronoun, unlike a strong one, cannot occur as a conjunct in a coordinated structure, cf., Kubiński (1982, 1987), Lubowicz (2000), Kupść (2000b).

(3.14) Jan umył siebie / *się i dzieci.
John washed self$_{strong}$ RM$_{deficient}$ and children
'John washed himself and the children.'

In other constructions, the RM cannot be coordinated either, see (3.15).[6]

(3.15) a. * Tę książkę czytało się i dziewczę z przyjemnością.
this book read RM$_{imp}$ and girl with pleasure
'One and a girl read this book with pleasure.' (putatively)

b. * Szklanka się i inne stłukła.
glass RM$_{mid}$ and others broke
'The glass broke itself and others.' (putatively)

Modification According to Cardinaletti and Starke (1994), deficient forms cannot be modified:

(3.16) Jan tylko siebie / #się umył.
John only self$_{strong}$ RM$_{deficient}$ washed
'John washed only himself.'

In (3.16), the adverb *tylko* 'only' can modify the strong pronoun *siebie* (see the gloss) but not the deficient form *się* ('#' means that the reading is unavailable). If the RM is used, the adverb modifies the verb but not the RM: the sentence means that John only washed (and he did nothing else), see Lubowicz (2000). Also in other constructions the RM cannot be modified, see (3.17) (sentences are interpretable only if the adverb *tylko* 'only' modifies the verb but not the RM).

(3.17) a. # Tę książkę tylko się czytało z przyjemnością.
this book only RM$_{imp}$ read with pleasure
'Only one read this book with pleasure.' (intended)

b. # Szklanka tylko się stłukła.
glass only RM$_{mid}$ broke
'The glass broke (only itself).' (intended)

Witkoś (1996b, 1998) observes that deficient pronouns, unlike strong ones, do not allow for constituent negation. As (3.18) shows, the RM cannot be preceded by the constituent negation *nie* 'not'.[7]

[6]The ungrammaticality of (3.15a) may be also explained by the fact that *dziewczę* 'a girl' is a nominative subject whereas, as we argue in §3.3.2, the RM in impersonal constructions is not a nominative subject. If this perspective is adopted, (3.15a) is inconclusive for the deficient vs. strong status of the RM in impersonal constructions.

[7]The RM and pronominal clitics cannot be preceded by the homophonous preverbal sentential negative marker *nie* either. As we show in Chapter 4, this is due to the fact that the sentential negative marker is a verbal prefix, unlike Polish pronominal clitics and the RM.

(3.18) a. Jan umył nie siebie / *się, lecz dziecko.
John washed not self$_{refl}$ RM$_{refl}$ but child
'John washed not himself but a child.'

b. * Tę książkę nie się czytało z przyjemnością.
this book not RM$_{imp}$ read with pleasure
'Not one enjoyed reading this book.' (putatively)

c. * Lód nie się stopił.
ice not RM$_{mid}$ melt
'Ice melt not itself.' (putatively)

Distribution The distribution of deficient forms is different from that of strong ones. For example, the sentence initial position of the RM is excluded, cf. (3.19), see §3.2.6 for other restrictions.[8]

(3.19) a. Siebie / *się wczoraj Jan umył.
self$_{strong}$ RM$_{deficient}$ yesterday John washed
'John washed himself yesterday.'

b. * Się czytało tę książkę z przyjemnością.
RM$_{imp}$ read this book with pleasure
'One read this book with pleasure.'

c. * Się szklanka stłukła.
RM$_{mid}$ glass broke
'The glass broke.'

d. * Się Jan boi Piotra.
RM$_{inh}$ John fears Peter
'John is afraid of Peter.'

Reference According to Cardinaletti and Starke (1994), strong forms, unlike deficient ones, may have only human reference. This distinction, however, is not fully respected by the Polish *siebie* and the anaphoric RM. In (3.20), both the RM and *siebie* may refer to dogs.

(3.20) Gdy psy$_i$ zobaczyły się$_i$ / siebie$_i$ nawzajem, zaczęły szczekać.
when dogs saw RM$_{deficient}$ self$_{strong}$ each other started$_{3.pl.nhum}$ bark$_{inf}$
'When the dogs saw each other, they started barking.'

Other constructions are not semantically reflexive and we do not consider the RM in (3.5), (3.6), (3.8) to be referential, see §3.3 for discussion. Thus, the reference of the RM in other constructions cannot be tested.

[8]In colloquial speech, *się* is sometimes used in the sentence initial position, see Misz (1966, p. 109), Dłuska (1974, p. 31), Rappaport (1998). Such sentences, however, usually convey a comic effect and we do not consider them grammatical.

Conclusion Since the RM successfully passes most of the tests proposed in Cardinaletti and Starke (1994) to distinguish deficient and strong forms, we conclude that 1) the RM is a deficient form irrespectively of its functions and 2) the presented contrasts between *siebie* and the anaphoric RM follow from the distinction between deficient and strong forms. As we will see in §3.2.7 and §3.5, there are other discrepancies between *siebie* and the anaphoric RM which do not strictly follow from the deficient vs. strong distinction.

3.1.4 The Status of the RM in the Literature

As we have seen in the previous subsection, the RM is a deficient form. Nevertheless, this does not tell us much about its grammatical status, i.e., whether the RM should be treated as an affix or a syntactic item. In the view of haplology of *się*, mentioned in (3.1), the grammatical status of the RM is an important issue. Before proceeding to the detailed discussion of properties of the RM in §3.2, let us briefly review opinions expressed so far in the literature.

The grammatical status of the RM is rarely dealt with explicitly. In most traditional grammar textbooks of Polish, the RM *się* is called the reflexive pronoun and it is coupled with the strong form *siebie*, e.g., Lehr-Spławiński and Kubiński (1952, p. 115), Szober (1953, p. 100), Jaworski (1974, p. 41). This suggests that the RM is treated as a syntactic item by these grammarians. If grammar textbooks ever mention other functions of *się*, the RM is either always called the reflexive pronoun, e.g., (Klemensiewicz, 1960, p. 53), Bąk (1995, p. 308), or the reflexive pronoun *się* (a semantic equivalent of *siebie*) is distinguished from the morph *się*, which is an indispensable part of certain verbs or syntactic constructions, e.g., Nagórko (1996, pp. 140–141), Grzegorczykowa *et al.* (1999, pp. 191–195). The grammatical status of the RM is not openly addressed.

Descriptive accounts often considerably differ in the treatment of the RM. In her monograph of constructions with *się*, Wilczewska (1966) mostly treats the RM as a derivational or inflectional morpheme or (only in two constructions) as a syntactic item. On the contrary, Saloni (1976) always treats the RM as a syntactic item but postulates two different lexemes *się*. Schenker (1985, 1993) calls the RM a particle, without explicitly specifying its grammatical status.

A similar diversity of opinions (or lack thereof) is reflected in formal approaches. In the metamorphosis grammars of Szpakowicz (1986) and Świdziński (1992) the RM is a syntactic item. Also Kański (1986) treats the RM as a syntactic item and provides a uniform semantic (lambda-categorial) analysis of all its functions. In the Arc Pair Grammar of Kubiński (1987), *się* is a syntactic reflexive pronoun. This means that the RM is always treated as a syntactic item (a pronoun) but only a subset of constructions with the RM is interpreted as semantically reflexive, Kubiński (1987, p. 42). In the approach of Dziwirek (1994), couched in the closely related framework of Relational Grammar, the RM is called the reflexive clitic. It is considered "a verbal morpheme (not unlike passive morphology)", Dziwirek (1994, p. 102), which suggests that she considers the RM an affix. Rappaport (1998) treats *się* as a phrasal affix, incorporating the A-Morphous approach of Anderson (1992) into his GB/Minimalist account of clitics. Although Witkoś (1998, p. 1) explicitly excludes *się* from his extensive Minimalist study of Polish clitics, his treatment of the RM might be close to that of Rappaport (1998). Witkoś (1998, p. 205, fn. 31) calls the RM the reflexive particle which "however, can be treated as a mobile part of verbal morphology". In the semantic accounts of Szymańska (1998) and Szymańska and Śpiewak (1998), the grammatical status of the RM is not ultimately specified but they tend to treat the RM as an affix, "a derivational morpheme marking the intransitivity of the verb", Szymańska and Śpiewak (1998, p. 131) (a similar view seems to be expressed also in Rozwadowska (1992)).

As this brief overview indicates, views on the grammatical status of the RM (if any) are quite diversed. In the next section, we examine the properties of the RM, aiming to establish its grammatical status.

3.2 Properties of the RM

Discussing the properties of the RM, we consider two options for *się*: a postlexical clitic (PLC, a syntactic item) and a phrasal affix (PA), adopting the terminology of Miller (1992a) introduced in §2.1.1. We apply a similar set of criteria as we used for establishing the grammatical status of pronominal clitics in Chapter 2. When syntactic vs. morphological properties of the RM are not crucial for the discussion, we just call it a 'clitic'.

It is not uncommon for one morpheme to behave differently in distinct environments, e.g., the French negative marker *ne* has different properties when it occurs with finite and non-finite verbs, see Recourcé (1995). Therefore, in the subsequent sections we provide examples with the RM in all its functions mentioned above.

3.2.1 Phonology

As we have seen in §3.1.3, the RM is a clitic: it is an atonic (unstressed) item, (3.9), (3.10), and always requires a host, (3.11), (3.12). Phonologically, the RM is usually an enclitic, i.e., it takes a preceding item as a phonological host, cf. Misz (1966, p. 103), Dłuska (1974, p. 29), Rappaport (1988, p. 321), Dukiewicz and Sawicka (1995, p. 174).

The enclitic status of the RM is supported by the fact that the RM cannot appear in sentence initial positions, cf. (3.19), but the initial positions in an embedded clause are perfectly grammatical, see (3.21).

(3.21) a. Jan powiedział, że się zatrzyma u znajomych.
John said COMP RM$_{inh}$ will.stop at acquaintances
'John said that he would stay at his friends'.'

b. Jan powiedział, że się szklanka stłukła.
John said COMP RM$_{mid}$ glass broke
'John said that the glass broke.'

c. Jan powiedział, że się dobrze czyta tę książkę.
John said COMP RM$_{imp}$ well reads this book
'John said that one reads this book easily.'

d. Jan powiedział, że się nie umyje.
John said COMP RM$_{refl}$ NM will.wash
'John said that he would not wash.'

There is a phonological word barrier, the inter-word sandhi, between the enclitic *się* and its phonological host, cf. Rappaport (1988, p. 310), Dukiewicz and Sawicka (1995, p. 174). The direct evidence confirming this fact comes from southern and western dialects of Polish. In these dialects, the last consonant of the word *brat* 'brother' in (3.22) is pronounced as [d], which is characteristic of inter-word voicing but not of the word-internal pronunciation (i.e., [t]), see Rappaport (1988, p. 310).

(3.22) Brat się spóźnia.
brother RM_{inh} is late
'The brother is late'.'

As observed in Dłuska (1974, pp. 29–31) and Rappaport (1988, 1998), under certain rhythmic conditions the RM can become a phonological proclitic. Apparently, this is possible in (3.23a), due to the heavy stress on *ten tylko* 'this only', or in (3.23b), where the RM is placed at the left margin of the intonational phrase, cf. Rappaport (1988, p. 321), Rappaport (1998, p. 470).

(3.23) a. ten tylko się dowie
this only RM_{inh} will.know
'Only this one will find out.'

b. W instytucie — się nie zgodzą.
in institute RM_{inh} NM agree-they
'At the institute, they won't agree to that.'

As noted in Dukiewicz and Sawicka (1995, pp. 173–174), the phonological barrier between the RM and its host is stronger than, e.g., a barrier between the negative proclitic *nie* 'not' and the verb. Therefore, *nie* is tighter phonologically bound to the verb than *się*.[9]

We take these facts to indicate that the RM does not form a phonological word with its host and, consequently, cannot be an affix.

3.2.2 Prosody

The RM shares prosodic properties with pronominal clitics (see §2.2.1): if the RM follows a word, the regular penultimate stress of this word is not affected, see (3.24).

(3.24) a. Piotr uBIErał się.
Peter dressed RM_{refl}
'Peter got himself dressed.'

b. Dobrze CZYta się tę książkę.
well reads RM_{imp} this book
'One reads this book very easily.'

c. Jak sprzedaWAła się ta książka?
how sold RM_{mid} this book
'How was this book selling out?'

d. Nie zaDAwaj się z nim.
NM deal RM_{inh} with him
'Stay away from him.'

Also if the RM succeeds a word irregularly stressed on the antepenultimate syllable, e.g., *graMAtyka* 'grammar' in (3.25), the lexical stress of this word remains intact. (For more examples of such irregularly stressed words see Rubach and Booij (1985).)

[9] In Chapter 4, we argue that the negative proclitic *nie* is a verbal prefix.

(3.25) a. Nie wszyscy graMAtycy się lubią.
not all grammarians RM$_{recp}$ like
'Not all grammarians like each other.'

b. GraMAtykom się z trudnością czytało tę pracę.
grammarians RM$_{imp}$ with difficulty read this paper
'Grammarians didn't find that paper easy to read.'

c. Ta graMAtyka się wydaje zbyt skomplikowana.
this grammar RM$_{inh}$ seems too complex
'This grammar seems to be too complex.'

d. Ta graMAtyka się nie przyjęła.
this grammar RM$_{mid}$ NM accepted
'This grammar was not accepted.'

Note also that, even if the RM becomes a proclitic (see (3.23)), it must remain unstressed if attached to a monosyllabic verb, see (3.26a). This behaviour of *się* should be contrasted with the verbal proclitic *nie* 'not', which is obligatorily stressed if it precedes a monosyllabic verb, (3.26b), and they form a prosodic word (see Chapter 4 for details).

(3.26) a. Tam, jak mówił, się ŚPI / *SIĘ śpi.
there as said-he RM$_{imp}$ sleeps RM$_{imp}$ sleeps
'There, as he said, one sleeps.'

b. On jeszcze NIE śpi / *nie ŚPI.
he yet NM sleeps NM sleeps
'He is not sleeping yet.'

As the penultimate and antipenultimate stress are the only productive lexical stress patterns in Polish (see Dukiewicz and Sawicka (1995)), we conclude that the RM does not form a prosodic word with its phonological host.

3.2.3 Coordination

We apply Miller's (1992a) coordination test to the RM in order to recognize its postlexical clitic (PLC) vs. phrasal affix (PA) properties.

The criteria of the test, applied also to pronominal clitics and quoted in §2.2.6, say that 1) if a prosodically deficient item has to be obligatorily repeated on each conjunct then the item has to be an affix (phrasal or inflectional); 2) if such an item must take wide scope over coordination and cannot occur in all conjuncts, it is necessarily a PLC. This test, however, cannot distinguish between a PA and a PLC if the item appears in each conjunct only optionally, Miller (1992b). The last case holds for the RM: the repetition of *się* is optional in coordinated phrases, cf. (3.27) vs. (3.28).

The RM can have a wide scope over coordination of hosts (verbs in both conjuncts require the RM).

(3.27) a. Jan umył się i ogolił.
John washed RM$_{refl}$ and shaved
'John washed (himself) and shaved (himself) quickly.'

b. Siedziało się i gadało do późna.
sat RM$_{imp}$ and chatted till late

'One was sitting and chatting till late.'

c. Dobrze pierze się i prasuje ten materiał.
well washes RM_{mid} and irons this cloth
'This cloth washes and irons well.'

d. Jan potykał się i ślizgał na zamarzniętej drodze.
John stumbled RM_{inh} and slid on frozen road
'John was stumbling and sliding on the frozen road.'

The RM can be repeated in coordination as well.

(3.28) a. Jan umył się i ogolił się.
John washed RM_{refl} and shaved RM_{refl}
'John washed (himself) and quickly shaved (himself).'

b. Siedziało się i gadało się do późna.
sat RM_{imp} and chatted RM_{imp} till late
'One was sitting and chatting till late.'

c. Dobrze pierze się i prasuje się ten materiał.
well washes RM_{mid} and irons RM_{mid} this cloth
'This cloth washes and irons well.'

d. Jan potykał się i ślizgał się na zamarzniętej drodze.
John stumbled RM_{inh} and slid RM_{inh} on frozen road
'John was stumbling and sliding on the frozen road.'

Since the repetition of the RM in each conjunct is optional, we conclude that the coordination test is indecisive with respect to the PLC vs. PA status of *się*.

3.2.4 Ellipsis

Data such as (3.29) can shed some light on the grammatical status of the RM.[10]

(3.29) a. Jan się umył, a Piotr ogolił.
John RM_{refl} washed and Peter shaved
'John washed himself and Peter shaved (himself).

b. Janowi dobrze się pisało tę książkę, a Tomkowi ją łatwo czytało.
$John_{dat}$ well RM_{imp} wrote this book and Tom_{dat} it easy read
'John enjoyed writing and Tom reading this book.'

c. Materiał dobrze się pierze, a sukienka łatwo prasuje.
cloth well RM_{mid} washes and dress easily irons
'The cloth washes well and the dress irons easily.'

d. Jan się potknął, a Tomek poślizgnął.
John RM_{inh} stumbled and Tom slid
'John stumbled and Tom slid.'

The reflexive marker behaves like a syntactic element with respect to syntactic ellipsis, i.e., it can be omitted in the sentence just like a complement can, cf. (3.29) and (3.30).

[10] We are grateful to Stefan Dyła for bringing such facts to our attention.

(3.30) Jan pisze książki, a Piotr czyta.
John writes books and Peter reads
'John writes books and Peter reads (books).'

Normally, such an omission is not attributed to (inflectional or derivational) affixes, e.g., Bresnan and Mchombo (1995). However, if examples in (3.29) are treated as a coordination of sentences which share an argument (namely *się*), both PA and PLC analyses can in principle account for the data (but see §3.2.6 for discussion). As discussed above, phrasal affixes and PLC can take wide scope over coordination of phrases, i.e., sentences, in the case of (3.29).

One of the crucial similarities between standard inflectional affixes and phrasal affixes (PA) is that both have a fixed position in their domain, i.e., a word or a phrase, respectively. Note, however, that the position of the RM in (3.29) is not fixed: in (3.29a) and (3.29d), the RM is the second, while in (3.29b–c), it is the third element in the sentence (the syntactic domain). On the other hand, there are no strict constraints on the linear position of PLCs. A more detailed discussion of a PA approach to the RM proposed in Rappaport (1998) is postponed to §3.2.6.

As discussed, e.g., in Saloni (1976, p. 121), Derwojedowa (1998, p. 103) and Kupść (1999d, p. 95), irrespectively of its function, the RM can be easily omitted in examples such as (3.31).

(3.31) a. D: A zajmiesz się nim do tego czasu?
and will-take-care-you RM$_{inh}$ him till this time
A: Zajmę.
will-take-care-I
D: 'Will you take care of him till this time?' A: 'I will.'

b. A: Umyłeś się już? B: Umyłem.
washes-you RM$_{refl}$ already washed-I
A: 'Have you washed already?' B: 'I have.'

c. A: Czy zupa się już ugotowała? B: ?Ugotowała.
Q soup RM$_{mid}$ already cooked cooked
A: 'Has the soup cooked yet?' B: 'It has.'

d. A: Spało się, co? B: ?Spało.
slept RM$_{imp}$ what slept
A: 'One was sleeping, weren't you?' B: 'I was.'

In order to maintain the PA analysis for (3.31), the domain of affixation would have to be extended to the whole discourse which is a very *ad hoc* solution. Note that neither inflectional, (3.32a), nor phrasal, (3.32b), affixes can be omitted in this way.

(3.32) a. * A: Przeczyta-ją mój tekst? B: Przeczyta.
will-read-they my text will-read
A: 'Will they read my text?' B: 'They will.' (intended)

b. * A: 'Is it your mother's book?' B: 'My mother.'

On the other hand, if *się* is a PLC, any theory of syntactic ellipsis will be able to account for (3.29) as well as for (3.31).

Although the coordination test is indecisive for *się*, we take ellipsis data as strong evidence in favour of the syntactic status of the RM.

3.2.5 Promiscuity

As in the case of pronominal clitics (see §2.2.4), promiscuous attachment of the RM usually is not explicitly dealt with. Certain restrictions, discussed further on in this subsection, were observed in Spencer (1991, p. 367) whereas Dukiewicz and Sawicka (1995, p. 173) mention that enclitics can attach to different expressions in the sentence but no examples are given.

As we said in §3.2.1, the RM is a phonological clitic and it requires an element to attach to. The examples considered so far show that the RM can be attached to a verb, cf. (3.27), (3.31a–b,d) (and also (3.4a), (3.5), (3.8), (3.20), (3.23), (3.24), (3.26a)), a noun, cf. (3.25), (3.31c) (and also (3.4b), (3.6)), an adverb, cf. (3.29b–c) but also a complementizer, cf. (3.21). Additionally, the RM can follow an adjective, (3.33), or certain prepositions, cf. (3.34) below, should prosodic or metric conditions be satisfied.

(3.33) a. Radosnemu się nastrojowi poddaliśmy.
joyful RM$_{inh}$ mood surrendered-we
'We surrendered to the joyful mood.'

b. Prawdziwe się problemy dopiero zaczną.
real RM$_{mid}$ problems only will.start
'The real problems are just about to begin.'

c. Długie się godziny przegadało.
long RM$_{imp}$ hours chatted
'One spend long hours chatting.'

d. Najstarsze się dziewczynki polubiły od razu.
oldest RM$_{recp}$ girls liked at once
'The oldest girls liked each other at once.'

(3.34) a. Dziewczyna dookoła się rozejrzała.
girl around RM$_{inh}$ looked
'The girl looked around.'

b. Dookoła się o tym nie mówiło.
around RM$_{imp}$ about this NM spoke
'Nobody was talking about it.' (lit.: 'One was not talking about it around.')

c. Dookoła się wszędzie zamykały okna.
around RM$_{mid}$ everywhere shut windows
'The windows shut everywhere around.'

d. Dookoła się wszyscy znali.
around RM$_{recp}$ everybody knew-they
'Everybody around knew each other.'

Two restrictions on attachment of clitics in Polish, mentioned in §2.2.4, were noticed in Spencer (1991, p. 367): the negative marker *nie* and prepositions do not host clitics in Polish. The latter observation can be falsified by (3.34) if *dookoła* 'around' is treated as a complementless preposition (an intransitive preposition in the terminology of Borsley and Rivero (1994)) rather than an adverb-like element. Similarly to pronominal clitics, however, *się* cannot be used as a complement of a preposition (the pronoun *siebie* must appear instead), (3.35), cf. Kubiński (1982), Lubowicz (2000), Kupść (2000b).

(3.35) Położył kurtkę obok siebie / *się.
put-he jacket next-to self$_{refl}$ RM$_{refl}$
'He put the jacket next to himself.'

Apparently the only counterexample to (3.35) in contemporary Polish is the fixed expression *samo przez się* 'by itself'. In Old Polish, however, weak forms as complements of prepositions were fully correct, e.g., Klemensiewicz *et al.* (1955).

Contrasts analogous to (3.35) hold for other Polish pronominal clitics, see §2.2.4, and are also present in other languages, Cardinaletti and Starke (1994). Therefore, we attribute the impossibility of *się* as a complement of a preposition to the general distinction between deficient and strong forms (§3.1.3). Note that the ungrammaticality of (3.35) cannot be explained by insufficient prosodic support of the preposition to host the clitic since *obok* 'next to' is a prosodic word.

As discussed in Borsley and Rivero (1994), PPs (and NPs) in Polish can be split by phrase external elements, e.g., (3.36a). Most transitive prepositions in Polish are prosodically deficient and they form a prosodic word with the first stressed element of their complement, see Rubach and Booij (1985, p. 315). The RM cannot be inserted in such PPs between the preposition and the first element of its complement but the RM can appear after this 'complex', see (3.36b).

(3.36) a. Janek z wieloma się dziewczynami spotykał.
John with many RM$_{inh}$ girls met
'John was dating many girls.'

b. * Janek z się wieloma dziewczynami spotykał.

Prosodic deficiency of prepositions, however, is only a partial explanation of the impossibility of PP split; as (3.37) shows PPs headed by prosodically full transitive prepositions cannot be split either.

(3.37) ?* Janek obok się Marysi usadowił.
John next RM$_{inh}$ Mary sat
'John sat next to Mary.'

As in §2.2.5, we assume that in PPs headed by transitive prepositions, the preposition forms an unseparable unit with the first element of its complement NP.

The fact that *się* is not hosted by *nie* 'not', e.g., (3.18), partly follows from the RM being a deficient form. As observed in Witkoś (1998, p. 152), the constituent negation cannot precede Polish weak pronominals. The homophonous sentential negative marker *nie* cannot host clitics either, see Borsley and Rivero (1994, p. 411), Borsley (1999, p. 41), Witkoś (1998, p. 232) (for verbal clitics) and Witkoś (1998, p. 156) (for pronominal clitics). Anticipating the conclusion of this section and the discussion in Chapter 4, this behaviour is explained by the fact that, in contrast to Polish pronominal clitics, the sentential negative marker *nie* is a verbal prefix, see also Witkoś (1998, ch. 4) and Kupść and Przepiórkowski (1997).

As observed in Nevis (1985), Zwicky (1987), Miller (1992a, ch. 4.2), among others, promiscuous phonological attachment is not a unique property of PLCs: a PA always has as its domain one and the same kind of syntactic phrase but the phonological host may vary since the PA is not attached to the head of the phrase, see §2.2.4 for some discussion.

We conclude that promiscuous phonological attachment of *się* does not *a priori* preclude it from being a PA or a PLC and this criterion is indecisive.

3.2.6 Placement

The distribution of Polish clitics in general, and the RM in particular, is relatively free, e.g., Misz (1966), Rappaport (1988, 1998), Spencer (1991, ch. 9), Borsley and Rivero (1994), Kupść (1999d,e), and especially Witkoś (1996b, 1998), and differs significantly from other Slavic languages, see, e.g., Halpern (1995) or Franks (1998) for an overview.

Simple Clauses Roughly speaking, the reflexive marker can occur to the left of the verb and if postverbal, it can be separated from the verb only by other clitics, e.g., Misz (1966), Spencer (1991, ch. 9), Rappaport (1998), Kupść (1999d). Various linear positions of *się* are illustrated in (3.38).[11]

(3.38) a. Sytuacja [się] jutro [się] poprawi [się].
situation RM$_{inh}$ tomorrow will.improve
'The situation will improve tomorrow.'

b. Wczoraj [się] Piotrek [się] późno [się] obudził [się].
yesterday RM$_{refl}$ Peter late woke
'Peter woke up yesterday.'

c. Książkę [się] dobrze [się] czytało [się].
book RM$_{imp}$ well read
'One enjoyed reading this book.'

d. Lody [się] prędko [się] roztopiły [się].
ice-cream RM$_{mid}$ quickly melt
'The ice-cream melt quickly.'

Polish clitics avoid postverbal clause final positions if possible but as (3.38) shows, this is a tendency rather than a principle of the grammar, cf. Misz (1966, p. 105), Spencer (1991, p. 368), Rappaport (1998, p. 468).

As (3.39) shows, when *się* follows the verb, the RM can be separated from the verb only by other clitics, Misz (1966, p. 104), Kupść (1999d, p. 98). No stressed syntactic items can occur between the postverbal RM and the verb, unlike in the case of the pronoun *siebie*, see (3.39a), cf. Rappaport (1998, p. 465), Kupść (1999d, p. 96) (for *się*) and Witkoś (1998, p. 156), Kupść (1999e, p. 112) (for pronominal clitics).[12]

(3.39) a. Jan obejrzał w lustrze siebie / [*się] dokładnie.
John watched in mirror self$_{refl}$ RM$_{refl}$ carefully
'John looked carefully at himself in the mirror.'

b. Czytało mi [się] dobrze [*się] książkę.
read me$_{cl}$ RM$_{imp}$ well book

[11] In all sentences below, only one RM can be present and we write the RM in brackets, [*się*], to indicate possible positions of the RM. For example, (3.38a) corresponds to three grammatical sentences with the RM occurring in three different linear positions: either immediately after the noun *sytuacja* 'situation', or the adverb *jutro* 'tomorrow', or the verb *poprawi* 'improves'.

[12] We use the notation [**się*] to indicate that the sentence is ungrammatical if the RM appears in a given linear position. For example, (3.39b) is grammatical if the RM immediately follows the clitic *mi* 'me' and ungrammatical if the RM immediately succeeds *dobrze* 'well'.

'I enjoyed reading the book.'

c. Podobał jej [się] brat [*się] wtedy.
liked her_{cl} RM_{inh} brother then
'She liked the brother at that time.'

d. Stłukła mu [się] szklanka [*się] wczoraj.
broke to-him_{cl} RM_{mid} glass yesterday
'He broke a glass yesterday.'

Unlike the strong pronoun *siebie*, cf. (3.19a), *się* cannot occur in the absolutely sentence initial position, cf. (3.19). As discussed in §3.2.1, this is due to the prosodic deficiency of the RM rather than to a syntactic restriction. On the other hand, the clause boundary constitutes a barrier for the leftward clitic placement. Thus, only (3.21) are grammatical and (3.40) with *się* in the main clause are impossible.

(3.40) a. Jan powiedział [*się], że [się] zatrzyma u znajomych.
John said RM_{inh} COMP will.stop at acquaintances
'John said that he would stay at his friends'.'

b. Jan powiedział [*się], że [się] szklanka stłukła.
John said RM_{mid} COMP glass broke
'John said that the glass broke.'

c. Jan powiedział [*się], że [się] dobrze czyta tę książkę.
John said RM_{imp} COMP well reads this book
'John said that one read this book easily.'

d. Jan powiedział [*się], że [się] nie umyje.
John said RM_{refl} COMP NM will.wash
'John said that he would not wash.'

Gerunds and Participles In Polish, gerunds (deverbal nominals) and active participles can also occur with the RM (as well as with pronominal clitics, see §2.2.5). Safarewiczowa (1954, p. 337) or Rappaport (1998, p. 471) discuss positions of the RM in gerund phrases and show that the RM must immediately follow the gerund, (3.41).[13]

(3.41) a. Autorzy opisują [*się] zachowanie [się] premiera.
authors describe RM_{inh} $behaviour_{gerund}$ RM_{inh} premier
'The authors describe the behaviour of the premier.'

b. I żadnych głupich myśli o [*się] wieszaniu [się] [...].
and no stupid thoughts about RM_{refl} $hanging_{gerund}$ RM_{refl}
'And no stupid ideas about hanging yourself.'

c. Wywołują one stałe [*się] zawężanie [się] wymiarów.
cause they constant RM_{mid} $narrowing_{gerund}$ RM_{mid} dimensions
'They cause constant narrowing of dimensions.'

In phrases headed by participles, *się* may also precede the participle as long as the RM remains within the phrase, (3.42).

[13]Gerunds and participles are not derived from verbs in impersonal constructions and examples with the impersonal RM cannot be provided here.

(3.42) a. Studentki chichotały [*się] co chwila [się] trącając [się]
students giggled RM$_{recp}$ every moment RM$_{recp}$ nudging$_{participle}$ RM$_{recp}$
łokciami.
elbows
'Students giggled nudging each other all the time.'

b. Głośno zaprotestował [*się] całkowicie [się] nie zgadzając [się] na
loudly protested-he RM$_{inh}$ absolutely RM$_{inh}$ NM agreeing$_{participle}$ RM$_{inh}$ for
takie traktowanie.
this treatment
'He protested loudly against such a treatment.'

c. Talerz zrobił dużo hałasu [*się] tłukąc [się].
plate made much noise RM$_{mid}$ breaking$_{participle}$ RM$_{mid}$
'The plate made lots of noise when it broke.'

Verb Clusters In contrast to (3.40), bare infinitive complements do not constitute a barrier for leftward clitic placement. The RM can be realised either in the infinitive complement it originates from or in the main clause, cf. (3.43).[14]

(3.43) a. Zaczniemy [się] bać [się] po wyborach.
will-begin-we RM$_{inh}$ fear$_{inf}$ after elections
'We'll begin to be afraid after the elections.'

b. Drzwi [się] mogły [się] zamknąć [się] od podmuchu wiatru.
door RM$_{mid}$ could close$_{inf}$ from blow of-wind
'The door could close because of the wind blow.'

c. Chcieli [się] poznać [się] lepiej.
wanted-they RM$_{recp}$ meet$_{inf}$ better
'They wanted to better know each other.'

Note that such a flexibility is not possible just in any infinitival environment: the phrase boundary of a bare infinitival adjunct in (3.44) constitutes a barrier for a further leftward placement of the RM.

(3.44) Poszedł [*się] na stację [*się] dokładnie [się] dowiedzieć [się] o pociągi.
went-he RM$_{inh}$ on station precisely inform$_{inf}$ about trains
'He went to the station to get detailed information about the trains.'

The phenomenon illustrated in (3.43) is an instance of clitic climbing (CC) discussed in §2.5. In (3.43), *się* is introduced by the infinitive verb and the RM is either realised locally or 'climbs' to the main clause. Note that climbing is equally possible both for the inherent RM, (3.43a), as well as other types of the RM, (3.43b–c). In this respect, Polish is similar to Italian but differs from French where inherent clitics do not climb to causative verbs, cf. Kayne (1975), Miller and Sag (1997), Abeillé *et al.* (1998).

The position of *się* in (3.43) is consistent with the general principles of clitic placement considered so far: the RM either occurs to the left of the verb it originates from or immediately

[14]The verb in the impersonal construction with *się* is always in the third neuter personal form and, hence, cannot be used as an infinitive complement.

follows this verb. Hence, it could be concluded that no syntactic climbing takes place and the position of *się* is the result of a linear constraint alone. This conclusion, however, is not supported by contrasts such as in (3.45) discussed in Rappaport (1998, p. 466).

(3.45) a. Bać zaczniemy się po wyborach.
fear$_{inf}$ will-begin-we RM$_{inh}$ after elections
'We'll begin to fear after the elections.'

b. * Boimy zacząć się wybory.
fear-we begin$_{inf}$ RM$_{inh}$ elections
'We fear to start the elections.'

(3.46) a. Zamknąć drzwi mogły się od podmuchu wiatru.
close$_{inf}$ door could RM$_{mid}$ from blow of-wind
'The door could have closed because of the wind blow.'

b. Poznać chcieli się lepiej.
meet$_{inf}$ wanted-they RM$_{recp}$ better
'They wanted to better know each other.'

In (3.45a) and (3.46), *się* occurs to the right of the verb from which it originates, i.e., *bać się* 'to fear', *zamknąć się* 'to close' and *poznać się* 'to know each other', respectively. The RM is separated from these verbs by prosodically full words, which is impossible in (3.45b) as well as in simple clauses, see (3.39). Note that it cannot be said that the linear position of the RM is licensed as long as *się* (and other clitics) follows *any* verb in VCs since the ungrammaticality of (3.45b) would remain unexplained. We follow Rappaport (1998) and explain the contrast in (3.45) on syntactic grounds. We assume that in (3.45a) and (3.46) the RM is raised from the verb at which it is introduced (the infinitive) to the main verb; hence, the RM immediately succeeds the verb from which it is realised, as in, say, (3.31a–b,d). On the other hand, in (3.45b) the RM is introduced by the main verb *boimy się* 'we fear' and cannot be realised as an argument of the 'lower' verb. Therefore, the RM follows the verb from which it is realised but they are separated by a non-clitic which does not respect the principle responsible for ordering the RM in general.

Position of the RM in Rappaport (1998) Data similar to (3.38), (3.39a) and (3.40)–(3.46) are discussed in Rappaport (1998) and are correctly predicted by his analysis. He treats *się* as a phrasal affix (PA) adopting the 'A-Morphous' approach to morphology proposed in Anderson (1992). On this approach, a PA is a spell-out of an abstract morphological feature in some phrasal domain. Similarly to inflectional affixes, a PA has a fixed position in its domain, i.e., the phrase. The position of an affix in the domain is characterized by two parameters: the anchor and the order. The anchor of PA can be either an onset, i.e., the first element (the first phonological word or the first constituent) in the domain, the head of the domain, or a coda, i.e., the final element in the domain. The order of a PA with respect to its anchor is either 'pre', i.e., to the left of the anchor, or 'post', i.e., to the right of the anchor.

It should be noted that Rappaport's analysis does not fully respect similarities between inflectional and phrasal affixes advocated in Anderson (1992): there is no fixed combination of parameters which would account for the distribution of the RM. Rappaport (1998) usually assigns the post-onset position to the Polish RM, i.e., after the first element in some verbal projection (the syntactic domain of the RM). Sometimes, however, the RM occurs after the head of the domain and the anchor parameter has to be changed to 'head'. Moreover, the RM occasionally becomes a proclitic and, thus, the order parameter has to be specified as 'pre' rather than 'post'. Therefore, although the distribution of the RM discussed in Rappaport (1998) *can be* expressed via the domain, anchor and order parameters, there is no fixed combination of these parameters that would uniformly capture *all* possible positions of the RM, as would be required of a PA in Anderson's (1992) or Miller's (1992a) terms. More importantly, the specification of RM positions proposed in Rappaport (1998) cannot account for the full range of data. Below we discuss some limitations of the Rappaport's (1998) approach.

First, the fact that the reflexive marker can be separated from the host by other clitics, e.g., (3.39b–d) and (3.47), seems problematic for Rappaport's (1998) analysis.

(3.47) a. Brat mi się podoba.
brother me$_{cl}$ RM$_{inh}$ likes
'I like the brother.'

b. Brat się mi podoba.

c. Wiele mu się dziewcząt podobało.
many him$_{cl}$ RM$_{inh}$ girls liked
'He liked many girls.'

His analysis accounts for the post-onset position of *się*, i.e., (3.47b), but in order to account for (3.47a) the RM should be treated as a verbal proclitic, i.e., if the RM appears in the pre-head position in VP. (3.47c) is even more problematic. In this case, *się* and *mu* appear within the NP *wiele dziewcząt* 'many girls' (similarly to (3.36a)). In order to maintain the analysis, the sequence *wiele mu* would have to be treated as a syntactic constituent or a phonological word. Since *mu* is the complement of the verb whereas *wiele* 'many' is the determiner of the noun *dziewcząt* 'girls', the sequence *wiele mu* does not form a syntactic constituent. As discussed in §2.2.2, there is a phonological word barrier between the pronominal clitic *mu* 'him' and its host. Thus, the sequence *wiele mu* is not a phonological word and the anchor parameter cannot be specified in phonological terms either.

An obvious alternative would be to represent the sequence of clitics in (3.47) via a unique feature spelled out in the post-onset position. In this case, the realisation of the RM would be captured by this general feature. Recall, however, that the order of clitics in Polish is not rigid, cf. (3.47a) vs. (3.47b). It is not clear how such a variation could be accounted for if both clitics are represented via a single feature.

Second, just like other clitics and even full words, Borsley and Rivero (1994, pp. 396–398), Kupść (1999e, p. 115), the RM can split NPs and PPs and occur phrase-internally, cf. (3.33), (3.36a), (3.47c), (3.48)–(3.50).[15]

(3.48) a. Koledze się Piotra przedstawił.
colleague RM$_{inh}$ Peter's introduced
'He introduced himself to Peter's colleague.'

b. [$_{IP'}$ [$_{NP}$ Koledze się Piotra] [$_{IP}$ przedstawił.]]

(3.49) a. O siostrze się Basi dobrze wyrażał.
about sister RM$_{inh}$ Barbara well spoke-he
'He was talking well about Barbara's sister.'

b. [$_{IP'}$ [$_{NP}$ O siostrze się Basi] [$_{IP}$ dobrze wyrażał.]]

[15] As discussed in §3.2.5, the RM cannot be inserted between the preposition and its complement.

(3.50) a. ? Młodej ładnej się dziewczynie przyglądał.
young nice RM_{inh} girl looked-he
'He was looking at a young nice girl.'

b. $[_{IP'}$ $[_{NP}$ Młodej ładnej się dziewczynie] $[_{IP}$ przyglądał.]]

Rappaport (1998) does not consider phrase internal positions of the RM, which might be attributed to his observation made in Rappaport (1988). Rappaport (1988, p. 322) notes the contrast in (3.51) which seems to illustrate that splitting a (sentence-initial) phrase is not equally available for all kinds of clitics (*m* and *by* are auxiliary clitics, see §2.5.5).

(3.51) a. ?* Ten się piosenkarz zgodził.
this RM_{inh} singer agreed
'This singer agreed.'

b. Tego*m* piosenkarza znał.
this AUX.PAST.1sg singer knew
'I knew this singer.'

(3.52) a. ?* Ten *by* piosenkarz zaśpiewał.
this AUX.COND.3sg singer sang
'This singer would sing.'

b. Tego się piosenkarza zaprosi.
this RM_{imp} singer will.invite
'One will invite this singer.'

Although we do not consider (3.51a) to be fully ungrammatical, unlike Rappaport (1988), there is a clear grammaticality difference between (3.51a) and (3.51b). In Rappaport (1988), this difference is attributed to the prosodic weight of the host rather than to a pure distinction between verbal and pronominal clitics. Indeed, such an explanation seems to be on the right track as the mirror examples in (3.52) indicate.

Let us come back to (3.48)–(3.50). The analysis of Rappaport (1998) can predict (3.48)–(3.49) but, as we show below, it does not account for (3.50). In Rappaport (1998), preposing is treated as a recursive adjunction to IP, which produces a clone of IP, an IP', cf. King (1995). The syntactic structure that would be obtained on this basis is indicated in the (b) examples.

As observed in Rubach and Booij (1985), most prepositions in Polish form a phonological and a prosodic word with the first element of their NP complement. Thus, the position of the RM in (3.48)–(3.49) can be specified as after the first phonological word (post-onset-of-IP'), i.e., *koledze* 'a colleague' and *o siostrze* 'about the sister', respectively. In (3.50), however, the sequence of adjectives neither forms a syntactic constituent nor can it be treated as a phonological word. Therefore, the post-onset position is not preserved. We admit that there is a slight grammaticality difference between (3.48)–(3.49) and (3.50); as this difference is not crucial, we do not consider (3.50) ungrammatical.

Furthermore, as mentioned in §3.2.4, Rappaport's (1998) PA analysis cannot fully account for the distribution of the RM in coordinated structures (this approach cannot account for the ellipsis data in (3.31) either). Similarly to preposing in (3.48), a coordinated structure is analysed as a clone of a conjunct in Rappaport (1998). If (3.29) (repeated for convenience below) is treated as a coordination of clauses the entire sentence will be represented as IP', a clone of the conjunct IPs. A schematic structure which would be associated with (3.29) is given below.

(3.29) a. [$_{IP'}$ [$_{IP}$ Jan się umył], a [$_{IP}$ Piotr ogolił]].
John RM$_{refl}$ washed and Peter shaved
'John washed himself and Peter shaved (himself).

b. [$_{IP'}$ [$_{IP}$ Janowi dobrze się pisało tę książkę], a [$_{IP}$ Tomkowi ją łatwo
John$_{dat}$ well RM$_{imp}$ wrote this book and Tom$_{dat}$ it easy
czytało]].
read
'John enjoyed writing and Tom reading this book.'

c. [$_{IP'}$ [$_{IP}$ Materiał dobrze się pierze], a [$_{IP}$ sukienka łatwo prasuje]].
cloth well RM$_{mid}$ washes and dress easily irons
'The cloth washes well and the dress irons easily.'

d. [$_{IP'}$ [$_{IP}$ Jan się potknął], a [$_{IP}$ Tomek pośliznął]].
John RM$_{inh}$ stumbled and Tom slid
'John stumbled and Tom slid.'

In such a setup, the post-onset position of the RM (in IP') is maintained in (3.29a) and (3.29d) but not in (3.29b–c). In the latter case, the sequences *Janowi dobrze* and *materiał dobrze*, respectively, are neither constituents, nor phonological words and the onset position is not preserved. Note that in (3.29) the syntactic domain of the RM has to be an IP' (the entire coordinated structure) since both conjuncts which require the RM are full IPs (clauses), unlike (3.27)–(3.28).

The above discussion shows that if *się* is considered a PA, assigning to it a rigid position within a phrasal domain is not plausible. Instead, the distribution of *się*, even if more restricted than that of full NPs, e.g., (3.39), (3.45b), is compatible with general principles of Polish linear order. Such a relatively free placement is more characteristic of syntactic items rather than of affixes and argues in favour of a PLC status of *się*.

3.2.7 Incomplete Paradigms?

As we mentioned in §3.1.2, the anaphoric (reflexive and reciprocal) pronoun *siebie* can be substituted by the RM, cf. (3.4) (repeated below).

(3.4) a. Janek zobaczył się / siebie w lustrze.
John saw RM$_{refl}$ self$_{refl}$ in mirror
'John saw himself in the mirror.'

b. Dzieci się / siebie straszą opowieściami o duchach.
children RM$_{recp}$ self$_{recp}$ frighten stories about ghosts
'Children frighten each other with stories about ghosts.'

As observed in §3.1.3, however, the RM is a deficient form and the distribution of *siebie* / *się* is different. Below, we show that the behaviour of the RM is even more idiosyncratic since there are verbs which allow only some of their forms to occur with a RM.

The complement *siebie* cannot be always substituted by the clitic *się*, (3.53), which seems to have been first noticed in Kupść (1999d, p. 100).[16]

[16] This observation is due to Adam Przepiórkowski (p.c.) and has been reported also in Kupść (2000b). Also Saloni (1976, p. 108, fn. 10) remarks on the impossibility or rarity of the RM with the verb *nienawidzić* 'to

(3.53) a. Jan siebie / ?*się lubi / rozumie / kocha / podziwia.
John self$_{refl/*recp}$ RM$_{?*refl/*recp}$ likes understands loves admires
'John likes / understands / loves / admires himself.'

b. Oni siebie / się lubią / rozumieją / kochają / podziwiają.
they self$_{refl/recp}$ RM$_{??refl/recp}$ like understand love admire
'They like / understand / love / admire each other / themselves.'

In (3.53a), the singular verb is markedly worse with *się* than with the strong form *siebie*. On the other hand, the plural verb form can be easily used with *się*, cf. (3.53b) (as well as with *siebie*). Thus, unlike *siebie*, *się* can be used only with plural forms of these verbs. This contrast could be taken to indicate that there are arbitrary gaps in the paradigms of verbs in (3.53), which would strongly support a PA analysis. Morphological formatives (phrasal and inflectional affixes) may exhibit such idiosyncratic behaviour while this is not typical for syntactic items (PLCs), cf. Zwicky and Pullum (1983). A closer look at (3.53) shows, however, that these 'gaps in the paradigms' are not arbitrary and can be uniformly predicted.

First, what is excluded (or strongly dispreferred) in (3.53) is just the reflexive interpretation of *się*, both in singular and plural forms, whereas the reciprocal reading of the RM is possible, (3.53b). Since the reciprocal interpretation is available only if several objects are involved, it is possible for both forms to co-occur with the plural verb, see (3.53b), but not if a single individual is present, (3.53a). Note that the reflexive reading is not excluded in general since *siebie* can be used in this sense both with singular and plural verbs, (3.53). Thus, verbs in (3.53) exclude the reflexive reading of *się* and the ban concerns entire paradigms. As we will see in §3.3.1, these facts are consistent with the claim, e.g., in Bogusławski (1977) and Lubowicz (2000), that the reflexive *siebie* and *się* do not have the same semantic function.

Second, the class of verbs which do not allow the reflexive *siebie / się* substitution seems to be semantically established. All verbs in (3.53) are the so-called subject psychological (psych) verbs, see Rozwadowska (1992): the experiencer of the mental state denoted by the verb is expressed as a subject NP. It seems that semantics of experiencer subject psych verbs excludes the reflexive interpretation of *się* (but not *siebie*, see (3.53)). Interestingly, a similar contrast occurs in Serbian/Croatian, see (3.54) (Wayles Brown (p.c.); examples in (3.54) were provided by Larisa Zlatić), and Dutch, see (3.55) from Lidz (2000).

(3.54) a. Ti volis / mrzis (samog) sebe.
you like hate (own) self
'You like / hate yourself.'

b. * Ti se volis / mrzis.
you RM$_{refl}$ like hate

(3.55) a. Max haat zichzelf.
Max hates selfself
'Max hates himself.'

b. * Max haat zich.
Max hates self
'Max hates himself.'

hate' whereas Rozwadowska (1992, p. 48) observes the impossibility of certain psych verbs in 'passive' (our middle) constructions. None of these authors, however, discusses the lack of *siebie / się* substitution with these verbs.

Therefore, the contrast in (3.53) can be taken to follow from semantics and not to be evidence supporting an affix status of the RM. We come back to this issue in §3.4.3.

Before completing this subsection, let us note that in impersonals the RM does not occur with all verb forms either, cf. (3.5) repeated below.

(3.5) a. Czyta się / *siebie tę książkę z przyjemnością.
reads$_{3.sg.neut}$ RM$_{imp}$ self this book with pleasure
'One reads this book with pleasure.'

b. Czytało mi się / *siebie tę książkę z przyjemnością.
read$_{3.sg.neut}$ me$_{dat}$ RM$_{imp}$ self this book with pleasure
'I enjoyed reading this book.'

Since it is systematically just one verb form which appears with the RM, i.e., the third neuter singular verb, we assume that there is no corresponding full paradigm with the impersonal RM and this unique form appears directly in the lexicon, see §3.4.5.

3.2.8 Conclusions

Let us summarize the discussion so far. In §3.1.3, we have seen that the RM is a deficient pronoun. This, however, does not yet determine whether the RM is a syntactic or a morphological item, i.e., PLC or PA. As far as phonology, prosody and ellipsis are concerned, the RM shows clear syntactic (PLC) behaviour. The RM does not have a fixed position in a sentence (the possible syntactic domain), which is not typical for a PA and supports a PLC analysis. The promiscuous attachment and the optional repetition of the RM in coordination are compatible both with a PA and a PLC status of the RM. Finally, the lack of certain verb forms with the RM apparently favours the morphological (PA) analysis of the RM. However, as discussed in §3.2.7, this fact can be explained on semantic grounds without attributing an affix status to the RM.

Taking into account the data presented in the preceding sections and anticipating the final discussion in §3.5, we assume that the RM is a syntactic item (PLC). In §3.5 we show that the haplology of the Polish RM is not morphologically or phonologically conditioned and, thus, does not contradict this conclusion.

3.3 Towards a Uniform Account of the RM

In §3.1.2, we have seen that the RM has numerous functions. This diversity has often led linguists to identify several RMs which occur in specific constructions, see Wilczewska (1966), Saloni (1976) and references in §3.1.4. On the other hand, Schenker (1985), Kański (1986), Kubiński (1987), among others, aim at providing a uniform account of the RM in all its functions. In this section, we build on the latter approaches and focus on common properties of the RM. In particular, we evaluate grammatical functions associated with the RM in Saloni (1976), i.e., the subject and complement status of *się*.

3.3.1 The RM vs. Complement

Many verbs which occur with an accusative, (3.56a–b), but also a genitive complement, (3.56c), may also occur with the RM.

(3.56) a. Dziewczyny sprzedają cukierki / się na ulicy.
girls sell candies$_{acc}$ RM$_{mid}$ on street
'Girls sell candies / themselves in the street.'

b. Jan$_i$ Piotra / się$_{i/*j}$ / siebie$_{i/*j}$ umył.
John Peter$_{acc}$ RM$_{refl}$ self$_{acc}$ washed
'John washed himself / Peter.'

c. Zosia$_i$ broniła Ani / siebie$_{i/*j}$ / się$_{i/*j}$ przed Kasią$_j$.
Sophia defended Anne$_{gen}$ self$_{gen,refl}$ RM$_{refl}$ before Kate
'Sophia defended Anne / herself against Kate.'

In Saloni (1976), the RM in sentences such as (3.56a–b) is treated on a par with the accusative complement of the verb (Saloni (1976, p. 108) explicitly excludes the RM in genitive from the discussion). Note also that the reflexive RM in (3.56b–c) has the same binding properties as the anaphoric complement *siebie* and is bound by the subject. (In Polish, as in Romance, reflexives can be bound only by the subject, see Kardela (1986), Reinders-Machowska (1991), Marciniak (1999) for binding in Polish.)

Saloni (1976) extends the analysis of the RM as NP[acc] also to verbs which occur with the inherent RM. The RM is treated either as a unique form of the accusative complement (similarly to fixed expressions), (3.57a), or substitutes NP[acc] of the corresponding transitive verb simultaneously affecting its meaning, (3.57b).

(3.57) a. Jan boi się Marysi.
John fears RM$_{inh}$ Mary$_{gen}$
'John fears Mary.'

b. Jan zadaje trudne pytania / się z Piotrem.
John gives difficult questions$_{acc}$ RM$_{inh}$ with Peter
'John asks difficult questions.' vs. 'John deals with Peter.'

However, there are reasons not to treat the RM on a par with the syntactic complement in (3.56)–(3.57).

As repeatedly noticed, (cf. Wilczewska (1966), Kański (1986), Rozwadowska (1992), Szymańska and Śpiewak (1998, fn. 15)) there is no full correspondence between the RM and the complement *siebie*. Neither middle nor inherent *się* is ever substituted by the complement *siebie*, cf. (3.6)–(3.8) on pp. 89–90. Moreover, the reverse is not always true, either: the complement *siebie* is not fully interchangeable with the anaphoric *się*, see §3.2.7. Due to such dissimilarities, the RM is sometimes treated as argument reducing operator or it is said to mark intransitivity or detransitivize verbs such as in (3.56)–(3.57) (cf. Wilczewska (1966), Schenker (1985), Kański (1986), Kubiński (1987), Szymańska and Śpiewak (1998), Lubowicz (2000)), analogously to the RM in Romance (cf. Grimshaw (1982), Alsina (1996), Abeillé *et al.* (1998)). The rest of this subsection is devoted to elaborate this hypothesis.

As discussed in Kubiński (1987), many reflexivized verbs are unaccusative or unergative predicates according to the classification of intransitives, e.g., Levin and Rappaport Hovav (1995). Some of the verbs enumerated in Kubiński (1987, pp. 44-45) are inherent reflexives, e.g., *ślizgać się* 'to slide', *ściemniać się* 'to darken', *czerwienić się* 'to blush', *uśmiechać się* 'to smile', *kłócić się* 'to quarrel', *modlić się* 'to pray', but many others have transitive counterparts and can be treated as middles, e.g., *gotować się* 'to boil', *palić się* 'to burn' *otwierać się* 'to open', *topić się* 'to melt', *zamykać się* 'to shut', *łamać się* 'to brake', or semantically reflexive verbs, e.g., *bawić się* 'to enjoy oneself', *zabić się* 'to kill oneself'. The fact that certain reflexivized verbs are intransitive does not by itself prove that reflexivized verbs are always intransitive. Let us have a closer look at verbs with the RM.

Intransitive verbs cannot be passivized and we use this observation to check intransitivity of reflexivized verbs. As shown in fig. 3.1 on p. 88, Polish anaphoric pronouns do not have nominative case and, hence, cannot be used as the nominative subject. Verbs which occur only with the RM cannot be passivized and the intransitivity of middle forms follows immediately. A little more has to be said about reflexivized verbs which additionally take NP complements, e.g., *bać się*+NP[gen] 'to fear' in (3.57a) or *domagać się*+NP[gen] 'to demand'. Although they cannot be passivized by the RM, certain verbs in Polish passivize a genitive complement (cf. Przepiórkowski (1999)) and, in principle, NP[gen] complement of *bać się* 'to fear' or *domagać się* 'to demand' might become a subject in passive.

In Polish only the direct object of an active verb can become the subject in passive. As (3.58) illustrates, the indirect object of the ditransitive verb *dać* 'to give' cannot become the subject (unlike in English, see the gloss).

(3.58) * Marysia była dana kwiaty (przez Janka).
Mary was given flowers by John

'Mary was given flowers by John.' (putatively)

Although it is usually the accusative direct object that is passivized in Polish, as mentioned above, a genitive or an instrumental direct object of certain verbs can be passivized, see Przepiórkowski (1999) for discussion. As the genitive NP complement of the reflexivized verbs *bać się* 'to fear' and *domagać się* 'to demand' is not a direct object of the verb, these verbs cannot be passivized by NP[gen] either, see (3.59).

(3.59) a. * Marysia jest bana (się) przez Jana.
Mary is feared RM_{inh} by John

b. * Sprawiedliwość jest domagana (się) przez Jana.
justice is demanded RM_{inh} by John
'Justice is demanded by John.' (intended)

On the basis of these facts, we subscribe to the view that in examples such as (3.56)–(3.57), *the RM marks intransitivity of a verb.* We do not treat the RM as an operator intransitivizing or detransitivizing the verb since the transitive counterpart of a reflexivized verb does not always exist, cf. inherent reflexives in (3.57).

The complement *siebie* does not have the nominative case, see fig. 3.1, and it cannot become the subject in passive either. Note, however, that semantic functions of the complement *siebie* and the RM *się* are distinct. As pointed out by Adam Przepiórkowski (p.c.), both nominals have different semantic properties, see (3.60).

(3.60) a. ? W trakcie zwiedzania *Mme Tussaud*, Gorbaczow nie rozpoznał siebie (swojej
during visiting Mme Tussaud Gorbachov NM recognized self self's
figury woskowej).
wax figure

'Gorbachov did not recognize himself (his wax figure) when he was visiting *Mme Tussaud.*'

b. * W trakcie zwiedzania *Mme Tussaud*, Gorbaczow nie rozpoznał się (swojej
during visiting Mme Tussaud Gorbachov NM recognize RM_{refl} self's
figury woskowej).
wax figure
'Gorbachov did not recognize himself (his wax figure) when he was visiting *Mme Tussaud.*' (putatively)

In (3.60b), the anaphoric RM cannot be interpreted as refering to the wax figure (only coreference with Gorbachov as a real person is possible). In (3.60a), however, the anaphoric pronoun *siebie* 'self' can be interpreted as refering either to a real person or to its wax figure, see Jackendoff (1992) and Lidz (2000) for similar examples. We do not attempt at providing an explanation of this fact here but we use (3.60) to illustrate different semantic properties of the two nominals.

In Lubowicz (2000), the RM is treated as an argument and θ-role reducing operator on transitive verbs, in contrast to *siebie* which is a semantic argument of a verb. This difference between *siebie* and the RM will allow us to explain the puzzling behaviour of the transitive (experiencer) subject psych verbs with respect to weak and strong reflexives, cf. (3.53) (repeated below for convenience).

(3.53) a. Jan siebie / ?*się lubi / rozumie / kocha / podziwia.
John $\text{self}_{refl/*recp}$ $\text{RM}_{?*refl/*recp}$ likes understands loves admires
'John likes / understands / loves / admires himself.'

b. Oni siebie / się lubią / rozumieją / kochają / podziwiają.
they $\text{self}_{refl/recp}$ $\text{RM}_{??refl/recp}$ like understand love admire
'They like / understand / love / admire each other / themselves.'

As (3.53) shows, the subject psych verbs can only occur with the complement *siebie* whereas a sentence with *się* is markedly worse in the reflexive reading.

According to Wechsler (1995), the specific property of the subject psych verbs is that the entity (the experiencer) expressed as the subject must have the notion (must know) of the entity (the "theme" or experienced) denoted by the object. If the reflexive RM does not correspond to a semantic argument, only the subject has a semantic role whereas the object's semantic role is 'lost'. Therefore, the sentence with the RM is not semantically well-formed since there is no entity the subject *Jan* could have the notion of and the semantic entailment associated with the psych verb is not fulfilled. On the other hand, according to Lubowicz (2000), the complement *siebie* does correspond to a semantic argument and the sentence is both semantically and syntactically well formed.

Note also that in inherently reflexive verbs, e.g., (3.57), or in middles, e.g., (3.6) on p. 89, the RM is not assigned any semantic role, unlike the NP complement of the corresponding transitive verb (if any). Therefore, the semantic difference between the RM and the NP complement holds for these verbs as well.

To sum up, in (3.56)–(3.57): 1) the RM is not an equivalent of the anaphoric complement *siebie*, 2) the RM marks intransitivity of the verb and 3) the RM is not assigned a semantic role by the verb, in contrast to *siebie*.

3.3.2 The RM vs. Subject

One function of the RM does not conform to the view presented in the previous subsection. In impersonal constructions (IC), cf. (3.61), unlike in (3.56)–(3.57), the RM may occur with a verb irrespectively of its transitivity. In (3.61a), the RM appears with an intransitive verb whereas in (3.5a) (repeated as (3.61b)), the accusative complement of the verb is preserved despite the presence of the RM.

(3.61) a. Żyje się / Jan szczęśliwie.
lives RM_{imp} John_{nom} happily
'One / John lives happily.'

b. Czyta się / Jan tę książkę z przyjemnością.
reads RM_{imp} John_{nom} this_{acc} book_{acc} with pleasure
'One / John reads this book with pleasure.'

Thus, the RM is not interchangeable with a complement in IC. On the other hand, the RM seems to substitute the nominative subject as the ungrammaticality of (3.62) shows.

(3.62) a. * Jan żyje się szczęśliwie.
John_{nom} lives RM_{imp} happily

b. * Jan czyta się tę książkę z przyjemnością.
John_{nom} reads RM_{imp} this_{acc} book_{acc} with pleasure

The mutual exclusiveness of the RM and NP[nom] in (3.62) has been taken, e.g., in Saloni (1976), to indicate that the impersonal RM is an item complementary to the defective nominative-less paradigm of the anaphoric pronoun, see fig. 3.1 on p. 88, and it is the nominative subject of impersonal constructions. On such an account, the RM has an unspecified human reference and is equivalent to the English 'one' (see the gloss in (3.61)), as proposed also for Romance, e.g., Burzio (1986), Dobrovie-Sorin (1998), Monachesi (1999). However, as admitted in Saloni (1976), this analysis does not account for morphologically parallel constructions where the agent is specified and overtly expressed as an NP[dat], e.g., (3.5b).

(3.5b) Czytało mi się tę książkę z przyjemnością.
reads me_{dat} RM_{imp} this_{acc} book_{acc} with pleasure
'I enjoy reading this book.'

The grammatical function of the RM in IC has been a topic of discussion, e.g., in Wilczewska (1966), Saloni (1976), Schenker (1985), Kański (1986), Kopcińska (1997), Rivero (1999). The contribution of the next two paragraphs is to re-examine some of the data which will shed new light on the function of the RM and NP[dat] in IC. First, we present a few arguments against the treatment of the RM as the subject in IC and then, we show that if NP[dat] is present, as in (3.5b), it is not the syntactic subject either. We argue that ICs do not have the overt syntactic subject (as proposed also in Schenker (1985)) but they have a covert (syntactically unrealised) subject.

Dativeless IC As noted in Saloni (1976, p. 106), Kański (1986, p. 182), IC are possible only with verbs which otherwise require the nominative subject; the verb in (3.63a) does not have a nominative subject and (3.63b) is ungrammatical. This, however, is a necessary but not a sufficient condition in Polish.

(3.63) a. Trzeba spać.
ought sleep$_{inf}$
'It's time to sleep.'

b. * Jan / się trzeba spać.
John$_{nom}$ RM$_{imp}$ ought sleep$_{inf}$
'It's time for John / one to sleep.' (putatively)

As observed in Rozwadowska (1992, p. 57), there are certain restrictions on ICs with experiencer object psych verbs such as *fascynować* 'to fascinate', (3.64a), or *interesować* 'to interest', (3.64b) (as indicated in (3.64), these verbs require a NP[nom] subject).

(3.64) a. Janek fascynuje profesorów swoją wiedzą.
John$_{nom}$ fascinates professors self's knowledge
'John fascinates professors with his knowledge.'

b. Janek interesuje ludzi, gdy opowiada o dalekich podróżach.
John$_{nom}$ interests people when talks about distant journeys
'John interests people when he tells about his distant journeys.'

(3.65) a. ?* Fascynuje się profesorów swoją wiedzą.
fascinates RM$_{imp}$ professors self's knowledge
'* One fascinates professors with one's knowledge.'

b. ?* Interesuje się ludzi, gdy jest się przybyszem.
interests RM$_{imp}$ people when is RM$_{imp}$ newcomer
'* One interests people when one is a newcomer.'

Such differences, ignored in the NP[nom] analysis of the impersonal *się* in Rivero (1999), remain puzzling if the RM in (3.65) is treated as a referential human NP[nom] subject similar to *Janek* 'John' in (3.64).

Some light on the subjecthood of IC can be shed by the behaviour of predicative adjectives. In Polish, as in Italian and Spanish, ICs are possible with many predicates, e.g., the copula, (3.66a), passive verbs, (3.66a), transitives, (3.61b), etc.[17]

(3.66) a. Było się miłym / *miły.
was$_{3.sg.neut}$ RM$_{imp}$ nice$_{inst}$ nice$_{nom}$
'One was nice.'

b. Często jest się zapraszanym / *zapraszany przez przyjaciół.
often is$_{3.sg.neut}$ RM$_{imp}$ invited$_{inst}$ invited$_{nom}$ by friends
'One is often invited by friends.'

On the basis of examples similar to (3.66), the Italian RM *si* has been considered to be the nominative subject (cf. Burzio (1986), and most recently Dobrovie-Sorin (1998)). In Polish, however, the morphology of the predicative adjective, *miłym* 'nice$_{inst}$' in (3.66a), or the participle, *zapraszanym* 'invited$_{inst}$' in (3.66b), distinguishes such examples from usual nominative subject clauses. In (3.66), the predicative adjective or participle is obligatorily marked with the instrumental case, unlike in (3.67) where the nominative subject is used.

[17] The use of IC with the copula has been reported in Kański (1986, p. 33) and Rivero (1999), whereas Rozwadowska (1992, p. 56) and Rivero (1999) remark on IC with passive forms. None of these uses is observed in the study of IC in Dziwirek (1994).

(3.67) Janek jest miły / ?*miłym.
John$_{nom}$ is nice$_{nom}$ nice$_{inst}$
'John is nice.'

In spite of the contrasts between (3.66) and (3.67), Rivero (1999) maintains the analysis of the RM as the nominative subject in IC. She takes the instrumental case marking on the adjective to indicate the deficiency of the subject.[18] In ICs the RM is supposed to be the human NP[nom] subject, arbitrary in reference and void of person, gender and number specification. Nevertheless, this analysis leaves certain facts unexplained.

First, such an approach is at odds with the general property of the nominative subject in Polish which agrees with the verb in number, person and gender, see Buttler *et al.* (1971). In ICs, the verb appears in the third person singular neuter form, see (3.66). If the RM is a nominative NP, it should agree with the verb, i.e., the RM should be a third person singular neuter NP, contrary to the conclusion in Rivero (1999, p. 9) that the RM "lacks phi-features for Gender, Number, Person".

Second, Rivero's (1999) approach cannot fully capture the behaviour of ICs with control (modal) verbs. Rivero (1999) observes that if modal verbs are used in ICs, (3.68), there is a similar contrast in case assignment as between (3.66) and (3.67).

(3.68) a. Powinno się być miłym / *miły.
should RM$_{imp}$ be$_{inf}$ nice$_{inst}$ nice$_{nom}$
'One should be nice.'

b. Jan powinien być miły / ?*miłym.
John$_{nom}$ should be$_{inf}$ nice$_{nom}$ nice$_{inst}$
'John should be nice.'

Again, the instrumental case on the adjective in (3.68a) is supposed to show the lack of person, gender and number specification of the nominative RM subject, in contrast to the nominative adjective in (3.68b) which signals the presence of the fully-fledged NP[nom] subject *Jan*.

As observed in Przepiórkowski (1999, pp. 231ff), the structural distance between the overt subject and the predicative adjective may relax the case marking on predicative adjectives. Although in (3.68b) the instrumental case of the adjective is excluded, if the adjective is separated from the nominative subject *Jan* by several verbs, as in (3.69b), the instrumental is possible as well. On the other hand, IC behave differently: irrespectively of the distance, only the instrumental can be used, cf. (3.68a) and (3.69a). This contrast remains unaccounted for if the RM is the nominative subject of IC as postulated in Rivero (1999).

(3.69) a. Powinno się chcieć próbować być szczęśliwym / *szczęśliwy.
should RM$_{imp}$ want$_{inf}$ try$_{inf}$ be$_{inf}$ happy$_{inst}$ happy$_{nom}$
'One should want to try to be happy.'

b. Jan powinien chcieć próbować być szczęśliwy / szczęśliwym.
John$_{nom}$ should want$_{inf}$ try$_{inf}$ be$_{inf}$ happy$_{nom}$ happy$_{inst}$
'John should want to try to be happy.'

[18] Rivero (1999) further motivates this argumentation by the analogous behaviour of predicative adjectives in infinitival and gerundial phrases, e.g., (3.74). We argue below that these apparent similarities are misleading.

The case (non-)agreement of predicative phrases in IC shows that the RM is not the accusative subject either. If the RM were an NP[acc]-subject (on a par with numeral phrases in Polish, see Przepiórkowski (1999)), the nominative subject-verb agreement is not expected. As discussed in Czuba and Przepiórkowski (1995), Przepiórkowski (1999), if the subject is not nominative, the verb always occurs in the third person singular neuter form. Recall that this is precisely the verb form used in IC. However, unlike in IC, the predicative adjective in a NP[acc]-subject clause appears in the accusative case (agrees with the subject), (3.70a), and the instrumental is possible only if the overt subject is structurally distant, (3.70b), in contrast to IC, cf. (3.66) and (3.69a).

(3.70) a. Kilku facetów było szczęśliwych / ?*szczęśliwymi.
a few guys$_{acc}$ was happy$_{acc}$ happy$_{inst}$
'A few guys were nice.'

b. Kilku facetów powinno chcieć próbować być ?szczęśliwych / szczęśliwymi.
a few guys$_{acc}$ should want$_{inf}$ try$_{inf}$ be$_{inf}$ happy$_{acc}$ happy$_{inst}$
'A few guys should want to try to be happy.'

Rivero (1999) uses binding and control data as arguments supporting the presence of an overt subject in IC. As (3.71) shows, the anaphoric *sobie* 'self' is interpretable and refers to (is bound by) an unspecified agent.[19] Also certain adjuncts, e.g., *po pijanemu* 'while drunk' can be controlled only by the subject, see Dziwirek (1994). The expression *po pijanemu* can occur in IC and it is controlled by some unspecified agent in (3.72) (see the gloss). According to Rivero (1999), examples such as (3.71)–(3.72) provide evidence that the RM is a syntactic exponent of this unspecified agent.

(3.71) Trudno się opowiada o sobie$_i$.
difficult RM$_{imp}$ tells about self
'It is difficult to talk about oneself.'

(3.72) Trudno prowadzi się samochód po pijanemu.
difficult drives RM$_{imp}$ car while drunk
'It is difficult to drive a car when one is drunk.'

Note, however, that in Polish the binder or the controller do not have to be overtly expressed. The infinitives or gerunds in (3.73) do not have the overt subject but, nevertheless, binding and control are possible.

(3.73) a. Mówić / mówienie o sobie było jej ulubionym zajęciem.
talk$_{inf}$ talking$_{gerund}$ about self was her favourite activity
'To talk / talking about herself was her favourite activity.'

b. Prowadzić samochód / prowadzenie samochodu po pijanemu było jej ulubioną rozrywką.
drive$_{inf}$ car driving$_{gerund}$ car while drunk was her favourite entertainment
'To drive / driving a car when she was drunk was her favourite entertainment.'

[19] Anaphora in Polish can be bound only by the subject, see (3.56a) and references on p. 111.

Therefore, we take (3.71) and (3.72) to indicate that IC must involve some kind of a subject but they are indecisive whether this subject is overt or covert. If, as suggested by our prior discussion, the RM does not have a syntactic function of the subject in ICs, what is the subject in (3.71) and (3.72)?

Let us come back to case assignment on predicative adjectives in ICs, cf. (3.66). Interestingly, predicative adjectives and participles in ICs behave on a par with predicative complements of infinitives or gerunds in (3.74).

(3.74) a. Być / bycie miłym / *miły ma sens.
be$_{inf}$ being$_{gerund}$ nice$_{inst}$ nice$_{nom}$ has sense
'To be / being nice makes sense.'

b. Być / bycie zapraszanym / *zapraszany sprawiało mu satysfakcję.
be$_{inf}$ being$_{gerund}$ invited$_{inst}$ invited$_{nom}$ made him satisfaction
'To be / being invited gave him satisfaction.'

(3.71)–(3.72) vs. (3.73) as well as (3.66) vs. (3.74) show that (dativeless) IC behave analogously to infinitive and gerund phrases (verb forms which have the syntactically unrealised subject (PRO)): a predicative adjective can only occur in instrumental.

Taking into account these similarities and given that the binder or controller do not have to be overtly expressed, cf. (3.73), we conclude that (dativeless) ICs have the syntactically unrealised subject (PRO), on a par with infinitives and gerunds (also Kardela (1986) treats PRO as the subject of ICs). Such an approach explains the discrepancies between ICs and usual nominative subject clauses, simultaneously capturing binding and control facts. If ICs do not have an overt syntactic subject, the presence of NP[nom] is correctly blocked, (3.62), and subject-verb agreement is not expected.

Finally, let us discuss briefly the behaviour of experiencer object psych verbs in ICs, see (3.64) and (3.65) (repeated below).

(3.64) a. Janek fascynuje profesorów swoją wiedzą.
John fascinates professors self's knowledge
'John fascinates professors with his knowledge.'

b. Janek interesuje ludzi, gdy opowiada o dalekich podróżach.
John interests people when talks about distant journeys
'John interests people when he tells about his distant journeys.'

(3.65) a. ?* Fascynuje się profesorów swoją wiedzą.
fascinates RM$_{imp}$ professors self's knowledge
'* One fascinates professors with one's knowledge.'

b. ?* Interesuje się ludzi, gdy jest się przybyszem.
interests RM$_{imp}$ people when is RM$_{imp}$ newcomer
'* One interests people when one is a newcomer.'

As observed in Rozwadowska (1992, p. 57), only non-agentive (causative) readings of experiencer object psych verbs as in (3.64) are fully acceptable in Polish; the agentive reading is strongly marked as confirmed by the ungrammaticality of (3.75) if the adverbial of purpose *celowo* 'in purpose' is used.[20]

[20] We are deeply indebted to Adam Przepiórkowski for the discussion and clarification of the data.

(3.75) a. * Janek celowo fascynuje profesorów swoją wiedzą.
John purposefully fascinates professors self's knowledge
'John in purpose fascinates professors with his knowledge.'

b. * Janek celowo interesuje ludzi, gdy opowiada o dalekich podróżach.
John purposefully interests people when talks about distant journeys
'John in purpose interests people when he tells about his distant journeys.'

In this respect Polish experiencer object psych verbs in (3.75) differ from their counterparts in other languages which are usually semantically ambiguous between agentive and non-agentive readings, see Grimshaw (1990), Wechsler (1995), Koenig and Davis (1999), Kordoni (1999). As noted in Rozwadowska (1992, p. 58), "it is impossible to form impersonal reflexives from sentences taking human subjects if that subject does not actively participate in the action but is a mere cause of emotion." Since experiencer object psych verbs are inherently non-agentive in Polish, see (3.75), their semantic specification is incompatible with that of impersonal constructions and, hence, (3.65) is semantically ill-formed.

Dative IC The agent in ICs can be specified and overtly expressed as NP[dat], (3.5b), which apparently sets Polish ICs apart from their counterparts in other languages, see Rivero (1999).[21]

(3.5b) Czyta mi się tę książkę z przyjemnością.
reads$_{3.sg.neut}$ me$_{dat}$ RM$_{imp}$ this$_{acc}$ book$_{acc}$ with pleasure
'I enjoy reading this book.'

Since the agent is overtly expressed as a dative NP (unlike in dativeless ICs), we do not consider the RM to be the subject here. Let us have a closer look at such constructions and examine properties of NP[dat].

As (3.76)–(3.77) show, NP[dat] serves as the binder of the anaphora *siebie* and the controller of the adjunct *po pijanemu* 'while drunk'.

(3.76) Jankowi$_i$ trudno się opowiada o sobie$_i$.
John$_{dat}$ difficult RM$_{imp}$ tells about self
'It is difficult for John to talk about himself.'

(3.77) Jankowi trudno prowadzi się samochód po pijanemu.
John$_{dat}$ difficult drives RM$_{imp}$ car while drunk
'It is difficult for John to drive a car while he is drunk.'

Moreover, the predicative adjective agrees with NP[dat] with respect to the case value rather than occurring in the instrumental case, cf. (3.78).

[21]In apparently similar constructions in Bulgarian, the verb occurs in plural and agrees with the plural NP 'books'; the dative NP "is not interpreted as Agent but Experiencer", in contrast to Polish (3.5b). Also in Spanish the dative NP is "interpreted as Ethical Dative / Dative of Interest not as the Agent" (as in Polish), see Rivero (1999, pp. 24ff) for more details.

(3.78) Jankowi dobrze się szło wypoczętemu / ?*wypoczętym / *wypoczęty.
John$_{dat}$ well RM$_{imp}$ walked relaxed$_{dat}$ relaxed$_{inst}$ relaxed$_{nom}$
'John enjoyed walking after some rest.'

Thus, it could be postulated that if NP[dat] is present, it is an overt syntactic subject of IC. Properties of NP[dat], however, distinguish it from overt subjects in Polish, cf. Dziwirek (1994, pp. 70–74).

First, Polish *co*-relative clauses require a resumptive pronoun if a complement is relativized, (3.79a); the presence of a resumptive pronoun, however, is unacceptable if the subject is relativized, see (3.79b). NP[dat] in IC behaves on a par with complements rather than subjects in *co*-relative clauses: the resumptive pronoun must be present, (3.79c).

(3.79) a. Książka$_i$, co *(ją$_i$) Marysia dała Jankowi...
book co$_{rel}$ her Mary gave John
'The book that Mary gave to John...'

b. Ten chłopiec$_i$, co (*on$_i$) tańczył z Ewą, chodzi teraz z Basią.
this boy co$_{rel}$ he danced with Eve walks now with Barbara
'The boy who danced with Eve now dates Barbara.'

c. Ten chłopiec$_i$, co dobrze *(mu$_i$) się tańczyło z Ewą, chodzi teraz z Basią.
this boy co$_{rel}$ well him$_{dat}$ RM$_{imp}$ danced with Eve walks now with Barbara
'The boy who enjoyed dancing with Eve now dates Barbara.'

Second, if subjects of conjunct clauses in a coordinated structure are coreferential, the subject of the second clause must be omitted, see (3.80a); complements coreferential with the subject of the first conjunct, however, cannot be omitted, (3.80b). The behaviour of NP[dat] in IC is analogous to that of complements: NP[dat] coreferential with the subject of the first conjunct cannot be omitted, (3.80c).

(3.80) a. Janek$_i$ chciał wstać o piątej, ale (*on$_i$) zaspał.
John wanted get$_{inf}$ up at five but he overslept
'John wanted to get up at five but he overslept.'

b. Janek$_i$ chciał wstać o piątej, ale Ewa *(mu$_i$) zabroniła.
John wanted get$_{inf}$ up at five but Eve him$_{dat}$ forbade
'John wanted to get up at five but Eve forbade him.'

c. Janek$_i$ chciał wstać o piątej, ale za dobrze *(mu$_i$) się spało.
John wanted get$_{inf}$ up at five but too well him$_{dat}$ RM$_{imp}$ slept
'John wanted to get up at five but he enjoyed sleeping too much.'

Finally, treating NP[dat] as an overt subject would contradict the generalization that the subject in Polish is always assigned a structural case since dative is a lexical case according to Przepiórkowski (1999).

We take these differences to indicate that NP[dat] does not have a syntactic function of the subject in IC but it is, nevertheless, a covert (the so-called logical, e.g., Holvoet (1991)) subject. Thus, NP[dat] is a legitimate binder and controller although its syntactic behaviour and that of overt subjects are distinct.

3.3.3 Conclusions

Data discussed in this section indicate that none of RMs take over the function of a syntactic argument in the corresponding non-reflexivized verb (if any). In §3.3.1, we showed that the RM is not a complement of the verb whereas in §3.3.2 we argued that the RM is not the subject of IC. Therefore, we do not see motivation for having the two distinct lexemes RM postulated in Saloni (1976): a nominative RM (for the subject of ICs) and an accusative RM (for a complement in other constructions). Instead, we assume that the common syntactic function of the RM is to restrict the valence of the verb: the RM either marks the absence of the direct object, cf. §3.3.1, or the absence of the syntactic subject, cf. §3.3.2 (see Schenker (1985) for a similar conclusion).

The semantic function of the RM is different from *siebie*. Unlike *siebie*, the RM is not assigned a semantic role and does not correspond to a semantic argument of the verb as evidenced by the behaviour of experiencer subject psych verbs. As discussed, e.g., in Schenker (1985), Szymańska (1998), the common semantic function of the RM is to modify the subject's involvement in the action or state denoted by the verb. In the case of IC, the agent is put in the background (the agent remains unspecified in dativeless ICs) whereas the subject is 'focused on' in other constructions with the RM.

3.4 HPSG Analysis

In this section, we provide an HPSG analysis of the reflexivized verbs considered above. First, we present a description of the RM and then we proceed to the analysis of verbs which occur with the RM. As the RM is a syntactic item (§3.2), its realisation is syntactic. Since the presence of the RM is related to the lexical semantics of verbs (§3.2.7), this fact is reflected in our analysis as well.

3.4.1 Representation of the Reflexive Marker

In §3.3 we discussed common syntactic and semantic properties underlying all RMs. We assume just one lexical entry for *się* whereas distinct functions of the RM, §3.1.2, will follow from the lexical semantics of reflexivized verbs rather than from the characteristics of the RM itself.

On the basis of the facts presented in §3.2, we treat *się* as a syntactic item. The lexical entry for the RM is given in (3.81).

(3.81)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{się} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \textit{clitic} & \\ \text{LOC} & \begin{bmatrix} \textit{local} & \\ \text{CAT}|\text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{CASE} & \textit{acc-gen} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{ana} & \\ \text{INDEX} & \textit{index} \\ \text{RESTR} & \{\} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

As discussed in §3.1.3 and §3.2, the RM is a postlexical clitic (a deficient form) and it should be distinguished from other NPs. We adopt the classification of *synsem* objects introduced in

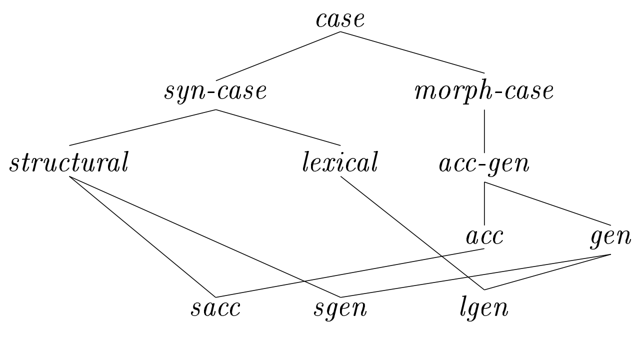
Chapter 2 and represent the RM as an NP whose SYNSEM component is an object of the sort *clitic*.

The anaphoric *się* has the same binding properties as other Polish anaphoric pronouns, see (3.56a) in §3.3.1, but if RM is used in other functions, the verb is not semantically reflexive, see §3.1.2, (3.56b), (3.57), (3.71a). This distinction is captured if we assume that the CONTENT (CONT) value of the RM is always an object of sort *anaphoric* (*ana*, a subsort of *nom_obj*, see §2.3.2) but its INDEX value is left unspecified in the lexical entry: the RM can be either a *referential* (*ref*) or a *non-referential* (*nref*) NP. According to the HPSG binding theory, see Pollard and Sag (1994, ch. 6), only referential NPs are subject to the syntactic binding principles. The separation of 'referentiality' from 'anaphoricity' allows us to account for the fact that only certain verbs with the RM are semantically reflexivized (see Kubiński (1987) for a similar solution couched within the framework of Arc Pair Grammar).

No reference restrictions are attributed to the anaphoric RM: it can be associated with a non-human object, e.g., *psy* 'dogs' in (3.20) on p. 93, or a human entity, e.g., *Zosia* in (3.56a). On the other hand, as mentioned in §3.1.3, we assume that in other constructions the RM is not referential. We do not provide any semantic restrictions on the interpretation of the RM in its lexical entry in (3.81) (RESTR is an empty set) and assume that the reference of *się* is constrained (if necessary) by semantics of a verb or pragmatic factors but not by the RM itself.

The RM can be a *structural* NP (see also Rivero (1999)): the RM can occur with verbs which require an accusative complement if the verb is non-negated, (3.56b), and genitive if the verb is negated, (3.40d) (the so-called genitive of negation, see p. 63). Additionally, the RM may occur with verbs which always take a genitive NP, (3.56c), and we assume that the RM can also occur in the *lexical genitive* case. Following the suggestion made in Przepiórkowski (1999, p. 47) for Russian, we assume a sort *acc-gen* in the sort hierarchy of Polish morphological cases and specify the CASE value of the RM to be an object of this sort, see (3.81). As a partial sort hierarchy in (3.82) indicates, *acc-gen* is a subsort of *case* and an immediate supersort of *acc* and *gen* sorts, which in turn have their own subsorts.

(3.82)



Although we do not treat the RM as a semantic argument of the verb, see §3.3, we assume that the RM is subcategorized for by the verb on a par with syntactic arguments, i.e., the RM is present on COMPS and ARG-ST lists. Thus, the general case principle for Polish proposed in Przepiórkowski (1999) takes care of resolving the CASE value of the RM.

Unlike its Romance counterparts (cf. Miller (1992a) (for French) and Monachesi (1995, 1999) (for Italian)), the Polish reflexive clitic *się* is a syntactic item and it is licensed in syntax by general principles and ID schemata introduced in Chapter 2 (§2.5.3). No lexical mechanisms are necessary for realisation of *się*. However, as in analyses of the Romance RM, e.g., Grimshaw (1982), Alsina (1996), Abeillé *et al.* (1998), Monachesi (1999), we assume that the presence of the RM in the subcategorization frame of the verb is lexically conditioned. Reflexivized verbs are either derived from their non-reflexive counterparts or directly marked with the RM in the lexicon if no corresponding non-reflexivized form is available. In the subsequent sections we provide a lexical analysis of verb forms with *się* considered so far.

3.4.2 Inherent Reflexives

Inherently reflexive verbs (called reflexiva tantum in Polish linguistic tradition) do not have corresponding forms without *się*, cf. (3.8) repeated below, and should be listed with the RM directly in the lexicon.

(3.8) a. Jan bał się / *siebie Piotra.
John$_{nom}$ feared RM$_{inh}$ self Peter$_{gen}$
'John was afraid of Peter.'

b. Nie zadawaj się / *siebie z nim.
NM deal RM$_{inh}$ self with him
'Stay away from him.'

Although some reflexiva tantum are morphologically derived from verbal stems, e.g., *najeść się* 'to eat till satisfaction' vs. *jeść* 'to eat' (there is no verb **najeść*), certain inherent reflexives are related to a nominal, e.g., *szarogęsić się* 'to boss around' (lit.: 'to greygoose'), or an adjectival stem, e.g., *zielenić się* 'to green'. Moreover, there are verbs for which no morphological derivation can be traced, e.g., *bać się* 'to fear' in (3.8a) (see Wilczewska (1966), Szymańska (1998) for more examples). Then, inherent reflexives are not syntactically coherent. Many verbs subcategorize only for the RM (apart from the subject), e.g., *spóźnić się* 'to be late' or *zielenić się* 'to green', but some verbs take NP (or PP) complements as well, e.g., (3.8) or (3.59). Reflexiva tantum do not form a uniform semantic class either. For example, *ślizgać się* 'to slide' is an unaccusative while *uśmiechać się* 'to smile' is an unergative verb, according to Kubiński (1987, pp. 44–45). Therefore, we assume that all inherent reflexives do not form a uniform group and are enumerated in the lexicon.

A partial lexical entry for *boi się* 'is afraid of', as in (3.8a), is given in (3.83).

(3.83)
$$\left[\begin{array}{l} \textit{word} \\ \text{PHON } \langle \text{boi} \rangle \\ \text{SYNSEM|LOC} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } \textit{verb} \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ } \boxed{1} \\ \text{COMPS } \boxed{2} \end{array}\right] \\ \text{ARG-ST } \boxed{1}\langle \text{NP}[\textit{str}]{:}\boxed{3}\rangle \oplus \boxed{2}\ \langle \text{RM}_{nref}, \text{NP}[\textit{lgen}]{:}\boxed{4}\rangle \end{array}\right] \\ \text{CONT} \left[\begin{array}{l} \textit{psoa} \\ \text{QUANTS } \langle\rangle \\ \text{NUCLEUS} \left[\begin{array}{l} \textit{fear-rel} \\ \text{ACTOR } \boxed{3} \\ \text{SOA } [\text{UNDERGOER } \boxed{4}] \end{array}\right] \end{array}\right] \end{array}\right] \end{array}\right]$$

Since the values of *word*'s valence attributes can be reconstructed from ARG-ST (see the principle in (1.12) on p. 18), we will usually specify just the latter.

As we said in §3.4.1, we represent the RM as a syntactic element subcategorized by the verb. Unlike in Kupść (1999d), we do not introduce a separate valence attribute to encode the syntactic requirement for *się*. The RM is present on the COMPS list and, subsequently, on the

ARG-ST list as well. In (3.83), and later on in this section, we use 'RM' as the abbreviation for the description of the SYNSEM component of the object described in (3.81).

We do not treat *się* in reflexiva tantum as a referential NP; the LOC|CAT|CONT|INDEX value of 'RM' in (3.83) is an object of the sort *nref*. Hence, the RM in (3.83) is not subject to binding principles and the predicate (correctly) will not be interpreted as semantically reflexive.

Finally, the RM is not assigned any semantic role in (3.83), unlike the two other elements on the ARG-ST list. (See Alsina (1996), Miller and Sag (1997) and Abeillé *et al.* (1998) for an analogous account of Romance intrinsic clitics.) We follow Davis (1997), Koenig and Davis (1999) and assume that the value of attributes specifying semantic roles is either *content*, $\boxed{3}$ and $\boxed{4}$ in (3.83), instead of an *index* object in Pollard and Sag (1994, p. 399), or a subsort of *qfpsoa* (a semantic relation) for SOA (the attribute which denotes the resulting state of affairs, cf. Koenig and Davis (1999)). This will allow us to correctly capture the semantic contribution of the RM in §3.4.3.

In (3.83), the CONT value appears in a complete form as specified in Pollard and Sag (1994, p. 399). Since quantification does not concern us here, we will drop the QUANTS and NUCLEUS attributes and simplify the verbal CONT (as in §2.3.4). Semantic relations (subsorts of *qfpsoa*, e.g., *fear-rel* in (3.83)) will be used directly as verbal CONT value. We largely adopt the encoding and the hierarchy of semantic relations as proposed in Davis (1997), Koenig and Davis (1999), Kordoni (1999) (see §3.4.6 for more details), which will allow us to express certain common semantic properties of reflexivized verbs. For example, semantics of the experiencer subject psych verb *bać się* 'fear' in (3.83) indicates that the ACTOR is linked to the subject whereas the object is the UNDERGOER of the resulting mental state described by a semantic relation specified by the value of the SOA attribute.

3.4.3 Semantic Reflexives

Most verbs which take a direct (accusative) NP complement can also occur with the reflexive *się*, cf. (3.4a) repeated below, or (3.84).

(3.4a) Janek zobaczył się / siebie / Tomka w lustrze.
John saw RM_{refl} self_{refl} Tom_{acc} in mirror

'John saw himself / Tom in the mirror.'

(3.84) Janek uderzył się / siebie / Tomka.
John hit RM_{refl} self_{refl} Tom_{acc}

'John hit himself / Tom.'

As we argued in §3.3.1, the reflexive RM is not an equivalent of the syntactic complement *siebie* and it is not a semantic argument of the verb. We account for this difference assuming that the verb subcategorizes for any NP[*str*], i.e., both the complement (*siebie* or other) or the RM can be used, but only the former is a semantic argument. As we show shortly, this distinction is crucial to explain the behaviour of subject psych verbs discussed in §3.2.7 and §3.3.1.

The relevant part of the lexical entry for the verb *uderzył* 'hit' in (3.84) is given in (3.85).

(3.85)
$$\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{word}}\\ \text{PHON} & \langle\text{uderzył}\rangle\\ \text{SS}|\text{LOC} & \left[\begin{array}{ll}\text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \textit{verb}\\ \text{ARG-ST} & \langle \text{NP}_{ref}[\textit{str}]{:}\boxed{1}, \text{NP}_{ref}[\textit{str}]{:}\boxed{2}\rangle\end{array}\right]\\ \text{CONT} & \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{hit-rel}}\\ \text{ACTOR} & \boxed{1}\\ \text{UNDERGOER} & \textit{content}\end{array}\right]\end{array}\right]\end{array}\right]$$

We assume that the verb *uderzył* in (3.85) requires two referential arguments, the first of which is interpreted as the ACTOR. The CONT value in (3.85), however, is not fully specified: the second semantic argument of the verb, UNDERGOER, does not correspond (is not linked) yet to any syntactic argument (an element of ARG-ST).

The RM can be a structural NP, cf. (3.81), hence it is compatible with the description of the second element on ARG-ST in (3.85), which predicts the syntactic presence of the RM in (3.84). As repeatedly stated above, we do not take (even the referential) RM to be a semantic argument of the verb. Although the RM is present on ARG-ST of the verb, it is not assigned any semantic role, see (3.86).

(3.86)
$$\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{word}}\\ \text{PHON} & \langle\text{uderzył}\rangle\\ \text{SS}|\text{LOC} & \left[\begin{array}{ll}\text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \textit{verb}\\ \text{ARG-ST} & \langle \text{NP}_{\boxed{2}}[\textit{str}]{:}\boxed{1}, \text{RM}_{\boxed{2}}\rangle\end{array}\right]\\ \text{CONT} & \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{hit-rel}}\\ \text{ACTOR} & \boxed{1}\\ \text{UNDERGOER} & \boxed{1}\end{array}\right]\end{array}\right]\end{array}\right]$$

We use objects of the sort *content* rather than *index* as values of semantic arguments, see §3.4.2. This modification allows us not to assign any semantic role to the RM and simultaneously account for its binding properties: the referential INDEX value of the RM, $\boxed{2}$, is shared with that of the subject NP (due to the HPSG binding principle proposed for Polish in Marciniak (1999)). In our account none of the semantic roles of the reflexivized verb is suppressed, unlike in Lubowicz (2000), but all semantic roles are assigned to the same syntactic argument (see Davis (1997) for a similar treatment of Dutch reflexive verbs). This reflects the 'subject-internalizing' and neutralisation of 'agent-patient' properties of reflexivized verbs advocated in Schenker (1985, 1993).

On the other hand, if any other NP is used (including the anaphoric complement *siebie*), it is assigned a semantic role, see (3.87).

(3.87)
$$\left[\begin{array}{ll}\multicolumn{2}{l}{\textit{word}}\\ \text{PHON} & \langle\text{uderzył}\rangle\\ \text{SS}|\text{LOC} & \left[\begin{array}{ll}\text{CAT} & \left[\begin{array}{ll}\text{HEAD} & \textit{verb}\\ \text{ARG-ST} & \langle \text{NP}[\textit{str}]{:}\boxed{1}, \text{NP}[\textit{str}]{:}\boxed{2}\rangle\end{array}\right]\\ \text{CONT} & \left[\begin{array}{ll}\multicolumn{2}{l}{\textit{hit-rel}}\\ \text{ACTOR} & \boxed{1}\\ \text{UNDERGOER} & \boxed{2}\end{array}\right]\end{array}\right]\end{array}\right]$$

Our analysis predicts that semantic contribution of the reflexive RM and a genuine NP complement *siebie* are different. Indeed, as indicated in (3.60), both nominals have distinct properties in 'statue' sentences. We use the same semantic relation, *hit-rel*, to represent semantics of the verb in (3.84) but the assignment of semantic roles depends on whether the RM or some other NP is used. The RM is not assigned any semantic role and, unlike other NPs, may force semantic arguments to 'merge' and relate to a single syntactic argument.

This approach is able to explain also the puzzling behaviour of subject psych verbs, see §3.3.1 ((3.53a) is repeated below).

(3.53a)	Jan	siebie	/ ?*się	lubi	/ rozumie	/ kocha	/ podziwia.
	John	self$_{refl}$	RM$_{?*refl}$	likes	understands	loves	admires

‘John likes / understands / loves / admires himself.’

Unlike in Kupść (1999d, 2000b), we do not exclude such sentences by restricting the combinatory potential of subject psych verbs to ‘non-RM’ NPs. We do not provide any syntactic restrictions on the ARG-ST but we rule out (3.53a) with the reflexive *się* on semantic grounds. Let us consider the partial lexical entry for the transitive subject psych verb *lubi* ‘likes’.

$$(3.88)\quad \begin{bmatrix} \textit{word} \\ \text{PHON } \langle \text{lubi} \rangle \\ \text{SS}|\text{LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{ARG-ST } \langle \text{NP}_{ref}[\textit{str}]{:}\boxed{1}, \text{NP}_{ref}[\textit{str}]{:}\boxed{2} \rangle \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \textit{like-rel} \\ \text{ACTOR } \boxed{1} \\ \text{SOA } \begin{bmatrix} \text{UNDERGOER } \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Both arguments of the verb (elements of the ARG-ST list) are referential NPs and, syntactically, the RM is compatible with the description of the second NP: the RM can be assigned a structural case, see (3.81). As argued in Wechsler (1995), subject psych verbs require a specific semantic entailment to hold: the experiencer of the mental state denoted by the subject must know (have a notion) of the entity denoted by the object (the *like-rel* relation in (3.88), just like *fear-rel* in (3.83), are subsorts of the *notion-rel* relation proposed in Koenig and Davis (1999)). If the reflexive RM occurred as a complement, it would be assigned a semantic role, contrary to our previous conclusion. On the other hand, as the complement *siebie* does correspond to a semantic argument, all semantic roles can be successfully assigned in (3.88) and (3.53a) with *siebie* ‘self’ is both syntactically and semantically well-formed.

3.4.4 Middles

As said in §3.1.2, we use the term ‘middles’ rather broadly and apply it to all constructions where the subject of a reflexivized verb corresponds to the direct (structural) object of the equivalent strictly transitive verb,[22] see (3.6) repeated below.

(3.6)	a.	Książka	się	dobrze	sprzedawała.
		book	RM$_{mid}$	well	sold

‘The book was selling well.’

	b.	Szklanka	się	zbiła.
		glass	RM$_{mid}$	broke

‘The glass broke.’

[22] Polish middles cannot be formed from ditransitive verbs, cf. (i), or transitive verbs which do not have the accusative complement, (ii).

(i)	*	Kwiaty	się	dają	Marysi.	
		flowers	RM$_{mid}$	give	to-Mary	
(ii)	*	Ten kraj	się	trudno	rządził.	
		this country	RM$_{mid}$	difficult	governed	

Therefore, not only do we ignore a semantic distinction between true middle, (3.6a), and inchoative (change of state), (3.6b), sentences but (as in Szymańska and Śpiewak (1998)) we also merge the different semantic shades of inchoatives distinguished in Rozwadowska (1992).

Although the syntax of middles resembles that of passives, see §3.1.2, there is a semantic difference between these two constructions, cf. Kański (1986) and Szymańska and Śpiewak (1998, p. 122).

(3.89) a. Książka została sprzedana.
book was sold
'The book was sold.'

b. Szklanka została zbita.
glass was broken
'The glass was broken.'

Middle (inchoative) sentences, (3.6), are usually perceived as more 'agentive' than true passives, (3.89). To put it differently, "the 'Patient' of the inchoative is not the same as the 'Patient' of the passive", cf. Szymańska and Śpiewak (1998, p. 122). Hence, we do not assume that in middles one of the semantic roles is suppressed or that only the 'Patient' is present. Instead, we preserve all semantic arguments of the active verb but we relate them to a single syntactic argument, i.e., the subject (similarly to semantic reflexives in §3.4.3). In this way, the RM is not assigned any semantic role by the verb, see, e.g., Alsina (1996), Abeillé *et al.* (1998) for a similar treatment of the Romance medio-passive RM.

As discussed in Rozwadowska (1992), Szymańska and Śpiewak (1998), Kupść (2000b), a purely syntactic account of middle constructions is too liberal. It is not our task to fully account for semantic properties of these constructions. Certainly, a more careful analysis is required to correctly specify and capture all semantic shades of what we call 'middles' here. Nevertheless, we incorporate into our analysis some observations with respect to mostly inchoative verbs made in Rozwadowska (1992) and Szymańska and Śpiewak (1998).

We use the following lexical principle to obtain middle verbs (among others) (rather than a lexical rule as in Grimshaw (1982), Abeillé *et al.* (1998) (for French medio-passives), Monachesi (1995, 1999) (for the Italian middle *si*) and Kupść (1999d, 2000b) (for Polish middles)).

(3.90)
$$\begin{bmatrix} \textit{word} & \\ \textsc{ss}|\textsc{loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textit{category} \\ \textsc{head}\ \textit{personal} \\ \textsc{arg-st}\ \langle \text{NP}[\textit{str}],\ \text{NP}[\textit{str}] \rangle \end{bmatrix} \\ \textsc{cont} & \begin{bmatrix} \textit{act\_und\_rel} \\ \textsc{actor}\ \textit{content} \\ \textsc{undergoer}\ \textit{content} \end{bmatrix} \end{bmatrix} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \textsc{ss}|\textsc{loc} & \begin{bmatrix} \textsc{cat}|\textsc{arg-st}\ \langle \text{NP}[\textit{str}]{:}\boxed{1},\ \text{NP}[\textit{str}]{:}\boxed{2} \rangle \\ \textsc{cont}\ \begin{bmatrix} \textit{act\_und\_rel} \\ \textsc{actor}\ \boxed{1} \\ \textsc{undergoer}\ \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \vee$$

$$\begin{bmatrix} \textsc{ss}|\textsc{loc} & \begin{bmatrix} \textsc{cat}|\textsc{arg-st}\ \langle \text{NP}[\textit{str}]{:}\boxed{3},\ \text{RM} \rangle \\ \textsc{cont}\ \begin{bmatrix} \textit{act\_und\_rel} \\ \textsc{actor}\ \boxed{3} \\ \textsc{undergoer}\ \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Rozwadowska (1992, p. 51) observes that only a patient-like object of a transitive verb can become the subject of an inchoative verb (an anti-causative or a passive reflexive verb in her terms). We encode this restriction in (3.90) by specifying the semantics (CONT) of the verb as *act_und_rel*, see §3.4.6 for details. This semantic relation, taken from Davis (1997) and Koenig and Davis (1999), has two attributes, ACTOR and UNDERGOER, which roughly correspond to the agent-like and the patient-like characteristics of NPs in Rozwadowska (1992). Therefore, (3.90) concerns only strictly transitive verbs with such semantic properties (in particular, it does not concern subject psych verbs discussed in §3.4.3).

The implication in (3.90) specifies assignment of the ACTOR and UNDERGOER semantic roles in a strictly transitive verb.[23] If the verb takes a genuine NP complement, both syntactic arguments are assigned semantic roles (as we will see shortly, the principles given in §3.4.6 ensure that the RM does not occur on the ARG-ST list in the first disjunct of the consequent in (3.90)). If the RM is present, however, it is not assigned any semantic role and both semantic arguments are mapped to the subject (the second disjunct). The latter description strongly resembles the description of the reflexivized verb *uderzył* '(he) hit' in (3.86). Indeed, the referentiality of the RM in (3.90) is not specified, hence both the referential and non-referential RM may appear. In the former case, the verb can be interpreted as semantically reflexive (hence, it can be a description of *uderzył* 'hit' in (3.86)) whereas in the latter a middle form results. Also semantic properties of the subject (and the verb itself) play a role in the interpretation of the reflexivized verb (the second disjunct). If the subject has agent-like properties, a semantically reflexive reading results. On the other hand, if the subject is more patient-like, the middle interpretation is obtained. Whichever reading is available depends on the lexical semantics of the verb since both semantic roles of a reflexivized verb are indistinguishable as they are linked to the same syntactic argument. This analysis allows us to capture semantic similarities as well as differences between middles (inchoatives) and semantic reflexives discussed in Schenker (1985, 1993) and Szymańska and Śpiewak (1998), among others.

3.4.5 Impersonals

As discussed in §3.3.2, impersonal constructions (IC) are very productive in Polish. They can be formed from intransitive, (3.61a), or (di)transitive active verbs, (3.61b), but also some reflexiva tantum, (3.91) (e.g., contra Rozwadowska (1992)), copula, (3.66a), passives, (3.66b), modals or control verbs, (3.68a).

(3.91) W tym kościele dobrze (mu) się modliło.
in this church well him$_{cl,dat}$ RM$_{imp}$ prayed
'One (he) prayed well in this church.' Dziwirek (1994, p. 64)

Moreover, the agent can be usually overtly expressed as NP[dat] (some exceptions noticed in Rivero (1999) are discussed below) and the presence of an adverbial or an adjunct is required to render the sentence fully grammatical, cf. Kański (1986).

Since only one verb form can appear in IC, we assume that there is no corresponding full paradigm with the impersonal RM, see §3.2.7. Due to syntactic and semantic differences

[23] Logically, more alternatives are possible, e.g., the subject can be linked to the UNDERGOER whereas the object to the ACTOR. These options, if attested in the language, may be added as disjuncts to the consequent of (3.90).

$$\begin{bmatrix} \textit{word} & \\ \text{SS}|\text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \textit{category} & \\ \text{HEAD}|\text{VFORM} & \textit{personal} \\ \text{ARG-ST} & \langle \text{NP}[\textit{str}]{:}\boxed{1}\rangle \oplus \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \mapsto$$

$$\begin{bmatrix} \text{SS}|\text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD}|\text{VFORM} & \textit{impersonal} \\ \text{VAL} & [\text{SUBJ}\ \langle\rangle] \\ \text{ARG-ST} & \langle \boxed{3}\text{NP}[\textit{str} \vee \textit{ldat}]{:}\boxed{1}\rangle \oplus \langle \boxed{4}\text{RM}_{nref}\rangle \oplus \boxed{2} \oplus \boxed{5}\textit{ne\_list}(\textit{adjunt}) \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{act\_rel} & \\ \text{ACTOR} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 3.2: The lexical rule for impersonals

between personal and impersonal verb forms, we use the lexical rule in fig. 3.2 to obtain this verb form, similarly to Monachesi (1999) (for Italian impersonals with *si*). We follow Meurers (1999) in specifying explicitly only differences between the input (the left-hand side of '$\mapsto$' in fig. 3.2) and the output (right-hand side of '$\mapsto$') of the lexical rule.[24]

The rule in fig. 3.2 provides impersonal forms in dativeless ICs ($\boxed{3}$ is assigned a structural case) and dative ICs ($\boxed{3}$ is a dative NP). As argued in §3.3.2, IC do not have the overt subject (hence [SUBJ ⟨⟩] in the output of the rule) but the agent is always semantically present. In dativeless ICs, the presence of a syntactically unrealized semantic argument is indicated by the behaviour of ICs with respect to case marking on predicative adjectives, binding and control. As discussed in §3.3.2, analogous properties of infinitives and gerunds are explained in Przepiórkowski (1999) by postulating the presence of PRO, an NP assigned the structural case *case-less* (*cless*). The case principle proposed for Polish in Przepiórkowski (1999, p. 438) ensures that the case value of the $\boxed{3}$NP[*str*] in fig. 3.2 is correctly resolved to *cless*. In contrast to other sorts associated with cases in Polish, the sort *cless* is not morphologically realized. The principle in (3.92), cf. Przepiórkowski (1999, p. 221), ensures that no overt *sign* is assigned *cless* CASE value.

(3.92) $$\begin{bmatrix} \textit{sign} \\ \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD}|\text{CASE}\ \boxed{1} \end{bmatrix} \rightarrow \boxed{1}\textit{morph-case}$$

On the other hand, in dative ICs $\boxed{3}$NP[*ldat*] is the first element on the ARG-ST list of the impersonal verb. Due to our specification of the relationship between ARG-ST and valence lists in §1.3.1, $\boxed{3}$NP is always present on the COMPS list of the verb since the SUBJ list is empty.

As discussed in §3.3.2, IC may be formed only from verbs which can occur with the nominative subject NP, see (3.63). This condition is reflected in the description of the input verb since only the first NP on ARG-ST of a personal verb can be assigned the nominative case according to the case principle of Przepiórkowski (1999).

In IC not only are all complements of the original verb maintained, $\boxed{2}$, but also the RM ($\boxed{4}$) and other elements ($\boxed{5}$) are added. As argued in Kański (1986), IC are fully grammatical only if a verbal modifier is present. In fig. 3.2, this requirement is reflected by adding a list of adjuncts, $\boxed{5}$, to the output's COMPS list, cf. van Noord and Bouma (1994), Bouma *et al.* (1998), Przepiórkowski (1999). The different HEAD|VFORM and VAL values of the input and output verbs correctly prevent a recursive application of the rule.

[24]Formally, description level lexical rules adopted here are objects of the sort *lex_rule* which has two appropriate attributes IN and OUT relating words in the input (the left-hand side of the rule) and the output (the right-hand side of the rule), see Meurers (1999, pp. 120ff).

This analysis allows us to correctly account for binding and control properties of IC. Whenever an anaphoric referential NP is an element of $\boxed{2}$, it is bound by the first element of the ARG-ST list, i.e., PRO or NP[dat], see (3.71) and (3.76). We treat the RM in ICs as a non-referential NP. Thus, binding principles do not apply to it and the verb is not a semantically reflexive predicate.

In contrast to Kupść (1999d, 2000b), the present approach allows us to capture the case assignment on predicative adjectives in IC, cf. (3.66), (3.68), (3.69), (3.70), (3.78).

(3.66) a. Było się miłym / *miły.
was RM$_{imp}$ nice$_{inst}$ nice$_{nom}$
'One was nice.'

b. Często jest się zapraszanym / *zapraszany przez przyjaciół.
often is RM$_{imp}$ invited$_{inst}$ invited$_{nom}$ by friends
'One is often invited by friends.'

Such facts are correctly accounted for if we adopt the case (dis-)agreement principle, proposed in Przepiórkowski (1999), cf. (3.93).[25]

(3.93)
$$\left[\textsc{ss}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{val}|\textsc{subj}\ \langle \text{XP}[\textsc{case}\ \boxed{1}]\rangle \\ \textsc{head}|\textsc{case}\ \boxed{2} \end{bmatrix}\right] \rightarrow$$
$$(\boxed{1}{=}\boxed{2}) \vee (\boxed{1}\mathit{str} \wedge \boxed{2}\mathit{lins})$$

This principle encodes the case (dis-)agreement between a predicative phrase (e.g., the adjective *miłym* 'nice' or the participle *zapraszanym* 'invited' in (3.66)) and its subject (an XP present on the SUBJ list). According to (3.93), either the predicative phrase agrees in case with its subject (the first disjunct in the consequent) or they have distinct case values (the second disjunct in the consequent): the predicative phrase appears in instrumental whereas its subject is assigned a structural case. (Note that verbal predicates are immune to (3.93) since the CASE attribute is not appropriate to verbs.) Let us see how this principle deals with case assignment on predicative AdjPs in ICs taking (3.66) (repeated above) as an example.

In (3.66), IC is dativeless and PRO is the first element on the ARG-ST list of the verb. The copula is a subject raising verb here, i.e., its subject (PRO) is shared with the subject of the predicative complement, i.e., the adjective in (3.66a) and the participle in (3.66b). Since PRO bears the structural (*cless*) case value, the predicative adjective or the participle must occur in the disagreeing instrumental case (*cless* is not a morphological case, it is never overtly realised, see (3.92), and the case-agreement option is impossible in (3.66)).

Similarly, the case marking in (3.78) (see below) is predicted.

(3.78) Jankowi dobrze się szło wypoczętemu / ?*wypoczętym / *wypoczęty.
John$_{dat}$ well RM$_{imp}$ walked relaxed$_{dat}$ relaxed$_{inst}$ relaxed$_{nom}$
'John enjoyed walking after some rest.'

The adjective *wypoczęty* 'relaxed' predicates of the agent, NP[dat], which is an element of the adjective's SUBJ list. Since dative is the lexical case in Polish, see Przepiórkowski (1999), only the case-agreeing option in (3.93) is available. The adjective is correctly predicted to appear in dative.

Let us briefly comment on semantic restrictions on ICs mentioned in §3.3.2.

[25] For expository reasons, (3.93) is slightly simplified here; see Przepiórkowski (1999§ 5.4) for details.

Although IC are quite productive in Polish, a syntactic specification of these constructions is insufficient. It should be stressed that we do not attempt to provide an elaborate semantic account of ICs here but we try to incorporate some of the constraints noted in Rozwadowska (1992) into our analysis.

In addition to syntactic restrictions mentioned above, we assume that IC imply an 'agentive' interpretation. In fig. 3.2, this is indicated as the *act_rel* value of the output's CONT attribute. This restriction allows us to prevent the incorrect sentence in (3.94) to be formed from the NP[nom]-subject non-agentive verb *boleć* 'to ache / hurt'.

(3.94) * (Nodze) bardzo się boli Jana.
leg$_{dat}$ much RM$_{imp}$ hurts John$_{acc}$

Although nothing prevents application of the rule to such a verb, the non-agentive subject *noga* 'a leg' does not satisfy semantic properties of the ACTOR and cannot be associated with this role.

A similar explanation holds for ICs with experiencer object psych verbs, see (3.65a) repeated below.

(3.65a) * Fascynuje się profesorów swoją wiedzą.
fascinates RM$_{imp}$ professors self's knowledge

'* One fascinates professors with one's knowledge.'

As mentioned in §3.3.2, Polish experiencer object psych verbs are inherently non-agentive. As ICs are agentive predicates, semantics of verbs such as *fascynować* 'fascinate' is incompatible with that of impersonal verbs: the subject cannot be assigned the agent role as required by the output of the lexical rule in fig. 3.2.

As observed in Rivero (1999), NP[dat] cannot be overtly expressed in IC originating from certain verbs. According to Rivero (1999), the copula *być* 'to be' and modal verbs may occur only in plain IC: the agent cannot be overtly realised, see (3.95).

(3.95) a. * Markowi jest się łatwo miłym.
Mark$_{dat}$ is RM$_{imp}$ easily nice
'Mark is easily nice.' (putatively)

b. * Markowi powinno się być łatwo miłym.
Mark$_{dat}$ should RM$_{imp}$ be$_{inf}$ easily nice
'Mark should be easily nice.' (putatively)

In fact, the ungrammaticality of (3.95b) cannot be blamed on the modal verb but rather on the copula since (3.96) is perfectly fine.

(3.96) Markowi powinno się dziś dobrze spać.
Mark$_{dat}$ should RM$_{imp}$ today well sleep$_{inf}$

'Mark should sleep well today.'

Therefore, we conclude that only the copula disallows the agent to be overtly expressed in IC. We assume here that (3.95) should be excluded on semantic grounds but we do not develop the issue any further.

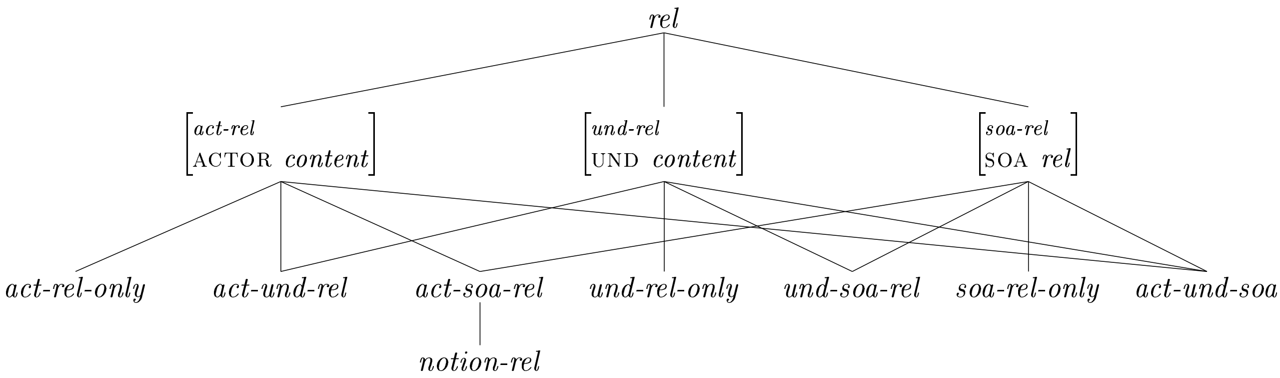


Figure 3.3: A sort hierarchy of semantic relations denoted by subsorts of *rel*

3.4.6 Assignment of Semantic Roles

The account of reflexivized verbs presented in the previous subsections heavily relies on the assumption that the RM is not assigned a semantic role. In this subsection, we explicitly incorporate this fact into our analysis. It should be stressed, however, that we do not attempt at specifying here linking constraints, i.e., which argument of the ARG-ST list is assigned which semantic role (but see Davis (1997), Koenig and Davis (1999), Kordoni (1999), among others).

As we mentioned in §3.4.2, we assume the encoding of semantic roles similar to that in Davis (1997), Koenig and Davis (1999), Kordoni (1999). More specifically, we assume the sort hierarchy for the subsorts of *rel* as in fig. 3.3 (the sort *rel* (*relation*) is a subsort of *qfpsoa*, which denotes semantic relations).

In order to ensure that the RM is not assigned a semantic role, we assume that values of attributes which encode semantic roles (ACTOR, UNDERGOER, SOA, ...) cannot be objects which denote the CONT value of the RM. In other words, if the value of an attribute denoting a semantic role is an object of the sort *ana* (a subsort of *content* denoting anaphoric objects, see §2.3.2), this object cannot be the CONT value of a clitic; see (3.97).

(3.97) a. $\forall \boxed{1}\ \forall \boxed{4}$

$$\boxed{4}\begin{bmatrix} \textit{rel} \\ \text{ACTOR}\ \boxed{1}\textit{ana} \end{bmatrix} \rightarrow \neg \left(\exists \boxed{2}\ \boxed{2}\begin{bmatrix} \textit{clitic} \\ \text{LOC}|\text{CONT}\ \boxed{1} \end{bmatrix} \right)$$

b. $\forall \boxed{1}\ \forall \boxed{4}$

$$\boxed{4}\begin{bmatrix} \textit{rel} \\ \text{UNDERGOER}\ \boxed{1}\textit{ana} \end{bmatrix} \rightarrow \neg \left(\exists \boxed{2}\ \boxed{2}\begin{bmatrix} \textit{clitic} \\ \text{LOC}|\text{CONT}\ \boxed{1} \end{bmatrix} \right)$$

The conditions in (3.97) do not concern the SOA attribute since its value is an object of the sort *rel* (fig. 3.3), rather than *content*, and it will not be related to the RM. Objects in the denotation of the SOA's value, however, may have the attributes ACTOR and UNDERGOER which are related to nominal objects (e.g., (3.88)). As the principles in (3.97) are imposed on objects of the sort *rel*, values of such 'embedded' attributes will not be token-identical with the CONT value of the RM either.

3.5 Haplology

Our discussion of the grammatical status of the RM in §3.2 was mostly motivated by the phenomenon illustrated in (3.1) (repeated below).

(3.1) Boję się głośno roześmiać.
fear-I RM aloud laugh_{inf}
'I'm afraid to laugh aloud.'

Both verbs in (3.1) are reflexiva tantum verbs, i.e., they obligatorily occur with *się*. In (3.1), however, a single occurrence of the RM is sufficient and the meaning of the sentence is preserved. We refer to such an omission without a consequence for the well-formedness of an expression as 'haplology' and discuss this phenomenon in detail in this section.

We consider several environments where haplology of the RM may occur: simple clauses, verbal complexes (Verb Clusters (VCs), such as (3.1)), coordinated structures and gerund phrases. Haplology of the Polish *się* (in various environments) has been previously noticed in the literature, e.g., Safarewiczowa (1954), Saloni (1976), Fowler (1993), Kopcińska (1997), Rappaport (1998), Kupść (1999d, 2000b), Derwojedowa (1998), Przepiórkowski (1999) but only Rappaport (1998) and Kupść (1999d) attempt to give a comprehensive analysis of this phenomenon. Moreover, none of the above mentioned authors has considered haplology in all environments.

The aim of this section is to examine haplology in the environments enumerated above and provide an HPSG account of this phenomenon. According to the definition in Crystal (1997), haplology is "[a] term used in phonology, in both synchronic and diachronic contexts, to refer to the omission of some of the sounds occurring in a sequence of similar articulations." We show, however, that 'haplology' of the Polish RM is not phonologically (or morphologically) conditioned and, thus, does not pose a challenge for the syntactic status of the RM advocated in §3.2. In the present context, the term 'haplology' is a misnomer but we retain this name in order to be consistent with the terminology introduced in Rappaport (1998) (and adopted in Kupść (1999d, 2000b), Derwojedowa (1998), Przepiórkowski (1999)). Since haplology (by its very nature) involves merging of several items which usually have different functions, we investigate how various types of *się* interact with this phenomenon.

3.5.1 Simple Clauses

In simple clauses, a single verb cannot appear with several *się*, irrespectively of their function. (See Saloni (1976, p. 107), Holvoet (1991, p. 146), Rozwadowska (1992, p. 62), Kopcińska (1997, p. 30), Szymańska and Śpiewak (1998, p. 126), Derwojedowa (1998, p. 100) and Rappaport (1998, p. 467) for a similar observation.)

(3.98) a. * Jan się bał się.
John_{nom} RM_{refl} feared RM_{inh}
'John was afraid of himself.' (putatively)

b. * Janowi łatwo się myje się w ciepłej wodzie.
John_{dat} easily RM_{imp} washes RM_{refl} in warm water
'John finds it easy to wash himself in warm water.' (putatively)

c. * Przyjemnie się nam się przechadza po parku.
nicely RM_{imp} $\text{us}_{dat,cl}$ RM_{inh} strolls after park

'We enjoy strolling in the park.' (putatively)

d. * Ziemniakom się szybko się ugotowało.
potatoes$_{dat}$ RM$_{imp}$ quickly RM$_{mid}$ cooked
'The potatoes cooked quickly.' (putatively)

Note that in each sentence in (3.98) RMs have different functions and, a priori, their co-occurrence should be possible.

As observed in Kupść (1999d, p. 103), the corresponding grammatical sentences with a single *się* do not always have a haplology meaning, i.e., a single RM does not always take over the function(s) of omitted one(s), cf. (3.99).

(3.99) a. Jan się bał.
John RM$_{inh}$ feared
'John was afraid.'

b. Janowi łatwo się myje w ciepłej wodzie.
John$_{dat}$ easily RM$_{imp/refl}$ washes in warm water
'John finds it easy to wash (himself) in warm water.' (ambiguous)

c. Przyjemnie się nam przechadza po parku.
nicely RM$_{imp/inh}$ us$_{dat,cl}$ strolls in park
'We enjoy strolling in the park.'

d. Ziemniakom się szybko ugotowało.
potatoes$_{dat}$ RM$_{imp/mid}$ quickly cooked
'The potatoes cooked quickly.'

(3.99a) means only that John was afraid (an 'object-drop' reading), not that he was afraid of himself, which would be expected if the RM took over functions of the inherent and reflexive RMs (haplology interpretation). On the other hand, (3.99b) is ambiguous between true haplology (John enjoys washing himself) and an 'object-drop' reading (washing of an unspecified object). Only the examples in (3.99c–d) have unambiguous haplology interpretations.

The ungrammaticality of (3.98) (and (3.98a–b) in particular) cannot be blamed on the fact that the presence of the RM blocks another anaphoric nominal since (3.100) are perfectly grammatical.

(3.100) a. Jan się bał siebie.
John$_{nom}$ RM$_{inh}$ feared self
'John was afraid of himself.'

b. Janowi łatwo się myje siebie w ciepłej wodzie.
John$_{dat}$ easily RM$_{imp}$ washes self in warm water
'John finds it easy to wash himself in warm water.'

Moreover, the contrast between (3.98a–b) and (3.100) shows that the RM is not always interchangeable with the syntactic complement *siebie*, which we have seen previously, e.g., §3.2.7. The verb *bał się* 'feared' requires an NP[gen] complement, see (3.83), but the RM cannot substitute a syntactic complement, §3.2.7. Hence, the syntactic complement in (3.99a) is missing and the 'object-drop' reading results. The verbs in (3.99c–d), i.e., the inherently reflexive *przechadzać się* 'to stroll' and the middle *ugotować się* 'to cook', obligatorily occur with the RM but they do not have NP complements. Since several RMs cannot co-occur (impersonals appear with their own RM, see e.g., (3.66) or (3.78) in §3.4.5), RMs have to be 'merged' and haplology results. As the (omitted) RMs are never complements, 'object-drop' readings are not available in (3.99c–d). The ambiguity of (3.99b) follows from the fact that the verb *myje* 'washes' either occurs with a syntactic complement and, then, an 'object-drop' reading results, or the impersonal form is related to the reflexivized *myje się* 'washes', which leads to haplology.

Also other languages observe restrictions on multiple occurrences of the RM. For example, in Standard Italian, a sequence of the RM *si* is impossible, cf. (3.101a), and (3.101b) has to be used instead.

(3.101) a. * Si si lava.
RM$_{imp}$ RM$_{refl}$ washes

b. Ci si lava.
ci RM washes
'One washes oneself.'

The impossibility of several RMs in Polish, see (3.98), cannot be attributed to phonology, as has been proposed for the Italian *si* in Wanner (1977), or morphology, cf. Bonet (1995), Monachesi (1999). In none of the examples in (3.98) are the two *się* adjacent or do they form a phonological sequence (as discussed in §3.2.6, the distribution of the Polish RM is relatively free and clitics do not have to cluster; hence (3.98) is not independently blocked by the requirement on clitics to cluster). The ungrammaticality of (3.98) is not morphologically conditioned either. As (3.99) vs. (3.102) show, which one of the co-occurring RMs is dropped does not depend on the morphosyntactic properties of the host (as discussed in §3.2.5, *się* is promiscuous and different hosts in (3.98) are fully legitimate).

(3.102) a. Jan bał się.
John feared RM$_{inh}$
'John was afraid.'

b. Janowi łatwo myje się w ciepłej wodzie.
John$_{dat}$ easily washes RM$_{imp/refl}$ in warm water
'John finds it easy to wash (himself) in warm water.' (ambiguous)

c. Przyjemnie nam się przechadza po parku.
nicely us$_{dat,cl}$ RM$_{imp/inh}$ strolls in park
'We enjoy strolling in the park.'

d. Ziemniakom szybko się ugotowało.
potatoes$_{dat}$ quickly RM$_{imp/mid}$ cooked
'The potatoes cooked quickly.'

In order to account for the fact that a single verb may occur with at most one RM, we modify the principle which relates valence attributes to the ARG-ST list. The revised version of (1.12) from §1.3.1 is given in (3.103).

(3.103) $$word \rightarrow \left[\textsc{synsem}|\textsc{local}|\textsc{category} \begin{bmatrix} \textit{category} \\ \textsc{valence} \begin{bmatrix} \textit{valence} \\ \textsc{subj}\ \boxed{1} \\ \textsc{comps}\ \boxed{2} \end{bmatrix} \\ \textsc{arg-st}\ \boxed{1} \oplus \boxed{2'} \end{bmatrix}\right] \wedge \texttt{arg2comps}(\boxed{2}, \boxed{2'})$$

$$\texttt{arg2comps}(\boxed{0}, \boxed{0}) \overset{\forall}{\Longleftarrow} \texttt{no\_member}(\mathrm{RM}, \boxed{0})$$

$$\begin{aligned} \texttt{arg2comps}(\boxed{0}, \boxed{1}) \overset{\forall}{\Longleftarrow}\ & \texttt{remove}(\boxed{2}\mathrm{RM}, \boxed{0}, \boxed{0'}) \wedge \texttt{no\_member}(\mathrm{RM}, \boxed{0'}) \wedge \\ & \texttt{shuffle}(\langle \boxed{2} \mid \boxed{3} \rangle, \boxed{0'}, \boxed{1}) \wedge \texttt{list\_RM}(\boxed{3}) \end{aligned}$$

The relation **arg2comps** specifies that if there is at least one RM on the sublist of ARG-ST which contains non-subject arguments ($\boxed{2'}$ in (3.103)), exactly one RM is present on the COMPS list; if this sublist contains no RMs, it equals COMPS.[26]

If we additionally assume that the RM is never an element of the SUBJ list, see (3.104), principles (3.103) and (3.104) conspire together to ensure that there can be at most one syntactically realized RM (and it is an element of the COMPS list $\boxed{2}$), irrespectively of the number of RMs present on the ARG-ST list.

(3.104)
$$\forall \boxed{1}\ \forall \boxed{2} \left(\begin{array}{l} \left(\left[\mathrm{SYNSEM}|\mathrm{LOC}|\mathrm{CAT}|\mathrm{SUBJ}\ \boxed{1}\right] \wedge \texttt{member}(\boxed{2},\boxed{1}) \right) \\ \rightarrow \neg(\boxed{2}\mathrm{RM}) \end{array} \right)$$

These two constraints block (3.98). The ungrammaticality of (3.98a) may be also predicted differently. The lexical entry of the verb *bać się* 'to fear' proposed in (3.83) assumes that its NP complement is assigned a semantic role. Since the RM is never assinged a semantic role, §3.4.6, its presence in (3.98a) is incompatible with the lexical specification of the verb. A similar argumentation, however, is not valid for (3.98b–d). As discussed above, the presence of the RM in reflexivized verbs in (3.98b–d) is obligatory (and there is no syntactic complement). Since the RM is required also in ICs, it is only by (3.103) (and (3.104)) that multiple RMs in (3.98b–d) are excluded. Note that these constraints prevent only the multiple occurrence of the RM but they do not force haplology. This is a desirable result since (3.98a) does not have a haplology interpretation (the second RM in (3.98a) cannot take the function of the complement).

As (3.99b–d) indicate, haplology occurs in impersonal constructions, irrespectively of functions of *się*. The lexical rule which provides impersonal forms (cf. §3.4.5) is repeated in (3.105).

(3.105)
$$\begin{bmatrix} \textit{word} \\ \mathrm{SS}|\mathrm{LOC} & \begin{bmatrix} \mathrm{CAT} & \begin{bmatrix} \textit{category} \\ \mathrm{HEAD}|\mathrm{VFORM} & \textit{personal} \\ \mathrm{ARG\text{-}ST} & \langle \mathrm{NP}[\textit{str}]{:}\boxed{1} \rangle \oplus \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \mapsto$$

$$\begin{bmatrix} \mathrm{SS}|\mathrm{LOC} & \begin{bmatrix} \mathrm{CAT} & \begin{bmatrix} \mathrm{HEAD}|\mathrm{VFORM} & \textit{impersonal} \\ \mathrm{VAL} & \left[\mathrm{SUBJ}\ \langle\rangle\right] \\ \mathrm{ARG\text{-}ST} & \langle \boxed{3}\mathrm{NP}[\textit{str} \vee \textit{ldat}]{:}\boxed{1} \rangle \oplus \langle \boxed{4}\mathrm{RM}_{nref} \rangle \oplus \boxed{2} \oplus \boxed{5}\textit{ne\_list}(\textit{adjunt}) \end{bmatrix} \\ \mathrm{CONT} & \begin{bmatrix} \textit{act\_rel} \\ \mathrm{ACTOR}\ \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Since the lexical rule in (3.105) adds the RM to the ARG-ST list of the input verb, several RMs can be present on the ARG-ST list of the resulting verb (but, thanks to (3.103), not on its COMPS list).

[26] The `remove` relation removes an arbitrary element from a list. The relation `list_RM` specifies a possibly empty list which consists of RMs. The definitions of `remove`, which is similar in all relevant respects to the ALE predicate `select/3` used in our implementation, and `shuffle` relations, defined as ALE predicates, are given in §5.2.4.

Our analysis predicts the ambiguity of (3.99b). Similarly to the verb *uderzyć* 'to hit' considered in §3.4.3, we assume that the verb *myje* 'washes' is either a regular transitive verb or a reflexivized verb. The two descriptions of the corresponding forms are given in (3.106).

(3.106) a.

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{myć} \rangle \\ \text{SS|LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{ARG-ST} & \langle \text{NP}[\textit{str}]{:}\boxed{0}, \text{NP}[\textit{str}]{:}\boxed{1} \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{wash-rel} & \\ \text{ACTOR} & \boxed{0} \\ \text{UNDERGOER} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

b.

$$\begin{bmatrix} \textit{word} & \\ \text{PHON} & \langle \text{myć} \rangle \\ \text{SS|LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{ARG-ST} & \langle \text{NP}_{\boxed{1}}[\textit{str}]{:}\boxed{0}, \text{RM}_{\boxed{1}} \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{wash-rel} & \\ \text{ACTOR} & \boxed{0} \\ \text{UNDERGOER} & \boxed{0} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

If the transitive form, (3.106a), is used as the input of the lexical rule in (3.105), the verb does not contain an RM and the RM, $\boxed{4}$, is added to its ARG-ST (and COMPS due to (3.103)). In this case, the syntactic complement of *myć* 'wash' in (3.106a) remains unrealised (the RM cannot substitute the complement and it is *added* to the ARG-ST list of the resulting impersonal verb; hence the RM cannot satisfy requirements of the initial verb). Due to semantics of the verb the object-drop reading results. Example (3.100b) illustrates a corresponding impersonal sentence with all arguments realised (the complement is realised as the pronoun *siebie*).

If the reflexivized verb *myje się* 'washes', (3.106b), is used as the input of the rule, the presence of the RM in the output forbids the repetition of the RM on the COMPS, see (3.103). The verb has no other non-subject arguments, see (3.106b). The two RMs on ARG-ST are represented as a single one on COMPS, (3.103), and haplology results.

Examples (3.99c–d) lack the ambiguity of (3.99b). This difference follows from the fact that both in the reflexiva tantum *przechadzać się* 'to stroll' in (3.99c), and in the middle form *ugotować się* 'to cook' in (3.99d), the RM is the unique realisation of the syntactic requirement of the verb, see §3.1.2. Only one RM is present on COMPS and haplology results.

3.5.2 Verb Clusters

Basics Haplology of *się* in verb clusters (VCs) has been observed in Schenker (1985), Fowler (1993), Rappaport (1998), Kupść (1999d), among others. This phenomenon is illustrated in the (b) examples below. Note that, unlike in simple clauses, the omission of *się* in VCs is optional and the (a) examples are also acceptable, cf. Kupść (1999d, p. 104).[27]

(3.107) a. Jan stara się mniej spóźniać się do pracy.
John tries RM$_{inh}$ less be$_{inf}$ late RM$_{inh}$ to work
'John tries to be less late to work.'

b. Jan stara się mniej spóźniać do pracy.

[27] Judgements vary here and some speakers find (a) worse than (b). For us and many of our informants, however, they are fully acceptable.

(3.108) a. Jan stara się golić się codziennie rano.
John tries RM_{inh} $shave_{inf}$ RM_{refl} every day morning
'John tries to shave himself every morning.'

b. Jan stara się golić codziennie rano.
'John tries to shave (himself) every morning.' (ambiguous)

(3.109) a. Janowi powinno się przyjemnie przechadzać się dziś po parku.
John should RM_{imp} nicely $stroll_{inf}$ RM_{inh} today in park
'John should enjoy strolling in the park today.'

b. Janowi powinno się przyjemnie przechadzać dziś po parku.

(3.110) a. Janowi powinno się łatwiej golić się wieczorem.
John should RM_{inh} easier $shave_{inf}$ RM_{refl} evening
'It should be easier for John to shave himself in the evening.'

b. Janowi powinno się łatwiej golić wieczorem.
'It should be easier for John to shave (himself) in the evening.' (ambiguous)

(3.111) a. Drzwi zdawały się powoli się otwierać.
door seemed RM_{inh} slowly RM_{mid} $open_{inf}$
'The door seemed to be opening slowly.'

b. Drzwi zdawały się powoli otwierać.

The number of reflexive markers that can be omitted is in principle unconstrained. In (3.112), a single *się* serves for 4 items.

(3.112) Po tych lekach powinno mu się zacząć udawać mniej
after these pills should him RM_{imp} $begin_{inf}$ $succeed_{inf}$(-RM_{inh}) less
obawiać spotkać ze znajomymi sprzed wojny.
$fear_{inf}$(-RM_{inh}) $meet_{inf}$(-RM_{refl}) with friends from-before war
'As a result of these pills he should begin to succeed to be less afraid of meeting friends from before the war.'

The above examples show that in VCs haplology (or lack thereof) is insensitive to the function of *się*. The optionality of haplology is related to the optional climbing of *się* (see §3.2.6 and the discussion below). If the governing verb is reflexivized, another RM cannot be syntactically realized on this verb, cf. §3.5.1. We assume that in such a case, the lower RM may remain unrealised, which results in haplology, cf. examples (b) in (3.107)–(3.111) and (3.112). On the other hand, no haplology occurs if the RM does not climb to a reflexivized verb and is realized downstairs, cf. examples (a) in (3.107)–(3.111).

The Syntactic Position of the RM Before presenting the analysis in detail let us briefly (re-)address the position of the RM in VCs. In §3.2.6, we argued (after Rappaport (1998)) that the position of *się* in VCs is the result of syntactic climbing and cannot be reduced to linear order (for a similar conclusion with respect to pronominal clitics in VCs see §2.5). This conclusion has been reached on the basis of examples such as (3.45): in both sentences the RM occurs postverbally and is separated from its semantic governor by a non-clitic; nevertheless only (3.45a) is grammatical.

(3.45) a. Bać zaczniemy się po wyborach.
fear$_{inf}$ will-begin-we RM$_{inh}$ after elections
'We'll begin to fear after the elections.'

b. * Boimy zacząć się wybory.
fear-we begin$_{inf}$ RM$_{inh}$ elections
'We fear to start the elections.'

In §3.2.6, we explained this contrast by the distinct syntactic position of the RM: in (3.45a) the clitic *się* climbs to the verb *zaczniemy* 'we will begin'; hence, the RM immediately follows the verb from which it is realised, in contrast to (3.45b), where the clitic is realised as a local argument of the verb *boimy się* '(we) fear'.

As the examples below indicate, similarly to CC, haplology depends on the syntactic position of the RM. The contrast between (3.113a) and (3.113b) shows that haplology is possible only if the lower RM is omitted (*staram się* '(I) try', *bać się* 'to fear' in (3.113a–b) and *roześmiać się* 'to laugh' in (3.113c) are reflexivized verbs but *przestać* 'to stop' in (3.113a–b) and *zaczynamy* '(we) begin' in (3.113c) are not).

(3.113) a. Staram się przestać bać egzaminów.
try-I RM$_{inh}$ stop$_{inf}$ fear$_{inf}$(-RM$_{inh}$) exams
'I'm trying to stop fearing examinations.' Rappaport (1998)

b. ?* Staram przestać się bać egzaminów.
try-I(-RM$_{inh}$) stop$_{inf}$ RM$_{inh}$ fear$_{inf}$ exams
'I try to stop to be afraid of exams.' (putatively)

c. ? Roześmiać zaczynamy się bać.
laugh$_{inf}$(-RM$_{inh}$) begin-we RM$_{inh}$ fear$_{inf}$(-RM$_{inh}$)
'We begin to fear laughing.'

As in (3.45b), the ungrammaticality of (3.113b) cannot be attributed to the principles ordering the RM in a sentence. In (3.113b) and (3.113c) the RM occurs in the same linear position but only the latter sentence is acceptable. We attribute such contrasts to the distinct syntactic position of the RM in these sentences.

If there are several contiguous reflexivized verbs (*starali się* '(they) tried' and *dostać się* 'to get into' in (3.114) are inherently reflexive verbs), haplology seems to be more permissible: the contrast between (3.114a–b) is weaker than between (3.113a–b). Nevertheless, haplology is fully grammatical only if the lower RM is missing, (3.114a).

(3.114) a. Do mieszkania starali się dostać ludzie.
to flat tried RM$_{inh}$ get$_{inf}$(-RM$_{inh}$) people
'People tried to get into the apartment.' Rappaport (1998)

b. ?? Do mieszkania starali dostać się ludzie.
to flat tried(-RM$_{inh}$) get$_{inf}$ RM$_{inh}$ people
'People tried to get into the apartment.' (putatively)

We take the above examples to indicate that haplology in VCs depends on the syntactic position of the RM: only the lower RM can be omitted.

Let us note finally (after Przepiórkowski (1999, p. 163)) that haplology is insensitive to negation: in (3.115) haplology is equally (optionally) possible across negation.

(3.115) a. Jan stara się nie spóźniać (się) do pracy.
John tries RM$_{inh}$ NM be$_{inf}$ late RM$_{inh}$ to work
'John tries not to be late to work.'

b. Jan stara się nie golić (się) rano.
John tries RM$_{inh}$ NM shave$_{inf}$ RM$_{refl}$ morning
'John tries not to shave in the morning.'

c. Janowi może się nie powieść (się) na egzaminach.
John$_{dat}$ may RM$_{imp}$ NM succeed$_{inf}$ RM$_{inh}$ on exams
'John might not succeed with his exams.'

d. Drzwi zdawały się wcale nie otwierać (się).
door seemed RM$_{inh}$ at-all NM open$_{inf}$ RM$_{mid}$
'The door seemed not to open at all.'

Clitic Climbing In order to account for haplology of the RM in VCs, we mostly rely on the analysis of clitic climbing (CC) presented in §2.5. Similarly to other Polish clitics, the RM may be realised locally, cf. (3.116a) or optionally climb to a governing non-reflexivized verb, cf. (3.116b) (or higher up). (In (3.116), there is only one reflexivized verb, namely *bać się* 'to fear'.)

(3.116) a. Janek zaczyna bać się ciemności.
John begins fear$_{inf}$ RM$_{inh}$ darkness
'John begins to be afraid of darkness.'

b. Janek zaczyna się bać ciemności.

Although negation does not stop haplology, see (3.115), it stops CC, cf. Przepiórkowski (1999, pp. 161ff) and the discussion in §2.5.2. The contrast between (3.117) and (3.118) indicates that climbing of the RM is blocked if a negated verb is present in VCs.

(3.117) a. Próbowałem ogolić się wczoraj.
tried-I shave$_{inf}$ RM$_{refl}$ yesterday
'I tried to shave yesterday.'

b. Ogolić próbowałem się wczoraj.
shave$_{inf}$ tried-I RM$_{refl}$ yesterday
'I tried to shave yesterday.'

(3.118) a. Próbowałem nie ogolić się wczoraj.
tried-I NM shave$_{inf}$ RM$_{refl}$ yesterday
'I tried not to shave yesterday.'

b. ?* Nie ogolić próbowałem się wczoraj.
NM shave$_{inf}$ tried-I RM$_{refl}$ yesterday
'I tried not to shave yesterday.'

c. ? Nie ogolić się próbowałem wczoraj.
NM shave$_{inf}$ RM$_{refl}$ tried-I yesterday
'I tried not to shave yesterday.'

The examples in (3.117) show that, when there is no negation, the lower verb can be preposed leaving the RM behind. The contrast in (3.118a–b) indicates that if the lower verb is negated, this is not possible. However, the result is clearly better if the negated verb is fronted with the clitic, compare (3.118b) and (3.118c). These examples support the conclusion that climbing of the RM (and clitics in general, see §2.5.2) is blocked by the intervening verbal negation.

An apparent counterexample to this conclusion is the example in (3.119).

(3.119)	Próbowałem	się	nie	ogolić	wczoraj.
	tried-I	RM_{refl}	NM	shave_{inf}	yesterday

‘I tried not to shave yesterday.’

In (3.119), it looks like the RM has escaped negation and climbed to a higher non-reflexivized verb. Nevertheless, a different account of such examples is plausible: as discussed in §3.2.6, the linear position of the RM is relatively free and the RM may appear on almost any element preceding the verb it originates from, including phrase internal positions. Therefore, the position of the RM can be attributed to the linear order rather than syntactic climbing (see Przepiórkowski (1999, p. 161) and §2.5.2 for the same explanation of analogous examples with pronominal clitics).

Having presented the data, we are now able to proceed to the analysis. First, we incorporate climbing of the RM into our previous account of CC, and then we provide an analysis of haplology.

An HPSG Account of CC As in Chapter 2, we encode the possibility of CC directly in the lexicon. The partial lexical entry for the verb *zaczyna* ‘begins’ in (3.116) is given in (3.120).

(3.120)
$$\begin{bmatrix} \textit{word} \\ \textsc{phon}\ \langle\text{zaczyna}\rangle \\ \textsc{ss|loc|cat}\ \begin{bmatrix} \textsc{head}\ \textit{verb} \\ \textsc{arg-st}\ \langle \text{NP}[\textit{str}],\ \begin{bmatrix}\textsc{loc|cat}\ \begin{bmatrix}\textsc{head|vform}\ \textit{inf} \\ \textsc{val|comps}\ \boxed{2}\end{bmatrix}\end{bmatrix}\rangle \oplus \boxed{2}\end{bmatrix}\end{bmatrix}$$

The verb inherits the COMPS value of its infinitive complement. Note that $\boxed{2}$ is not further specified and it can be either the empty or a non-empty list. Unrealised arguments of the infinitive (if there are any) are inherited by the verb *zaczyna* ‘begins’. Therefore, if the lower verb requires *się*, the RM becomes an argument of the main verb, cf. (3.116b), unless the RM has been realised on the lower verb, (3.116a).

We encode raising of arguments by adding the COMPS list of the infinitival complement to the ARG-ST list of the main verb, see §2.5.3. This allows the infinitive complement of the verb in (3.120) to be either lexical or phrasal. As discussed, e.g., in Dziwirek (1994), Przepiórkowski and Kupść (1997c), Witkoś (1998), Kupść (1999d,2000), Przepiórkowski (1999,2000), ‘clause union’ effects (binding, case assignment, negative concord, word order, clitic climbing) are optional in VCs, which disallows us to assign a unique syntactic structure to VCs (see §2.5.2 for some discussion). Since the RM is a syntactic item, see §3.2, it is present on the COMPS list, cf. §3.4, and can be raised on a par with other arguments.

One of the syntactic analyses of (3.116b) is given in fig. 3.4 (assuming a hierarchical structure of VCs and raising of the RM). (We use **S**, **C** and **H** to indicate the SUBJ-DTR, COMP-DTRS and the HEAD-DTR, respectively.)

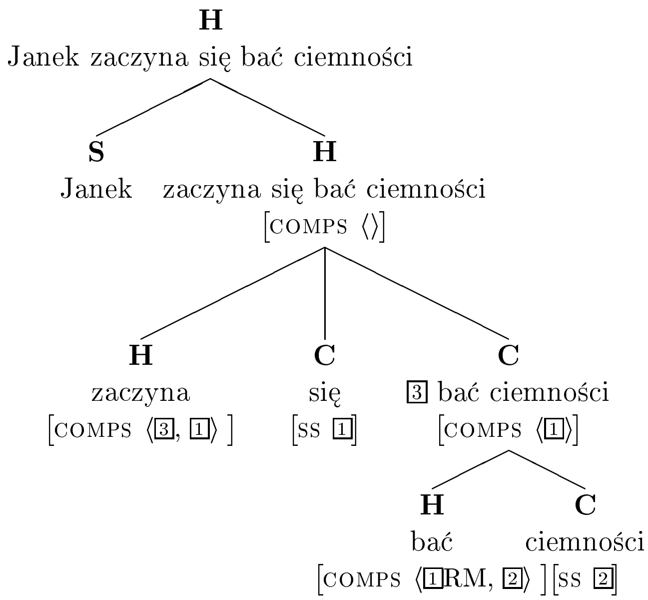


Figure 3.4: A syntactic analysis of (3.116b)

The possibility of clitic climbing is encoded directly in lexical entries of the verbs which form VCs. This allows us to prevent climbing of the RM in other verbal environments, e.g., CC out of a subordinate clause is impossible, see (3.40) repeated below.

(3.40) a. Jan powiedział [*się], że [się] zatrzyma u znajomych.
John said RM_{inh} COMP will.stop-he at acquaintances
'John said that he would stay at his friends'.'

b. Jan powiedział [*się], że [się] szklanka stłukła.
John said RM_{mid} COMP glass broke
'John said that the glass broke.'

c. Jan powiedział [*się], że [się] dobrze czyta tę książkę.
John said RM_{imp} COMP well reads this book
'John said that one read this book easily.'

d. Jan powiedział [*się], że [się] nie umyje.
John said RM_{refl} COMP NM will.wash-he
'John said that he would not wash.'

As argued above and in §2.5.2, CC out of a negated verbal phrase is impossible, see (3.118). Since the RM is present on the ARG-ST list, the principle adopted in §2.5.3 from Przepiórkowski (1999) will also prevent raising of the RM across negation; (2.97) is repeated below for convenience.

(2.97) $$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{NEG} \ + \end{bmatrix} \\ \text{ARG-ST} & \langle \boxed{0} \rangle \oplus \boxed{1} \end{bmatrix} \rightarrow \boxed{1}\ \text{list}(\text{XP}^{-})$$

According to (2.97), no non-subject argument of a negated verb can be raised to a higher verb (XP^{-} abbreviates an XP element of the ARG-ST list marked as [RAISED $-$], i.e., an element which does not appear on any higher ARG-ST). Therefore, if the infinitival complement of the main verb is negated, as in (3.118), the principle (2.97) prevents its complements from raising to the main verb. Since the RM is an element of the ARG-ST list its raising is blocked as well. The ungrammaticality of (3.118b) is correctly predicted and only (3.118a,c) may result.

The account of linear positions of pronominal clitics, presented in §2.5.4, holds for the RM: *się* is a clitic present on the COMPS list of the verb. In particular, the rightward linear position of the RM with respect to the verb *bać się* 'to fear' in (3.45a) is correctly captured by the LP constraints proposed in §2.5.4 (see (2.100) repeated below).

(3.45a) Bać zaczniemy się po wyborach.
fear$_{inf}$ will-begin-we RM$_{inh}$ after elections

'We'll begin to fear after the elections.'

(2.100) a. If the DOM list $\boxed{4}$ contains *dom_ objs* $\boxed{0}$, $\boxed{1}$, such that $\boxed{0}$ represents a verb and $\boxed{1}$ represents a non-raised argument clitic of this verb and $\boxed{0}$ precedes $\boxed{1}$ then

either $\boxed{0}$ immediately precedes $\boxed{1}$ or

there is a *dom_ obj* $\boxed{2}$, which represents a non-raised argument clitic of the same verb, such that $\boxed{2}$ immediately precedes $\boxed{1}$.

b.
$$\mathtt{lp}_4(\boxed{4}) \overset{\forall}{\Longleftarrow} \forall\boxed{0}\ \forall\boxed{1}\ \forall\boxed{5}
\left(\begin{array}{l}
\exists\boxed{3}\ \exists\boxed{7} \\
\left(\begin{array}{l}
\mathtt{member}\left(\boxed{0}\begin{bmatrix} \textit{dom_ obj} \\ \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{VAL}|\text{COMPS } \boxed{5} \end{bmatrix} \end{bmatrix}, \boxed{4}\right) \wedge \mathtt{member}\left(\boxed{1}\begin{bmatrix} \textit{dom_ obj} \\ \text{SS } \boxed{3} \end{bmatrix}, \boxed{4}\right) \\
\wedge\ \mathtt{member}\left(\boxed{7}\begin{bmatrix} \textit{arg} \\ \text{ARG } \boxed{3}\textit{clitic} \\ \text{RAISE } - \end{bmatrix}, \boxed{5}\right) \wedge \mathtt{prec}(\boxed{0},\boxed{1},\boxed{4})
\end{array}\right) \\
\rightarrow \exists\boxed{2}\ \exists\boxed{6}\ \exists\boxed{8}
\left(\begin{array}{l}
\mathtt{imprec}(\boxed{0},\boxed{1},\boxed{4}) \vee \\
\mathtt{imprec}\left(\boxed{2}\begin{bmatrix} \textit{dom_ obj} \\ \text{SS } \boxed{6} \end{bmatrix}, \boxed{1},\boxed{4}\right) \wedge \mathtt{member}\left(\boxed{8}\begin{bmatrix} \textit{arg} \\ \text{ARG } \boxed{6}\textit{clitic} \\ \text{RAISE } - \end{bmatrix}, \boxed{5}\right)
\end{array}\right)
\end{array}\right)$$

The domain representation of the sentence in (3.45a), consistent with (2.100), is given in (3.121).

(3.121)
$$\begin{bmatrix} \text{SS VC} \\ \text{DOM } \left\langle \begin{bmatrix} \textit{dom\_ obj} \\ \text{PHON } \langle \text{bać} \rangle \\ \text{SLCV}|\text{COMPS } \left\langle \begin{bmatrix} \textit{arg} \\ \text{ARG } \boxed{1} \\ \text{RAISE } + \end{bmatrix} \right\rangle \end{bmatrix}, \begin{bmatrix} \textit{dom\_ obj} \\ \text{PHON } \langle \text{zaczniemy} \rangle \\ \text{SLCV}|\text{COMPS } \left\langle \begin{bmatrix} \textit{arg} \\ \text{ARG } \boxed{1} \\ \text{RAISE } - \end{bmatrix} \right\rangle \end{bmatrix}, \begin{bmatrix} \textit{dom\_ obj} \\ \text{PHON } \langle \text{się} \rangle \\ \text{SS } \boxed{1}\textit{clitic} \end{bmatrix}, \ldots \right\rangle \end{bmatrix}$$

The RM originates from the infinitive *bać się* 'to fear' but it is raised to the main non-reflexivized verb *zaczniemy* '(we) will begin'. The LP constraint in (2.100) orders the clitic with respect to the verb from which it cannot be raised higher up (*zaczniemy* '(we) will begin'). The linear position of *się* with respect to the verb *bać się* 'to fear' is licensed (see §2.5.4 for a similar example with a pronominal clitic).

An HPSG Account of Haplology As discussed above, haplology is fully acceptable only if the lower RM is omitted, cf. (3.113a). No haplology occurs if another *się* is realised on a lower verb, (3.122) (see also (3.107)–(3.111)).

(3.122) a.

Staram	się	przestać	się	bać	egzaminów.
try-I	RM_{inh}	stop_{inf}	RM_{inh}	fear_{inf}(-RM_{inh})	exams

'I'm trying to stop fearing examinations.'

b.

Staram	się	przestać	bać	się	egzaminów.
try-I	RM_{inh}	stop_{inf}	fear_{inf}(-RM_{inh})	RM_{inh}	exams

'I'm trying to stop fearing examinations.'

In (3.122), there are two reflexivized verbs: *starać się* 'to try' and *bać się* 'to fear'. In (3.122a), the position of the second RM is the result of the syntactic climbing of *się* associated with *bać się*. In (3.122b), no climbing takes place and the second RM is realised locally.

As proposed in §3.4, we assume that the syntactic requirement for the RM is specified in the lexicon, in subcategorization frames of reflexivized verbs. A partial lexical entry for the inherently reflexive verb *staram się* '(I) try' in (3.122) is given in (3.123).

$$(3.123)\quad \begin{bmatrix} \textit{word} \\ \text{PHON } \langle \text{staram} \rangle \\ \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD } \textit{verb} \\ \text{ARG-ST } \langle \text{NP}[\textit{str}], \text{RM}, \begin{bmatrix} \text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD}|\text{VFORM } \textit{inf} \\ \text{VAL}|\text{COMPS } \boxed{1} \end{bmatrix} \end{bmatrix} \rangle \oplus \boxed{1} \end{bmatrix} \end{bmatrix}$$

The principle in (3.103) on p. 135 ensures that if the RM present on COMPS of a lower verb is raised to the ARG-ST of the reflexivized verb, only one RM will appear on the COMPS list of the latter verb.

Note that our analysis of haplology differs from that of CC: in contrast to other clitics, the RM is only raised to the ARG-ST of a reflexivized verb but it is absent from its COMPS list (but not from the COMPS of a non-reflexivized verb, see (3.120)). Haplology results if the RM associated with the infinitival complement remains unrealised (the Immediate Dominance SCHEMA2' introduced in §2.5.3 licenses phrases with an unsaturated COMPS list). On the other hand, the lower RM (if any) may be realised on the lower verb and a non-haplology sentence is obtained, see (3.122).

A schematic analysis of haplology in (3.113a) (assuming the hierarchical structure of VC) is presented in fig. 3.5.

In the phrase *bać egzaminów* 'to fear exams', the required RM, $\boxed{4}$, is unrealised. Assuming that the non-reflexivized verb *przestać* 'to stop' has the lexical entry analogous to (3.120), it inherits all unrealised complements of its infinitive complement. In fig. 3.5, this infinitival complement is the phrase $\boxed{2}$ whose COMPS contains one element, $\boxed{4}$. This element is inherited by *przestać* 'to stop' and added to its arguments, see (3.120). The RM is missing in the infinitival phrase *przestać bać egzaminów* 'to stop fear exams' ($\boxed{3}$) and remains unrealised. The reflexivized verb *starać się* 'to try' already contains the RM ($\boxed{1}$) on its COMPS (ARG-ST) list, cf. (3.123). The lower *się* $\boxed{4}$ is inherited on the ARG-ST list but it is absent on COMPS (there are two RMs on ARG-ST of *stara* 'tries', see (3.103)). The lower RM remains unrealized which results in haplology. Nothing, however, prevents the lower RM from being realised locally on *bać* 'to fear' or on the intermediate non-reflexivized verb *przestać* 'to stop'. Thus, the lack of haplology in (3.122) can be also accounted for.

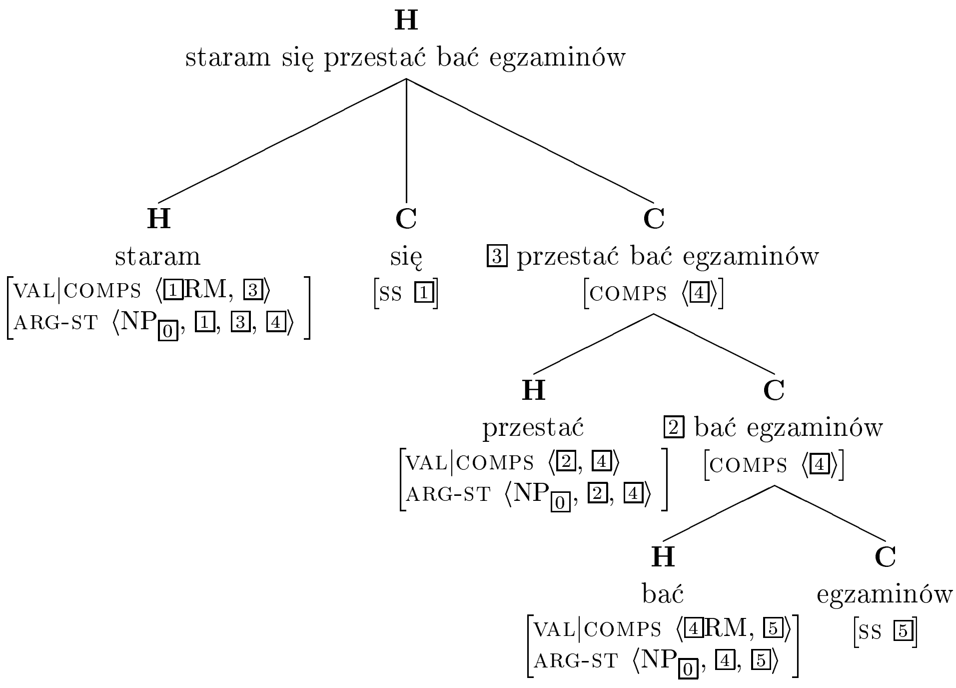


Figure 3.5: A syntactic structure for (3.113a)

In order to capture the different behaviour of haplology and CC with respect to verbal negation, cf. (3.115) and (3.118), we have to slightly reformulate how values of the RAISE attribute are specified. The principle in (2.97), repeated on p. 142, prevents non-subject arguments of a negated verb from occurring on any higher ARG-ST: an argument is [RAISE +] if and only if it is present on an immediately higher ARG-ST, see Przepiórkowski (1999, p. 94). If this definition is adopted, haplology across negation will be impossible. If an RM is unrealized on the negated infinitival complement (thus present on its COMPS), it will be inherited by the governing reflexivized verb, cf. (3.123). The RM will be present on ARG-ST of this verb, i.e., marked [RAISE +], contrary to (2.97). Note, however, that although the RM is raised to ARG-ST, it will be absent on the COMPS list of the reflexivized verb, cf. (3.103); so it is not raised in syntax.

We extend non-raising to such cases and assume that an argument of ARG-ST of some *sign* is [RAISE +] if and only if it belongs to the immediately higher ARG-ST **and** is present on the COMPS list related to this ARG-ST.

Let us check whether this modification provides correct predictions. CC out of negated phrases should be blocked, see (3.118b). The principle (2.97) requires complements of *nie golić* 'not to shave', i.e., the RM in (3.118b), to be absent from COMPS of the higher verb *próbowałem* '(I) tried' (the RM must be marked [RAISE −]). Simultaneously, however, the lexical description of the non-reflexivized verb *próbowałem* '(I) tried' (which is analogous to (3.120)), indicates that complements (COMPS) of the infinitive complement are obligatorily inherited and added to ARG-ST. The list of inherited arguments is also always added to COMPS since there is no RM on ARG-ST of *próbowałem*, cf. (3.120). These two requirements are not contradictory only if all arguments of the negated verb are realised locally, i.e., if the added COMPS list is empty. Therefore, CC out of a negated phrase is still blocked as desired.

In the case of haplology in (3.115a), the reflexivized verb *stara się* 'tries' adds the unrealized RM of the negated phrase to its ARG-ST. Since ARG-ST of *stara się* already contains an RM, see (3.123), the inherited RM is not present on the COMPS list: due to (3.103), at most one RM can occur on COMPS. Hence, the lower RM is marked [RAISE −] as required by (2.97) but also remains unrealised and haplology results.

Therefore, although climbing of the RM is effectively blocked by (2.97), haplology is not: the RM may remain unrealised.

Haplology is encoded lexically, which correctly excludes it from other verbal environments, e.g., the presence of a complementizer blocks haplology, see (3.124) (see also Rappaport (1998, p. 467), Derwojedowa (1998, p. 105), Przepiórkowski (1999, p. 158)).

(3.124) a. Boję się, żeby *(się) nie spóźnić.
fear-I RM_{inh} COMP RM_{inh} not be_{inf}-late
'I fear not to be late.'

b. Maria stara się, żeby Jan codziennie *(się) golił.
Mary tries RM_{inh} COMP John everyday RM_{rfl} shaved
'Mary tries to remind John to shave everyday.'

c. Mówiło się, że miejscowa ludność opowie *(się) za niepodległością.
told RM_{imp} COMP local people will.support RM_{inh} for independence
'It was said that local people would vote for the independence.'

d. Uważa się, że ten materiał *(się) dobrze prasuje.
claims RM_{imp} COMP this cloth RM_{mid} well irons
'It is said that this cloth irons well.'

As before, the ambiguity of (3.108b) and (3.110b) follows from the double interpretation of the verb *golić* 'to shave'. Similarly to *uderzyć* 'to hit' and *myć* 'to wash', discussed previously, *golić* 'to shave' can be either a 'regular' transitive verb, with the lexical description analogous to (3.106a), or a semantically reflexivized verb, (3.106b). If the former is used, the NP complement of the transitive *golić* must be inherited by the main verb, see (3.123). The lexical rule for impersonals in (3.105) adds the RM to the ARG-ST (and COMPS) list of the resulting transitive impersonal form but the NP complement of the 'original' verb remains unrealised: there is no direct object in (3.108b) and (3.110b), and an 'object-drop' reading results. If the reflexivized verb *golić się* 'to shave' appears in (3.108b) and (3.110b), only one RM appears on the COMPS list of the governing reflexivized verb *staram się* 'to try'. One RM remains unrealised and the haplology interpretation is obtained. On the other hand, the lower RM can be realised locally as well and (3.108a) or (3.110a) without haplology are possible.

3.5.3 Coordination and Ellipsis

In Rappaport (1998), coordinated phrases with the RM omitted in the second conjunct are treated as instances of haplology. In this subsection, we re-examine coordination data, cf. §3.2.3, focusing this time on the interaction of different types of *się*. We do not provide an HPSG analysis of these data since coordination still awaits a satisfactory account in HPSG.

In §3.2.3, as in Rappaport (1998), coordination of verbs that require the same type of *się* was considered. If the RM has the same function in both conjuncts, its repetition is optional, cf. (3.27)–(3.28), repeated below. (See Derwojedowa (1998, p. 100) for similar data.)

(3.27) a. Jan umył się i ogolił.
John washed RM$_{refl}$ and shaved
'John washed (himself) and shaved (himself) quickly.'

b. Siedziało się i gadało do późna.
sat RM$_{imp}$ and chatted till late
'One was sitting and chatting till late.'

c. Dobrze pierze się i prasuje ten materiał.
well washes RM$_{mid}$ and irons this cloth
'This cloth washes and irons well.'

d. Jan potykał się i ślizgał na zamarzniętej drodze.
John stumbled RM$_{inh}$ and slid on frozen road
'John was stumbling and sliding on the frozen road.'

(3.28) a. Jan umył się i ogolił się.
John washed RM$_{refl}$ and shaved RM$_{refl}$
'John washed (himself) and quickly shaved (himself).'

b. Siedziało się i gadało się do późna.
sat RM$_{imp}$ and chatted RM$_{imp}$ till late
'One was sitting and chatting till late.'

c. Dobrze pierze się i prasuje się ten materiał.
well washes RM$_{mid}$ and irons RM$_{mid}$ this cloth
'This cloth washes and irons well.'

d. Jan potykał się i ślizgał się na zamarzniętej drodze.
John stumbled RM$_{inh}$ and slid RM$_{inh}$ on frozen road
'John was stumbling and sliding on the frozen road.'

If the data in (3.27) are treated as haplology, (3.28) illustrates that, as in VCs, haplology in coordinated structures is optional. As discussed in §3.2.3, the omission of the RM in (3.27) can be interpreted differently, namely as syntactic ellipsis.[28] As we will see shortly, these two mechanisms are indeed necessary to fully account for the omission of the RM in coordination.

If the function of *się* in both conjuncts is different, the omission of the RM does not seem to be equally possible, cf. (3.126b) and (3.127b).

(3.125) a. Janek ubierał się i uśmiechał się do siebie.
John dressed RM$_{refl}$ and smiled RM$_{inh}$ to self
'John was getting dressed and was smiling to himself.'

b. Janek ubierał się i uśmiechał do siebie.

(3.126) a. Jan ubrał się ciepło i przyjemnie spacerowało mu się po parku.
John dresses RM$_{refl}$ warmly and nicely strolled him$_{cl}$ RM$_{imp}$ in park
'John got warmly dressed and enjoyed strolling in the park.'

b. * Jan ubrał się ciepło i przyjemnie spacerowało mu po parku.

[28]In fact, one more possibility is feasible for (3.27): coordination of verbs which share a syntactic element (so-called Right Node Raising). Since this option is not available in the case of coordination of sentences, see (3.131), we do not consider this alternative here.

(3.127) a. Dziecko potykało się i trudno mu się wchodziło po schodach.
child stumble RM_{inh} and difficult him_{cl} RM_{imp} climb after stairs
'The child stumbled and had difficulty in climbing the stairs.'

b. * Dziecko potykało się i trudno mu wchodziło po schodach.

(3.128) a. Gałąź się złamała, ale szybko zazieleniła się od nowa.
branch RM_{mid} broke but quickly greened RM_{inh} from new
'The branch broke but it greened quickly again.'

b. Gałąź się złamała, ale szybko zazieleniła od nowa.

If impersonal constructions are treated as independent clauses, the ungrammaticality of (3.126b) and (3.127b) can be explained by the fact that haplology cannot cross a simple clause boundary, which we have seen previously in (3.124). On the other hand, the grammaticality of (3.125) and (3.128) is predicted if these examples are considered a coordination of VPs rather than full clauses.[29]

Such an account seems to be further supported by the impossibility of haplology when different types of *się* occur in coordinated sentences with explicitly distinct subjects, cf. (3.129b).

(3.129) a. Jan się elegancko ubrał i Maria uśmiechnęła się, gdy go zobaczyła.
John RM_{refl} smartly dressed and Mary smiled RM_{inh} when him saw-she
'John got smartly dressed and Mary smiled when she saw him.'

b. * Jan się elegancko ubrał i Maria uśmiechnęła, gdy go zobaczyła.

Derwojedowa (1998, p. 103) argues that haplology is impossible only out of a subordinate-like clause and provides (3.130) as grammatical.

(3.130) ?? Jan się elegancko ubrał, a Maria uśmiechnęła, gdy go zobaczyła.
John RM_{refl} smartly dressed and Mary smiled when him_{cl} saw-she
'John got smartly dressed and Mary smiled when she saw him.'

Derwojedowa (1998, p. 103) claims that the conjunction *i* 'and' in (3.129b) is used as the conjunction *więc* 'so, thus' which introduces a subordinate-like clause, see Świdziński (1992). Thus, (3.129b) is ruled out by the impossibility of haplology across a subordinate clause boundary, see (3.124). Apparently, the conjunction *a* 'and' in (3.130) does not serve as a subordinate-like conjunction and, hence, haplology is possible. Although we do not find (3.130) fully acceptable (see our judgement), we do not deny that additional factors might relax the omission of different RMs in coordinated sentences. As this requires a detailed examination of properties of various conjunctions (and also an analysis of coordination), which goes far beyond the scope of this study, we do not consider the issue any further.

The examples in (3.131) seem to be counterexamples to the conclusion that a simple clause boundary constitutes a barrier for haplology: these sentences are fully acceptable although *się* appears only in the first clause.

[29]Due to various 'clause union' effects, we do not consider bare infinitive complements in VCs to be independent clauses.

(3.131) a. Jan się umył, a Piotr ogolił.
John RM$_{refl}$ washed and Peter shaved(-RM$_{refl}$)
'John washed (himself) and Peter shaved (himself).

b. Janek się potknął, a Tomek przewrócił.
John RM$_{inh}$ stumble and Peter fell-down(-RM$_{inh}$)
'John stumbled and Peter fell down.'

c. Janowi dobrze się pisze książki, a Marii przyjemnie je czyta.
John well RM$_{imp}$ writes books and Mary nicely them reads(-RM$_{imp}$)
'John enjoys writing books and Mary enjoys reading them.'

Note, however, that unlike in (3.126)–(3.127) and (3.129), the same type of the RM is omitted in (3.131). Thus, (3.131) do not necessarily contradict clause boundedness of haplology but may be explained by a different mechanism, namely by syntactic ellipsis.

The examples in (3.31) (repeated below) do not conform to the view that haplology is a clause bound phenomenon either. Not only is the RM omitted in a different clause but this clause constitutes an independent utterance.

(3.31) a. D: A zajmiesz się nim do tego czasu?
and will-take-care-you RM$_{inh}$ him till this time

A: Zajmę.
will-take-care-I

D: 'Will you take care of him till this time?' A: 'I will.'

b. A: Umyłeś się już? B: Umyłem.
washes-you RM$_{refl}$ already washed-I

A: 'Have you washed already?' B: 'I have.'

c. A: Czy zupa się już ugotowała? B: ?Ugotowała.
Q soup RM$_{mid}$ already cooked cooked

A: 'Has the soup cooked yet?' B: 'It has.'

d. A: Spało się, co? B: ?Spało.
slept RM$_{imp}$ what slept

A: 'One was sleeping, weren't you?' B: 'I was.'

Again, as in (3.131), the omitted RM has the same function in both clauses, unlike in (3.126)–(3.127) and (3.129). Therefore, we do not consider such data to be counterexamples to our previous conclusion and take syntactic ellipsis to be responsible for (3.31).

To conclude, we treat haplology and syntactic ellipsis as two independent phenomena which may give the same result in coordination of Vs or VPs. Haplology is insensitive to the type of *się* but it is possible only up to a VP level. On the other hand, syntactic ellipsis always affects RMs of the same type but it is possible also at the sentence level and across a clause boundary. (A similar conclusion that there are two distinct mechanisms responsible for the omission of *się* in coordinated VPs and Ss has been also suggested by Stefan Dyła (p.c.).)

3.5.4 Gerunds

As discussed in §3.2.6, Polish deverbal nouns (i.e., gerunds) and participles may occur with the RM, see (3.41) and (3.42) repeated below.

(3.41) a. Autorzy opisują [*się] zachowanie [się] premiera.
authors describe RM$_{inh}$ behaviour$_{gerund}$ RM$_{inh}$ premier
'The authors describe the behaviour of the premier.'

b. I żadnych głupich myśli o [*się] wieszaniu [się] [...].
and no stupid thoughts about RM$_{refl}$ hanging$_{gerund}$ RM$_{refl}$
'And no stupid ideas about hanging yourself.'

c. Wywołują one stałe [*się] zawężanie [się] wymiarów.
cause they constant RM$_{mid}$ narrowing$_{gerund}$ RM$_{mid}$ dimensions
'They cause constant narrowing of dimensions.'

(3.42) a. Studentki chichotały [*się] co chwila [się] trącając [się] co chwila.
students giggled RM$_{recp}$ every moment RM$_{recp}$ nudging$_{participle}$ RM$_{recp}$ every moment
'Students giggled nudging each other all the time.'

b. Głośno zaprotestował [*się] całkowicie [się] nie zgadzając [się] na takie traktowanie.
loudly protested-he RM$_{inh}$ absolutely RM$_{inh}$ NM agreeing$_{participle}$ RM$_{inh}$ for this treatment
'He protested aloud against such a treatment.'

c. Talerz zrobił dużo hałasu [*się] tłukąc [się].
plate made much noise RM$_{mid}$ breaking$_{participle}$ RM$_{mid}$
'The plate made lots of noise when it was breaking.'

The above examples illustrate that the RM (just like other clitics) cannot climb out of a gerund or a participle phrase (the former property of the RM has been observed in Rappaport (1998, p. 467) and Przepiórkowski (1999, p. 160), among others).

Although the judgements are much less clear than one would wish, it seems that haplology is (optionally) possible also across a gerund phrase boundary, cf. (3.132), see Przepiórkowski (1999, p. 161), but not across a participle phrase, cf. (3.133).

(3.132) a. Bałem się tego ciągłego wylegiwania ?(się) w łóżku.
feared-I RM$_{inh}$ this constant lolling$_{gerund}$ about RM$_{inh}$ in bed
'I was afraid of this constant lolling about in bed.'

b. Spodziewał się potwierdzenia ?(się) zarzutów.
expected-he RM$_{inh}$ confirmation$_{gerund}$ RM$_{mid}$ of accusations
'He was expecting the confirmation of accusations.'

c. Powszechnie oczekuje się polepszenia (się) sytuacji.
commonly awaits RM$_{imp}$ improvement$_{gerund}$ RM$_{mid}$ of situation
'It has been commonly awaited that the situation will improve.'

d. Zachęca się mieszkańców do wyprowadzenia ?(się) z tego domu.
encourages RM$_{imp}$ inhabitants to moving$_{gerund}$-out RM$_{refl}$ from this house
'The inhabitants are encouraged to move out of this house.'

(3.133) a. Studentki śmiały się trącając *(się) co chwila.
students laughed RM_{inh} nudging$_{participle}$ RM_{recp} every moment
'Students laughed nudging each other all the time.'

b. Rozgniewał się na Marysię nie zgadzając *(się) na takie traktowanie.
got-he angry RM_{inh} on Mary NM agreeing$_{participle}$ RM_{inh} for this treatment
'He got angry with Mary not agreeing for such a treatment.'

c. Talerz rozbił się na drobne kawałki tłukąc *(się).
plate broke RM_{mid} on small pieces breaking$_{participle}$ RM_{mid}
'The plate broke into pieces.'

As observed, e.g., in Safarewiczowa (1954), many nouns derived from reflexivized verbs may optionally drop the RM. Nominals in (3.132b–d) have such double forms (although their meanings may differ), i.e., a gerund with or without the RM is possible: *potwierdzenie się : potwierdzenie* 'confirmation', *polepszenie się : polepszenie* 'improvement' and *wyprowadzenie się : wyprowadzenie* 'moving out'. Hence, it could be suggested that the lack of the RM in (3.132b–d) has nothing to do with haplology but follows from the double nature of these nominals: the RM is optionally realised in gerund phrases. Safarewiczowa (1954, p. 341) notes, however, that in the case of nominals derived from inherent reflexives, the RM is obligatory. The verbal nominal in (3.132a) is related to the inherent reflexive verb *wylegiwać się* 'to loll about' and, thus, the omission of the RM is not independently possible, see (3.134).

(3.134) ?* Polubił długie wylegiwanie w łóżku.
liked-he long lolling$_{gerund}$ about in bed
'He liked long lolling about in bed.'

Therefore, if one accepts the omission of the RM in examples such as (3.132a), it should be attributed to haplology.

The RM may be missing from the gerund phrase which is an immediate complement of a reflexivized verb, as in (3.132a–c), but apparently also if the gerund phrase is deeper embedded, see (3.132d) and (3.135) ((3.135a) is taken from Przepiórkowski (1999, p. 163)).

(3.135) a. Przyzwyczaiłem się do wylegiwania ?(się) w łóżku.
got-I used RM_{inh} to lolling$_{gerund}$ about RM_{inh} in bed
'I got used to this lolling about in bed.'

b. Bał się częstych myśli o wylegiwaniu ??(się) w łóżku.
feared-he RM_{inh} often thoughts about lolling$_{gerund}$ about RM_{inh} in bed
'He was afraid of persistent thoughts of lolling about in bed.'

Since the presence of the RM on the gerund *wylegiwanie się* 'lolling about' in (3.135) is otherwise obligatory, see (3.134), the examples above illustrate optional haplology across NP and PP barriers. As in VCs, it is not just any RM that can be omitted: the higher RM cannot be missing, see (3.136).

(3.136) a. * Bałem tego ciągłego wylegiwania się w łóżku.
feared-I this constant lolling$_{gerund}$ about RM_{inh} in bed

‘I was afraid of of this constant lolling about in bed.’ (putatively)

b. * Przyzwyczaiłem do ciągłego wylegiwania się w łóżku.
got-I used to constant lolling$_{gerund}$ about RM$_{inh}$ in bed
‘I go used to constant lolling about in bed.’ (putatively)

As stated previously, haplology across gerund phrases is less acceptable than in VCs. Nevertheless if one wishes to incorporate also the gerund data, we provide an account of haplology which relies on the intuition that ‘the presence of a higher RM licenses the omission of a lower RM’.

In order to express the configurational relations of command and domination in HPSG terms we follow definitions proposed for Polish in Marciniak (1999). The definition of *local s-command* introduced in Marciniak (1999, p. 142) is given as follows:

(3.137) Let Y and Z be *synsem* objects, where Y is *referential*. Then Y *locally s-commands* Z in case either:

i. Y and Z belong to the same ARG-ST list and Y is the first element of this ARG-ST list; or
ii. Y locally s-commands some X and Z belongs to the ARG-ST list of X; or
iii. Y locally s-commands a transparent X and X is the head projection of Z (i.e., the HEAD values of X and Z are token-identical).

The class of *transparent synsem*s specified in Marciniak (1999, p. 141), comprises *synsem* objects of noun and preposition phrases but also bare (complementizerless) infinitival phrases. Thus, clause iii. of (3.137), could be rephrased as ‘Y locally s-commands an NP, a PP or a VP[inf] which is the head projection of Z.’ Clause i. of (3.137) has been introduced specifically for Polish binding which is subject oriented. In our discussion of haplology, we will not refer to the first element of ARG-ST but rather to the RM. We define the relation of *local r-command* (local reflexive marker command) based on (3.137):

(3.138) Let Y and Z be distinct *synsem* objects. Then Y *locally r-commands* Z in case either:

i. Y and Z belong to the same ARG-ST list and Y is an RM; or
ii. Y locally r-commands some X and Z belongs to the ARG-ST list of X; or
iii. Y locally r-commands a transparent X and X is the head projection of Z (i.e., the HEAD values of X and Z are token-identical).

As discussed in §3.4.1, we assume that the RM may be referential or not. Since haplology is possible irrespectively of the type of the RM, in (3.138) we drop the requirement of the commander’s referentiality present in (3.137).

Most relevant for our present discussion are clauses ii. and iii. of (3.138). Clause ii. allows us to locally r-command arguments of a lexical argument whereas clause iii. expresses local r-command of arguments of a phrasal argument. As we will see shortly, this is exactly what is necessary to account for haplology in gerund phrases discussed above.

Namely, we assume that in an unembedded sign (i.e., corresponding to an utterance) an RM $\boxed{1}$ can remain unrealised (it is present on COMPS $\boxed{2}$) in a maximal head projection, $\boxed{3}$, of some lexical head, $\boxed{7}$, with [HEAD $\boxed{0}$], only if $\boxed{1}$ is locally r-commanded on the ARG-ST list, $\boxed{6}$, of the lexical head, see (3.139).[30]

(3.139) *unembedded-sign* $\rightarrow$

$$\forall \boxed{0} \forall \boxed{1} \forall \boxed{2} \forall \boxed{3} \forall \boxed{6} \forall \boxed{7}$$

$$\left(\begin{array}{l} \texttt{max\_head\_proj}(\boxed{3}\begin{bmatrix} \textit{sign} \\ \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \textit{category} \\ \text{HEAD } \boxed{0} \\ \text{VAL}|\text{COMPS } \boxed{2} \end{bmatrix} \end{bmatrix}, \boxed{7}\begin{bmatrix} \textit{word} \\ \text{SS}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD } \boxed{0} \\ \text{ARG-ST } \boxed{6} \end{bmatrix} \end{bmatrix}) \\ \wedge\ \texttt{member}(\boxed{1}\text{RM}, \boxed{2}) \rightarrow \\ \hfill \texttt{locally\_r-commanded}(\boxed{1}, \boxed{6}) \end{array}\right)$$

Due to the argument composition analysis presented in §2.5, in verb clusters not only the RM but also other arguments can be unrealised. Hence, $\boxed{2}$ can contain several elements but always at most one RM, see (3.103). For NPs an PPs, we assume that the unique complement which can remain unrealised is the RM, see (3.140).

$$(3.140) \quad \forall \boxed{2} \left(\begin{array}{l} \begin{bmatrix} \textit{phrase} \\ \text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix} \text{HEAD } \textit{noun} \vee \textit{prep} \\ \text{VAL}|\text{COMPS } \boxed{2}\textit{ne\_list} \end{bmatrix} \end{bmatrix} \\ \rightarrow \boxed{2}\langle \text{RM} \rangle \end{array} \right)$$

The principle in (3.139) does not force us to drop the RM but it specifies the condition when it is possible. Therefore, the optionality of haplology in gerund phrases is predicted. If the RM is not subcategorized for by a gerund, the COMPS list is empty.

Note that (3.139) is a general principle which licenses the non-realisation of any locally r-commanded RM. The lack of haplology across participle phrases, see (3.133), is correctly predicted since such phrases are not arguments of a verb, i.e., they are not members of the verb's ARG-ST list and they are immune to (3.138). Haplology is also impossible across subordinate complement phrases introduced by a complementizer, see (3.124). This fact is correctly accounted for since these phrases do not have transparent *synsem*s, unlike bare VP[inf], and constitute a barrier for local r-command.

Let us illustrate the analysis with several examples, see (3.132a).

(3.132a) Bałem się tego ciągłego wylegiwania ?(się) w łóżku.
feared-I RM_{inh} this constant lolling_{gerund} about RM_{inh} in bed
'I was afraid of this constant lolling about in bed.'

(3.141) a. bałem: [ARG-ST ⟨NP[*str*], $\boxed{1}$RM, NP[*lgen*]⟩]
b. wylegiwania: [ARG-ST ⟨NP, $\boxed{2}$RM, PP[*w*]⟩]

There are two reflexivized elements in (3.132a): the verb *bałem się* '(I) feared' and the gerund *wylegiwania się* 'lolling about'. Their ARG-ST lists are given in (3.141). We follow Marciniak (1999) and Przepiórkowski (1999) and assume that gerunds have a subject: the first element of the ARG-ST list in (3.141b). The NP[*lgen*] argument in (3.141a) is the gerund complement

[30] To be fully formal, we have to ensure that ARG-ST appears only on *word*s, see fn. 7. Additionally, we assume that HEAD values are identical only if they are components of objects which belong to the same head projection.

of the verb, i.e., *wylegiwania (się) w łóżku* 'lolling about in bed'. The RM $\boxed{1}$ in (3.141a) locally r-commands the phrase NP[*lgen*] as they are members of the same ARG-ST, see clause i. of (3.138). Then, NP[*lgen*] is the (maximal) head projection of the gerund *wylegiwanie się* 'lolling about'. As all NPs, this NP has a transparent *synsem* (see above) and, additionally, it is locally r-commanded. Thus, according to clause iii. of (3.138), $\boxed{1}$ locally r-commands also the gerund. Applying clause ii. of (3.138), $\boxed{1}$ locally r-commands members of gerund's ARG-ST list. Since this ARG-ST contains an RM, the RM ($\boxed{2}$) is locally r-commanded by $\boxed{1}$. Therefore, $\boxed{2}$ may be omitted in the head projection of *wylegiwanie* 'lolling about'. The omission is not obligatory, however, and we obtain both a haplology and a non-haplology sentence in (3.132a).

A similar reasoning allows us to capture optionality of haplology in (3.135a).

(3.135a) Przyzwyczaiłem się do wylegiwania ?(się) w łóżku.
got-I used RM_{inh} to $lolling_{gerund}$ about RM_{inh} in bed
'I got used to this lolling about in bed.'

(3.142) a. przyzwyczaiłem: [ARG-ST ⟨NP[*str*], $\boxed{1}$RM, PP[*do*]⟩]
b. do: [ARG-ST ⟨NP[*lgen*]⟩]
c. wylegiwania: [ARG-ST ⟨NP, $\boxed{2}$RM, PP[*w*]⟩]

The ARG-ST lists of two reflexivized elements in (3.135a) are given in (3.142a) and (3.142c). According to clause i. of (3.138), the PP *do wylegiwania (się) w łóżku* 'to lolling about in bed' is locally r-commanded by the RM $\boxed{1}$. PPs have transparent *synsem*s and, due to clause iii. of (3.138), the preposition *do* 'to' is locally r-commanded by $\boxed{1}$ as well. According to clause ii. of (3.138), the element of the preposition's ARG-ST, i.e., the NP *wylegiwania (się) w łóżku* 'lolling about in bed', is locally r-commanded by the RM $\boxed{1}$. The same argumentation as for (3.132a) shows that the RM $\boxed{2}$ is locally r-commanded by the RM $\boxed{1}$. Having said that, the omission of the RM in the gerund phrase is licensed. Again, dropping of the RM is not compulsory and both a phrase with or without *się* can be obtained.

Let us have a quick look at the ungrammatical sentences in (3.136). The missing RM is not locally r-commanded and its omission is not licensed by the principle in (3.139). Analogously, (3.134) is predicted to be ungrammatical since there is no local r-commander to license the omission of the obligatory RM.

The analysis of haplology in VCs, e.g., (3.107), is given below.

(3.107) Jan stara się mniej spóźniać (się) do pracy.
John tries RM_{inh} less be_{inf} late RM_{inh} to work
'John tries to be less late at work.'

(3.143) a. stara: [ARG-ST ⟨NP, $\boxed{1}$RM, VP[*inf*]⟩]
b. spóźniać: [ARG-ST ⟨NP, $\boxed{2}$RM, PP[*do*]⟩]

For the sake of concreteness, we assume that the main verb *stara się* 'tries', subcategorizes for a phrasal (rather than a lexical) infinitival complement, see (3.143a). The RM $\boxed{1}$ locally r-commands the VP *mniej spóźniać do pracy* 'to be less late at work', clause i. of (3.138), but also the infinitive *spóźniać się* 'to be late', clause iii. Applying clause ii., the RM $\boxed{2}$ is locally r-commanded by $\boxed{1}$. Hence, the optional haplology in (3.107) is correctly predicted by (3.139).

This analysis nicely captures optional haplology both in gerund phrases and VCs. Note that an argument composition analysis of haplology, as proposed for VCs in §3.5.2, would be a rather controversial solution if applied to NPs and PPs. There is, however, one problem with this account. The recursivity of (3.138) allows us to omit an RM fairly deeply embedded.[31] This is a desired result in VCs, e.g., (3.112), but haplology in gerund phrases is more restricted, see the judgements in (3.132) and (3.135). Nevertheless, as the analysis is configurational (see clause iii. of (3.138)), the depth of the locally r-commanded RM in gerunds can be syntactically restricted. Since data showing haplology across gerund phrases are not very clear in general, we leave the precise formulation of this specific constraint for future study.

3.5.5 Summary of Haplology

In this section, we examined haplology of the RM in various syntactic environments. Since the omission of the RM is not phonologically or morphologically conditioned, we take haplology to be a syntactic phenomenon. Such an approach is not surprising in the light of the conclusion reached in §3.2 that the RM is a syntactic item. Although haplology is optional in VCs, coordination and gerund phrases, it behaves differently in simple clauses. A single verb may appear with at most one RM but haplology interpretation does not always result. This distinction stems from the fact that the RM cannot replace a complement, cf. §3.2.7and §3.3.1, which independently excludes repetition of the RM in certain cases.

Coordination data show that haplology is not the only phenomenon responsible for the omission of the RM. Haplology may involve RMs which have different functions but it is a clause bound phenomenon. On the other hand, syntactic ellipsis may operate across a clause boundary, which allows us to drop an RM in a different clause providing the omitted RM has the same function as the remaining one.

In order to block multiple realisation of the RM, we provide a general constraint which permits at most one RM to be realised as a syntactic requirement of a verb (present on its COMPS list), although more than one RM may be present on ARG-ST. Such an analysis allows us to retain the lexical rule for impersonals in its previous form. An RM can be added to the ARG-ST list of a (reflexivized) impersonal form but this will not show up in syntax. Haplology in VCs is accounted for analogously: a reflexivized verb can accept an additional RM on ARG-ST but this will not be reflected on valence attributes.

The analysis of haplology in VCs differs from CC: unlike pronominal clitics, the RM is not raised in syntax to a reflexivized verb. This distinction allows us to capture the different behaviour of both phenomena with respect to negation: haplology, unlike CC, is possible across negation.

Haplology is possible also out of gerund phrases. As an analysis à la argument composition for haplology across NPs and PPs is rather implausible, we propose an account à la binding. This approach is configurational and relies on the intuition that 'in the presence of a higher RM, a lower RM can be omitted'. As the same principle licenses the omission of the RM in VCs and gerund phrases, the analysis uniformly captures optional haplology in both cases.

[31] The recursivity of (3.137), our definition in (3.138) is based on, is fully motivated and necessary for binding in Polish, see Marciniak (1999).

Chapter 4

Sentential Negative Marker

This chapter is concerned with the sentential negative marker (NM) *nie* 'not' which is often called a clitic.[1] We show that, contrary to orthography, the sentential NM is a verbal prefix rather than a syntactic item. Such an account is supported by prosody, idiosyncratic behaviour of certain verbs with respect to negation, coordination and word order facts. Once we treat verbal negation as a morphological category, negated and non-negated verbs should be related to prevent the lexicon from expanding. In HPSG, this can be achieved via lexical rules or underspecification and lexical constraints. We consider both accounts and show that due to some irregularities of verbal negation the latter is more adequate. We propose an HPSG morphosyntactic analysis of verbal *nie*-prefixation and compare our analysis with other approaches.[2]

4.1 Multiple Functions of *nie*

In Polish, unlike in many other languages, many different kinds of negation are expressed by one negative morpheme *nie* 'not'. It may express the plain sentential negation, (4.1a), or constituent negation, (4.1b).[3]

(4.1) a. Janek **nie** lubi Marysi.
John NM likes Mary$_{gen}$
'John does not like Mary.'

b. Janek lubi **nie** Marysię, lecz Tomka.
John likes not Mary$_{acc}$ but Tom
'John likes not Mary but Tom.'

Note that the difference between (4.1a) and (4.1b) is not just a matter of scope of negation. The verbal negation, (4.1a), may affect the case value of a complement, unlike constituent

[1]This chapter is based on two papers (co-)authored by me, i.e., Kupść and Przepiórkowski (1997) and Kupść (2000).

[2]Other phenomena related to verbal negation in Polish such as negative concord or genitive of negation are not the central topic of our discussion here. For different approaches to negative concord in Polish see, among others, Przepiórkowski and Kupść (1997c,a,b, 1999), Przepiórkowski and Świdziński (1997), Richter and Sailer (1999a), Błaszczak (2000); for genitive of negation in Polish see, e.g., Witkoś (1996a, 1998), Przepiórkowski (1996, 1997, 1999, 2000a), Przepiórkowski and Świdziński (1997).

[3]The sentential negative marker is glossed as 'NM' whereas *nie* in its other functions is glossed as 'not'. For convenience, all occurrences of *nie* in examples are written in **boldface**.

negation, (4.1b) (see §4.2.2 and §4.2.5 for a more detailed discussion).

Moreover, the grammatical status of *nie* may vary: *nie* appears as a morpheme attached to a noun or an adjective if used for the so-called lexical negation, (4.2a–b), but it is an autonomous syntactic item if used as a short negative answer, (4.2c), or a question tag, (4.2d).

(4.2) a. Ogarnął ją **nie**pokój.
embraced her_{acc} not-peace
'Fear overwhelmed her.'

b. Janek jest **nie**miły.
John is not-nice
'John is unpleasant.'

c. Czy Janek lubi Marysię? **Nie**.
Q John likes $Mary_{acc}$ not
'Does John like Mary? No.'

d. Janek lubi Marysię, **nie**?
John likes Mary not
'John likes Mary, doesn't he?'

Properties of *nie* in lexically negated nouns, (4.2a), or adjectives, (4.2b), differ from the sentential NM: it is only the latter that may influence the case value of the complement. On the other hand, unlike the negative answer, (4.2c), or a question tag, (4.2d), the sentential NM cannot be used alone, see discussion in §4.2.6.

In this chapter, we concentrate on the properties of the sentential NM. We leave aside other functions of *nie* since such a general discussion would lead us too far away from the main topic of this study, i.e., an analysis of Polish clitics.

4.2 NM as a Verbal Prefix

Verbal negation, in spite of its ubiquity and productivity, is a rather neglected phenomenon in Polish. Traditional grammars, e.g., Krasnowolski (1911), Szober (1953) and Klemensiewicz (1968), do not mention negation at all, while many other grammars do not deal with the nature of negation *per se*: the negative element *nie* is called a 'particle', cf. Bąk (1984, p. 190), Jaworski (1986, p. 47), Bartnicka and Satkiewicz (1990, p. 156) or 'modulant' in Jodłowski (1976, p. 21). These notions, however, are vague and merge syntactic and morphological categories.[4]

Many formal accounts do not treat verbal negation in a uniform way. In most papers related to Government and Binding, e.g., Willim (1990), Borsley and Rivero (1994), Śpiewak and Szymańska (1995) or Witkoś (1996a), the negative marker *nie* is assumed to be a clitic (essentially a syntactic category) but Dornisch (1997) and Witkoś (1998) treat verbal negation as a morphological category. Also Saloni and Świdziński (1985) consider verbal negation to be a morphological phenomenon. In the Metamorphosis Grammar of Świdziński (1992) as well as in Saloni and Świdziński (1998), however, this claim is withdrawn and verbal negation is analysed as a purely syntactic phenomenon.

[4]See, e.g., Gołąb *et al.* (1968, pp. 413–414) and Polański (1993, p. 386) for the definition of 'particle' as used by Polish grammarians and Zwicky (1985) for some cross-linguistic discussion.

The most extensive study of the properties of verbal negation is given in Kupść and Przepiórkowski (1997) and the discussion presented in this section is based on this paper.[5]

4.2.1 Prosody

As we have seen previously in §2.2.1 and §3.2.2, the regular lexical stress in Polish falls on the penultimate syllable of a word, cf. (4.3).[6]

(4.3) LUbisz poLUbisz
like-you will.like-you
'you like / will like'

Rubach and Booij (1985) show that if a monosyllabic verb is preceded by the NM *nie*, they are treated as a single prosodic word for the purpose of lexical stress assignment. The negated verb in (4.4a) follows the regular penultimate lexical stress pattern (see also §2.2.1 and §3.2.2), unlike the verb preceded by the interrogative particle in (4.4b).[7]

(4.4) a. **NIE** wiesz
NM know-you
'You don't know.'

b. czy WIESZ
Q know-you
'Do you know?'

The contrast in (4.4) is immediately explained if the NM is a verbal prefix, unlike the syntactic interrogative particle *czy*.

4.2.2 Frame Variation

Morphological phenomena often affect valency of a predicate. In this section, we discuss several examples of frame variation triggered by verbal negation in Polish. First, we present two non-conclusive arguments and then, we present what we take to be a strong argument for the morphological status of the NM.

Genitive of Negation The well-known frame variation triggered by negation is the Genitive of Negation (GoN). In Polish, this phenomenon is very regular: whenever a non-negated verb subcategorizes for an accusative complement, (2.75a), its negated counterpart expects a genitive NP, (2.75b) (see §2.5.2 for some discussion).[8]

(2.75) a. Janek lubi Marysię.
John likes Mary$_{acc}$
'John likes Mary.'

b. Janek **nie** lubi Marysi.
John NM likes Mary$_{gen}$

[5] A similar set of data, apart from irregularities of negation, is considered in Miller (1992a) and Recourcé (1995) to show that the French NM *ne* is prefixed to finite verbs.

[6] The stressed syllable is indicated in UPPER case.

[7] Prosodic properties of the NM are taken as an argument in favour of its affix nature in Witkoś (1998) but also in Willim (1990) although she finally does not treat the NM as an affix.

[8] See Przepiórkowski (1999, ch. 5) for a few apparent counterexamples and their discussion.

'John doesn't like Mary.'

Saloni and Świdziński (1985, pp. 141–143) use GoN as an argument in favour of a morphological status of negation.[9] They admit, however, the weakness of this argument since GoN occurs also in (complementizerless) infinitival complexes (verb clusters, VCs) where the affected complement is not an immediate argument of the negated verb, see (2.76) repeated below:[10]

(2.76) a. Jan lubi jeść ser.
John likes eat$_{inf}$ cheese$_{acc}$
'John likes to eat cheese.'

b. Jan **nie** lubi jeść sera.
John NM likes eat$_{inf}$ cheese$_{gen}$
'John does not like to eat cheese.'

The productivity of GoN, as well as its presence in infinitival complements, allows one to analyse this phenomenon in syntax (cf., e.g., Willim (1990), Witkoś (1996a, 1998), Przepiórkowski and Kupść (1997c), Przepiórkowski and Świdziński (1997), Przepiórkowski (1999)) and, hence, does not provide a strong argument for a morphological analysis of negation.

Indicative vs. Subjunctive Mood An interesting frame variation occurs in the paradigms of verbs such as *przypuszczać* 'to presume' or *sądzić* 'to judge'. Their non-negated forms allow only a sentential complement marked by the indicative complementizer *że* whereas their negative counterparts can occur with a subjunctive sentential complement introduced by the subjunctive complementizer *żeby* as well, see Saloni and Świdziński (1985, p. 144) and Greszczuk (1993, pp. 125–129):

(4.5) a. Jan przypuszcza, że / *żeby Piotr przyszedł.
John presumes COMP$_{ind}$ COMP$_{subj}$ Peter came
'John thinks that Peter came.'

b. Jan **nie** przypuszcza, że Piotr przyszedł.
John NM presumes COMP$_{ind}$ Peter came
'John doesn't think that Peter came.'

c. Jan **nie** przypuszcza, żeby Piotr przyszedł.
John NM presumes COMP$_{subj}$ Peter came
'John thinks that Peter didn't come.'

Moreover, each choice of the complementation in (4.5b–c) is associated with a slight difference in meaning. If the indicative *że*-complement is used, as in (4.5b), only a sentential negation reading is possible. On the other hand, if the subjunctive *żeby*-complement is used, as in (4.5c), the sentence is understood as if the subordinate clause rather than the main clause was negated (see the gloss).

[9] Also Menantaud (1989) considers (and rejects) GoN as an argument for negation being the morphological category of a verb.

[10] Also Menantaud (1989), Witkoś (1996a, 1998), Śpiewak and Szymańska (1995), Przepiórkowski and Kupść (1997a,c), Przepiórkowski and Świdziński (1997) and Przepiórkowski (1999) consider GoN in such infinitival complexes.

These frame alterations are taken by Saloni and Świdziński (1985) to support the lexical (morphological) analysis of negation. This argument, however, is not very strong: a similar behaviour is attested also in other languages (see Laka (1990) for Basque and Spanish, Abeillé and Godard (1997) for French) and such frame variations might be attributed to more general semantic properties of negation.

Copula *być* The most striking example of an alteration triggered by negation is the copula *być* 'to be', more specifically, its existential/locative version, cf. Krasnowolski (1911, p. 16). There are two negative forms of *być* in the present tense: the regularly inflected *nie być* 'not to be', (4.7a), and the irregular *nie ma* ('there isn't', literally: 'not has'), (4.7b).[11] It is the suppletive form *nie ma* in (4.7b) that is a true sentential negation of (4.6); (4.7a) has only a contrastive meaning.

(4.6) Książka jest na stole.
book$_{nom}$ is on table
'(the) book is on (a/the) table.'

(4.7) a. Książka **nie** jest na stole (tylko na półce).
book$_{nom}$ NM is on table only on shelf
'(the) book is not on (a/the) table (but on the shelf).'

b. Książki **nie** ma na stole (*tylko na półce).
book$_{gen}$ NM has on table only on shelf
'(the) book is not on (a/the) table.'

Negation changes the subcategorization frame of the copula: the nominative subject in (4.6) becomes a genitive NP in (4.7b). Since it is the nominative subject rather then the accusative complement that is affected and the existential/locative copula is the only Polish verb in which negation affects the subject, we do not treat the alteration in (4.7b) as an instance of GoN in the subject position, unlike Witkoś (1996a, 1998).

The irregular negation of the copula cannot be considered an opposition of two paradigms, i.e., *być* 'to be' vs. *nie mieć* 'not to have'. Irrespectively of the person and number specification of the subject, see (4.6) and (4.8), it is only the third neuter singular form *nie ma* (lit.: 'not has') that is used for negation of *być* in the present tense, (4.7b) and (4.9a):

(4.8) Chłopcy są w szkole.
boys$_{nom}$ are in school
'(the) boys are at school.'

(4.9) a. Chłopców **nie** ma w szkole.
boys$_{gen}$ NM has in school
'(the) boys are not at school.'

b. * Chłopców **nie** mają w szkole.
boys$_{gen}$ NM have$_{3rd,pl}$ in school

[11] The idiosyncratic behaviour of the existential copula *być* is considered an argument in favour of the morphological status of *nie* also in Saloni (1976), Saloni and Świdziński (1985) and Kupść and Przepiórkowski (1997).

Lit.: 'They don't have boys at school.'

c. * Chłopcy **nie** mają w szkole.
boys$_{nom}$ NM have$_{3rd,pl}$ in school
Lit.: 'Boys don't have at school.'

Moreover, in the future and past tenses, the third neuter singular form of *być* is used again in negation (the same nominative vs. genitive variation occurs), (4.10), while any form of *mieć* is impossible, (4.11). As the ungrammatical (4.11a) shows, the third person neuter form of *nie mieć* 'not have' cannot be used instead.

(4.10) Chłopców **nie** było / **nie** będzie dziś w szkole.
boys$_{gen}$ NM were$_{3rd,sg,neut}$ NM will be$_{3sg,sg,neut}$ today in school
'(the) boys were not / will not be at school today.'

(4.11) a. * Chłopców **nie** miało / **nie** będzie miało dziś w szkole.
boys$_{gen}$ NM had$_{3rd,sg,neut}$ NM [will be have]$_{3rd,sg,neut}$ today in school
Lit.: 'It didn't / won't have boys at school today.'

b. * Chłopcy **nie** mieli / **nie** będą mieli dziś w szkole.
boys$_{nom}$ NM had$_{3rd,pl}$ NM [will be have]$_{3rd,pl}$ today in school
Lit.: 'The boys didn't / won't have at school today.'

Observe that *nie ma* is the only option available when negating the existential meaning of (4.6): $\exists x\ book(x) \wedge on\_table(x)$. That is, $\neg\exists x\ book(x) \wedge on\_table(x)$ can be expressed by (4.12a), but not by (4.12b) which means: $\exists x\ book(x) \wedge \neg\ on\_table(x)$.

(4.12) a. Na stole **nie** ma książki.

b. Książka **nie** jest na stole.

The copula *być* 'to be' is also used in 'existential' constructions, such as (4.13a) and (4.14a). Since negation of the existential *być* can be obtained only via *nie ma*, the morphologically regular negation is impossible, cf. (4.13b) and (4.14b).

(4.13) a. Jest czego żałować. **Nie** ma czego żałować.
is something regret$_{inf}$ NM has something regret$_{inf}$
'It is something to regret.' vs. 'There is nothing to be sorry about.'

b. * **Nie** jest czego żałować.

(4.14) a. Jest gdzie się kąpać. **Nie** ma gdzie się kąpać.
is where RM bathe$_{inf}$ NM has where RM bathe$_{inf}$
'There is a place to take a bath.' vs. 'There is no place to take a bath.'

b. * **Nie** jest gdzie się kąpać.

Before concluding this subsection, let us note yet another difference between the two negative forms of the copula *być* 'to be'. As stated above, if (4.6) is regularly negated, it has only a contrastive, (4.7b), rather than the sentential negation meaning, (4.7a). This observation is further supported by the fact that, unlike the regular *nie jest*, the suppletive *nie ma* may license *n*-words, e.g., *nic* 'nothing' in (4.15).

(4.15) a. Niczego tam **nie** ma.
nothing$_{gen}$ there NM has
'There is nothing there.'

b. * Nic tam **nie** jest.
nothing$_{nom}$ there NM is

Polish is a negative concord language. Roughly, words such as *nikt* 'nobody' or *nic* 'nothing' (which we call *n*-words after Laka (1990)), are licensed in a clause if the verb is overtly negated, i.e., preceded by the NM.[12] However, as argued in Przepiórkowski and Kupść (1999), licensing of *n*-words is essentially a semantic phenomenon and the presence of the morphosyntactic NM is insufficient. There are at least three Polish environments where the NM is present but semantic negation does not result (the so-called pleonastic negation, see Brown (1996, 1999), Brown and Franks (1995) for similar facts from Russian): negative yes/no questions, the temporal expression *dopóki nie* 'until', see (4.16a), and the adverbial *omal nie* 'almost'.

(4.16) a. Będziemy czekać dopóki jej **nie** przeczytają.
will-be-we wait until her$_{gen}$ NM read-they
'We'll be waiting until they have read it.' (= '... as long as they read it.')

b. Będziemy czekać dopóki ją czytają.
will-be-we wait until her$_{acc}$ read-they
'We'll be waiting as long as they are reading it.'

In (4.16a), *nie* is a true NM since the NP *jej* occurs in the genitive rather than in the usual accusative case, as in (4.16b). As indicated in the glosses, however, none of (4.16) has a semantic negation reading.

Przepiórkowski and Kupść (1999) argue that clauses headed by overtly negated verbs are potentially ambiguous between expressing *propositional* and *eventuality* semantic negation and that it is only the latter that licenses *n*-words. Therefore, the contrast in (4.15) boils down to the fact that *nie ma* is ambiguous and, hence, may express eventuality negation whereas *nie jest* expresses only the propositional negation (and, thus, it does not license *n*-words, see (4.15b)).

We conclude that in the case of the existential/locative copula *być* 'to be' negation affects not only the valence of the verb, but its morphology as well, and must be analysed as a lexical phenomenon. If some verbs have to be negated in the lexicon and if all verbs can be negated in the lexicon, we consider it desirable to view verbal negation uniformly as a lexical phenomenon.

4.2.3 Idiosyncratic Forms

Polish verbal negation is not a homogeneous phenomenon. Some verbs do not have negated forms at all whereas there are verbs which occur in negated forms only.

'Negative' Verbs As occasionally noticed, mainly in orthographic dictionaries, e.g., Jodłowski and Taszycki (1968), Szymczak (1975), but also in Bugajski (1983, p. 81) and, especially, Kupść and Przepiórkowski (1997), there are verbs in contemporary Polish which

[12]See Przepiórkowski and Kupść (1997c,a, 1999), Richter and Sailer (1999a), Błaszczak (2000) for some analyses of Polish negative concord.

do not have forms without *nie*, e.g., *nienawidzić / *nawidzić* 'to hate', *nie cierpieć* 'to dislike/detest',[13] (4.17a), *niepokoić / *pokoić* 'to worry', (4.17b), *niewolić / *wolić* 'to keep imprisoned', (4.17c), *nie omieszkać / *omieszkać* 'not to neglect', *niedomagać / *domagać* 'to suffer from',[14] *niecierpliwić się / *cierpliwić (się)* 'to be impatient'.[15] The fact that *nie* is written separately in, say, *nie cierpieć* 'to detest', unlike in *nienawidzić* 'to hate', is just an orthographic convention and should not be taken into account while considering the status of the NM.

(4.17) a. **Nie**nawidzę / **nie** cierpię Marysi.
hate-I NM suffer-I Mary$_{gen}$
'I hate / detest Mary.'

b. Zachowanie Piotra **nie**pokoi Marysię.
behaviour of-Peter worries Mary$_{acc}$
'Peter's behaviour worries Mary.'

c. **Nie**wolili ją w podziemiach zamku.
imprisoned-they her$_{acc}$ in underground of-castle
'They kept her imprisoned in the castle.'

Certain verbs may change their meaning if overtly negated, e.g., *nie dostawać* 'to lack' vs. *dostawać* 'to receive', *nie dosypiać* 'to lack sleep' vs. *dosypiać* 'to sleep till some time', *nie dojadać* 'to starve' vs. *dojadać* 'to finish eating', etc. Although non-negated forms of these verbs exist, as indicated in the glosses, they have different meanings. Thus, these *nie*-verbs are either irregular forms which cannot appear without *nie* or they are forms of regularly negated verbs.

Verbs which obligatorily occur with *nie*, such as in (4.17), do not form a uniform class. Some of them behave as non-negated verbs do, e.g., they can be overtly negated, whereas others already contain the NM. As (4.17a) shows, verbs *nienawidzić* 'to hate' or *nie cierpieć* 'to detest' require a genitive complement, an instance of genitive of negation (GoN) as argued in Przepiórkowski (1999) on the basis of examples such as (4.18).

(4.18) Jan **nie**nawidzi / **nie** cierpi pisać listów / *listy.
John hates NM suffers write$_{inf}$ letters$_{gen}$ letters$_{acc}$
'John hates / dislikes writing letters.'

As mentioned in §4.2.2, in infinitival verbal complexes GoN can be triggered on a distance, (2.76). Both *nie cierpieć* and *nienawidzić* may take a bare infinitival complement. As (4.18) shows, they trigger GoN as well: the complement of the non-negated verb *pisać* 'to write' is a genitive NP rather than the usual accusative NP.

On the other hand, *niewolić* 'to imprison' and *niepokoić* 'to worry' in (4.17b–c) behave as non-negated verbs do and they can be overtly negated, (4.19). Once negated, they trigger GoN:

[13] There exists also a regularly negated verb *cierpieć* 'to suffer'. This verb has a different meaning and should be distinguished from the idiomatic *nie cierpieć* 'to detest' in (4.17a).

[14] The regularly negated verb *domagać się* 'to demand' exists but it obligatorily occurs with the reflexive marker *się* and has a different meaning; thus, it cannot be the non-negated counterpart of (non-reflexive) *niedomagać* 'to suffer from'.

[15] Przepiórkowski (1999) mentions also predicatively used expressions *nie sposób* '(be) not possible' and *niepodobna* '(be) unlikely'.

the accusative complement of a non-negated verb, (4.17b–c), appears in genitive, (4.19a–b), if the verb is negated.[16]

(4.19) a. Zachowanie Piotra **nie nie**pokoi Marysi.
behaviour of-Peter NM worries Mary_{gen}
'Peter's behaviour does not worry Mary.'

b. **Nie nie**wolili jej tam.
NM imprisoned-they her_{gen} there
'They didn't keep her imprisoned there.'

Apparently, *nienawidzić* or other negated verbs can be negated as well, cf. (4.20).

(4.20) a. Tomek **nie nie**nawidzi klasówek, toleruje je.
Tom not hates class-tests tolerates-he them
'Tom does not hate class-tests, he tolerates them.'

b. ? Tomek **nie nie** lubi klasówek, toleruje je.
Tom not NM likes class-tests tolerates-he them
'Tom does not dislike class-tests, he tolerates them.'

Note however that this kind of double negation is not a multiplication of the NM. The first *nie* is constituent negation in (4.20), as indicated by the sharp contrast with (4.21) where, unlike in (4.20), no contrastive completion of the sentence is possible.

(4.21) Tomek **nie**nawidzi / **nie** lubi klasówek, *toleruje je.
Tom hates NM likes class-tests tolerates-he them
'Tom hates / does not like class-tests.'

Although (4.18) strongly suggests that *nie cierpieć* 'to detest' and *nienawidzić* 'to hate' contain the NM, its presence might be questioned since none of (4.17)–(4.18) has the sentential negation meaning. Remember, however, that the NM does not always have the semantic sentential negation effect, see (4.15b) and (4.16). We assume that, as in (4.16), the NM is present in (4.17a) and (4.18), it triggers GoN but has no sentential (eventuality) negation impact.

Since verbs in (4.17) do not have non-negated counterparts, it might also be suggested that *nie* is a morpheme of lexical negation, see (4.2a–b), rather than the NM added to a verb. In the case of *niewolić* 'to imprison' and *niepokoić* 'to worry', it indeed is the case: these verbs trigger GoN only if overtly negated, cf. (4.17b–c) vs. (4.19). Moreover, as noted in Jodłowski and Taszycki (1968), they are derived from the nouns *niepokój* 'anxiety' and *niewola* 'prison', respectively, which confirms that *nie* belongs to the morphological nominal base and it is not the NM. Furthermore, both verbs can be prefixed: there exist verbs *z**nie**wolić* 'to imprison (perfective)' or *za**nie**pokoić* 'to worry (perfective)'. Note that similar perfective forms of the regularly negated verb *nie kochał* '(he) didn't love' or the irregular negative verb *nie cierpiał* '(he) detested' are impossible, see (4.22).

(4.22) a. * Jan z**nie**kochał Marysię.
John z-not-loved Mary

b. * Jan z**nie**cierpiał Marysię.
John z-not-suffered Mary

[16] Also *niecierpliwić się* 'to be impatient' can be overtly negated. Since it does not have an accusative complement, GoN is not observable.

Interestingly, however, *nienawidzić* 'to hate' can be prefixed as well, (4.23a), which might suggest that also *nie* in *nienawidzić* 'to hate' is a morpheme of lexical negation rather than an NM added to a verb, unlike in *nie cierpieć* 'to detest'. Nevertheless, this is a premature conclusion. Although regularly negated verbs in principle cannot be prefixed, (4.22a), there is one exception, see (4.23b).[17]

(4.23) a. Jan z**nie**nawidził Marysię.
John z-hated Mary$_{acc}$
'John started hating Mary.'

b. ? Jan z**nie**lubił Marysię. (colloquial)
John z-not-liked Mary$_{acc}$
'John has disliked Mary.'

The verb ***znie**lubić* 'to dislike' in (4.23b) sounds colloquial and slightly comic but it sharply contrasts with the clearly ungrammatical ***znie**kochać* (lit.: 'z-not-love'), (4.22a). Since *nie lubić* 'not like' is a regularly negated verb, see (2.75), *nie* in (4.23b) is an NM added to the verb and it is not part of the verbal stem. This indicates that the perfective prefixation is not strictly related to whether *nie* is part of the verbal stem or not. Therefore, we do not take the contrast between (4.22b) and (4.23a) to be a strong argument for the distinct status of the NM in *nie cierpieć* 'to detest' and *nienawidzić* 'to hate', respectively.

We conclude that in verbs such as *niewolić* 'to imprison' or *niepokoić* 'to worry' *nie* is just part of the morphological base. On the other hand, *nienawidzić* 'to hate' or *nie cierpieć* 'to detest' contain a true NM: it triggers GoN but it does not have the sentential negation effect. Moreover, as a diachronic observation shows, the NM is added to the verb in *nienawidzić* 'to hate' and *nie cierpieć* 'to detest': in XVth and XVIth century Polish, both verbs had non-negated counterparts, see Brückner (1957).

'Positive' Verbs Saloni and Świdziński (1985, p. 143) mention that the polite imperative form *proszę* '(I) ask' has no negated forms, see (4.24).[18]

(4.24) a. Proszę przyjść o siódmej.
ask-I come at seven
'Please, come at seven.'

b. * **Nie** proszę przyjść o siódmej.
NM ask-I come at seven

As *proszę* '(I) ask' in (4.24) is used performatively here and such verbs are not negated in general, this might be an independent explanation of the ungrammaticality of (4.24b).

Other verbs which do not have negative forms in Polish are weak auxiliaries, i.e., conditional and past tense auxiliary verbs, see §2.5.5. Unlike their counterparts in Bulgarian and Serbian/Chroatian, cf. Borsley and Rivero (1994), Borsley (1999), Polish weak auxiliaries cannot be preceded by the NM, (4.25b) and (4.26b), and the NM has to be attached to the participle, (4.25a,c) and (4.26a,c) (see also Dornisch (1997), Witkoś (1998); see §2.5.5 for similar examples).

[17] There seem to be a few more such perfective negated verbs, namely, *zaniemówić* 'to be speechless' (lit.: za-not-speak) or *zaniedbać* 'to neglect' (lit.: za-not-care). It is possible, however, that they are derived from the lexically negated nouns *niemowa* 'a mute' and *niedbałość* 'carelessness' as their semantics seems closer to these nouns rather than negated verbs.

[18] This verb is idiosyncratic also in this respect that it has only first person forms.

(4.25) a. **Nie** widzieli -śmy Marii.
NM saw AUX.PAST-we Mary_{gen}
'We did not see Mary.'

b. * **Nie**-śmy widzieli Marii.

c. Wcale-śmy **nie** widzieli Marii.
at-all-AUX.PAST-we NM saw Mary
'I haven't seen Mary at all.'

(4.26) a. **Nie** widzieli -byśmy Marii.
NM saw AUX.COND-we Mary
'We would not see Mary.'

b. * **Nie** byśmy widzieli Marii.

c. Wcale byśmy **nie** widzieli Marii.
at-all AUX.COND-we NM saw Mary
'We wouldn't see Mary at all.'

Weak auxiliaries in (4.25)–(4.26) are treated in Borsley (1999) as syntactic enclitics (see §2.5.5 for some discussion). They can be detached from the participle and follow almost any preceding element, cf. (4.25c) and (4.26c), but not *nie*, see (4.25b) and (4.26b). Note that the ungrammaticality of (4.25b) and (4.26b) cannot be explained by prosody or phonology alone. The weak auxiliaries in (4.25)–(4.26) are syllabic verbs and, in principle, might support the negative proclitic. As shown in (4.4a), the NM can be attached to monosyllabic verbs. Moreover, as noted in Halpern (1995, p. 215), it is not impossible for enclitics in general to take a preceding proclitic as a phonological host. Taking into account these facts, we attribute the data in (4.25b) and (4.26b) to the fact that Polish weak auxiliaries cannot be negated (unlike their verbal complements, (4.25c) and (4.26c)).[19]

As noted in §2.5.5, weak auxiliaries contrast with the (strong) future auxiliary with respect to negation. The future auxiliary *będzie* 'will be' can be overtly negated, (2.122a), but its participial (or infinitive) complement cannot, (2.122b), see Dornisch (1997), Witkoś (1998, p. 232).[20]

(2.122) a. Piotr **nie** będzie jutro śpiewał / śpiewać.
Peter NM AUX.FUT tomorrow sing sing_{inf}
'Peter will not sing tomorrow.'

b. # Piotr będzie jutro **nie** śpiewał / nie śpiewać.

The periphrastic future tense form, as in (2.122a), is considered a syntactic construction in Borsley (1999). Therefore, if we require the future auxiliary to subcategorize for a non-negated (infinitive or participle) complement, the ungrammaticality of (2.122b) is correctly excluded. Note that negation of the infinitive or participle is not ruled out in general: the negated bare participles in (4.25c), (4.26c) or the negated infinitive in (4.27) are perfectly grammatical.

[19] The same possibility, i.e., that the NM can be attached only to certain verbs, is also considered in Witkoś (1998, p. 232).

[20] (2.122b) is grammatical only in the contrastive meaning, i.e., if *nie* is not the NM:

(i) Piotr będzie jutro **nie** śpiewał / **nie** śpiewać, lecz tańczył / tańczyć.
Peter AUX.FUT tomorrow not sing not sing_{inf} but dance dance_{inf}
'Peter will not sing but dance tomorrow.'

(4.27) Piotr wolał **nie** śpiewać.
Peter preferred NM sing$_{inf}$
'Peter would rather not sing.'

Conclusion The data presented in this subsection clearly show that Polish verbal negation is not fully productive. There are verbs which lack non-negated or negated forms. A straightforward account of such idiosyncrasies is to enumerate defective forms directly in the lexicon. This, however, leads us to treating verbal negation as a lexical (morphological) phenomenon.

4.2.4 Coordination

The behaviour of the NM in coordination differs from that of Polish clitics. If both conjuncts are to be negated, *nie* must be repeated in coordination, (4.28b), and cannot be omitted, (4.28c) (in contrast to auxiliary clitics, cf. *bym* 'I'd' in (4.28a), Polish pronominal clitics, cf. §2.2.6, or the RM, cf. §3.2.3, which can have wide scope over coordination).

(4.28) a. Chętnie bym [czytał i pisał] książki.
willingly aux.cond-I read and write books$_{acc}$
'I'd willingly read and write books.'

b. **Nie** czytam i **nie** piszę książek.
NM read-I and NM write-I books$_{gen}$
'I don't read and write books.'

c. * **Nie** [czytam i piszę] książek.
NM read-I and write-I books$_{gen}$

The obligatory repetition of the Polish NM in (4.28b) does not follow from general properties of negation. Note that the English *not* or the French *ne pas* 'not' may negate both conjuncts even though they occur only once:

(4.29) a. Dana will not [walk and talk]. Kim and Sag (1995)

b. Paul dit ne pas [lire le journal ou regarder la télevision].
Paul says not read the newspaper or watch the television
'Paul pretends (to) not read the newspaper or watch TV.'
Abeillé and Godard (1997)

According to Miller (1992a), the obligatory repetition of an item in coordination is a property of affixes but not of syntactic items.

Note, however, that also certain (syntactic) adverbs cannot take wide scope over coordination. French manner adverbs or *très* 'very' modify the word directly and must be repeated in coordination, cf. Abeillé and Godard (1997). As observed in Śpiewak and Szymańska (1995), there exists a class of Polish adverbs that modify the verb directly, e.g., *prawie* 'almost/nearly', (4.30), *naprawdę* 'really', etc.

(4.30) a. Tomek prawie skończył swoją książkę.
Tom almost finished self's book
'Tom has almost finished his book.'

b. * Prawie Tomek skończył swoją książkę.

c. ? Tomek skończył prawie swoją książkę.

d. * Tomek skończył swoją książkę prawie.

Therefore, this would be a natural class which *nie* should belong to if it were a negative adverb (similar to the French *pas* or the English *not*, e.g., Pollock (1989)). However, unlike in the case of *nie*, *prawie*-adverbs may also optionally take wide scope over coordination of verbs, see (4.31).

(4.31) a. Tomek prawie [przeczytał] i zrozumiał książkę Jasia.
Tom almost read and understood book of-John
'Tom almost read and understood John's book.'

b. Tomek prawie [przeczytał i zrozumiał] książkę Jasia.

Hence, the parallelism between such scope elements and the NM cannot be preserved. *Nie* should be analysed as a verbal prefix.

4.2.5 Word Order

In Polish, a relatively free word order language, sentential negation is obtained only if the NM *nie* immediately precedes the verb, cf. (4.32).

(4.32) a. Często **nie** czytam książek.
often NM read-I books$_{gen}$
'I often don't read books.'

b. Adam **nie** mieszka już w Tybindze.
Adam not lives yet in Tübingen
'Adam doesn't live in Tübingen any more.'

Placing *nie* elsewhere results in negating the immediately following element: 'Adam' in (4.33a), 'in Paris' in (4.33b), 'often' in (4.34a) and 'books' in (4.34b).[21]

(4.33) a. **Nie** Adam mieszka w Tybindze, tylko Ania.
not Adam lives in Tübingen, only Ania.
'It is not Adam who lives in Tübingen but Ania.'

b. Ania mieszka **nie** w Paryżu, tylko w Tybindze.
Ania lives not in Paris, only in Tübingen.
'Ania does not live in Paris but in Tübingen.'

(4.34) a. **Nie** często czytam książki.
not often read-I books$_{acc}$
'It is not often that I read books.'

b. Często czytam **nie** książki, ale gazety.
often read-I not books$_{acc}$, but newspapers$_{acc}$
'What I often read is not books, but newspapers.'

Note that constituent negation, (4.33)–(4.34), differs from sentential negation, (4.32), since it does not trigger GoN. In (4.33)–(4.34), the verb is not negated and the NP complement occurs in accusative, unlike in (4.32). Moreover, the sentential NM takes only a verb as its host, unlike the constituent negation *nie* which can precede a noun, (4.33a) and (4.34b), a

[21] Similar facts are also discussed in Śpiewak and Szymańska (1995) and Witkoś (1996a).

PP, (4.33b), an adverb, (4.34a), etc. It should be noted that the high degree of selection with respect to the host contrasts the NM with Polish clitics, see §2.2.4, §2.5.5 and §3.2.5.

Unlike Polish clitics (see §2.2.5, §3.2.6), the sentential NM *nie* has a fixed position and always immediately precedes the verb, (4.32). Moreover, unlike in Romance and also some Slavic languages, e.g., Bulgarian or Macedonian, nothing (including clitics) can separate the NM from its verbal host, cf. (4.35b–c).

(4.35) a. Janek by go **nie** spotkał.
John cond.aux him$_{cl}$ NM met
'John would not meet him.'

b. * Janek **nie** by go spotkał.
John NM cond.aux him$_{cl}$ met

c. * Janek by **nie** go spotkał.
John cond.aux NM him$_{cl}$ met

Also Polish "expletives" show that the relationship between *nie* and the verb is stronger than between any pair of adjacent words. While the vulgar expletive *kurwa* ('fucking', literally: 'whore') can be inserted between any adjacent words (including clitics), it cannot be inserted between the NM and the verb, see (4.36c).

(4.36) a. Może, kurwa, go **nie** wal tym łomem?!
maybe EXPL him$_{cl}$ NM batter this crowbar
'Why don't you fucking stop battering him with this crowbar?!'

b. Może go, kurwa, **nie** wal tym łomem?!

c. * Może go **nie**, kurwa, wal tym łomem?!

d. Może go **nie** wal, kurwa, tym łomem?!

e. Może go **nie** wal tym, kurwa, łomem?!

Observe that such a strict adjacency requirement is not the property of *prawie*-adverbs. Expletives, (4.37a), as well as clitics, (4.37b), can separate the adverb from the verb.

(4.37) a. Tomek prawie, kurwa, skończył swoją książkę.
Tom almost EXPL finished self's book
'Tom has almost, fucking, finished his book.'

b. Tomek prawie ją skończył.
Tom almost her$_{cl}$ finished
'Tom has almost finished it.'

Such contrasts are immediately explained if the Polish NM differs from both clitics and *prawie*-adverbs in being a verbal prefix.[22]

[22] In Polish, unlike in English, expletive insertion within words is in general infelicitous. Thus, the English *fan-fucking-tastic*, Sproat (1992, p. 49, fn. 20), cannot be rendered by the Polish * *fanta-kurwa-styczny*, in spite of the fact that infixation occurs before the prosodic foot (the syllable *-sty-*) which is a precondition for such an infixation in English.

4.2.6 Ellipsis

Data presented so far support rather strongly a morphological analysis of the NM. Kupść and Przepiórkowski (1997) admit, however, one problem with such an analysis, namely the behaviour of *nie* with respect to syntactic ellipsis. As (4.38) shows, *nie* can stand alone when the verb is omitted. Such a behaviour is unexpected if the NM and the verb form a single lexical unit.

(4.38)	Marysia	lubi	Janka,	a	Tomek	**nie**	[lubi	Janka].
	Mary	likes	John	but	Tom	not	likes	John
	'Mary likes John but Tom does not.'							

The sequence *Tomek nie* 'Tom not' can be completed to a full sentence (indicated in square brackets), which suggests that *nie* is the NM here. Witkoś (1998, p. 238) argues, however, that *nie* in (4.38) should be considered a different item than the verbal prefix NM.

This hypothesis is based on the fact that *nie* can be such an orphan under certain circumstances only. If the preceding sentence is negative, the verb alone cannot be elided, see (4.39). Note also that (4.38) is grammatical only if the contrastive conjunction, such as *a* 'but' or *lecz* 'but', is used; the omission of the verb alone is impossible if a different conjunction is used, see (4.40).

(4.39)	* Marysia	**nie**	lubi	Janka	i	Tomek	**nie**.		
	Mary	NM	likes	John	and	Tom	not	likes	John
	'Mary does not like John and neither does Tom.'								(putatively)

(4.40)	* Marysia	lubi	Janka	i	/ lub	Tomek	**nie**.	
	Mary	likes	John	and	or	Tom	not	
	'Mary likes John and / or Tom does not.'							

We conclude, after Witkoś (1998), that *nie* in (4.38) is not the NM but an overt exponent of the (negative) polarity of the sentence, necessary to obtain a desired contrast. Similarly, *tak* 'yes' in (4.41) indicates the affirmative polarity of the sentence contrasted with the preceding negative one.

(4.41)	Marysia	**nie**	lubi	Janka,	a	Tomek	tak	[#lubi	Janka].
	Mary	NM	likes	John	and	Tom	yes	likes	John
	'Mary does not like John but Tom does.'								

Therefore, we do not take data such as (4.38) to be an obstacle for a morphological analysis of the NM.

4.2.7 Summary

The data presented in §4.2.1–§4.2.5 strongly support the morphological analysis of the NM *nie*. Such an account allows us to straightforwardly explain prosodic properties of monosyllabic negated verbs, the idiosyncratic behaviour of the copula *być* 'to be', the lack of full productivity of *nie*-attachment to verbs, as well as coordination and word order facts. Assuming that *nie* is a syntactic clitic cannot account for lexical exceptions (the suppletive negative form of *być*, the lack of 'negative' or 'positive' forms of certain verbs), strict adjacency to the verb (exceptional in the domain of Polish clitics) and coordination facts (a wide scope reading should be possible). The analysis of *nie* as a negative adverb would additionally leave prosodic properties of the NM unexplained.

Due to restrictions on ellipsis discussed in §4.2.6, we treat *nie* left alone in such constructions as an item different from the sentential NM: *nie* in (4.38) expresses the negative polarity of the contrasted sentence, similarly to *tak* 'yes' which indicates the affirmative polarity of the sentence in (4.41). As in Witkoś (1998) and Przepiórkowski (1999), we do not take data such as in (4.38) to be an obstacle for the morphological analysis of the NM.

4.3 A Lexical Rule Approach

Apart from the irregularities enumerated in §4.2.2 and §4.2.3, verbal negation is a highly productive process and negated and non-negated verbs should be related in the grammar. In this section, we discuss a potential HPSG-encoding of this idea, namely a lexical rule (LR) approach. For the moment, we concentrate on the formulation of the LR itself and we leave aside morphological aspects of the analysis. We will come back to these issues in §4.4, where an alternative (and more efficient) analysis is proposed.

We follow Przepiórkowski and Kupść (1997a) and use the morphosyntactic NEG attribute to signal whether the NM is present. Since the sentential NM can be attached only to verbal objects (unlike lexical or constituent negation, cf., (4.2a) and (4.1b), respectively), we assume that NEG is appropriate only to the sort *verb*. We take the value of NEG to be boolean, indicated as '+' if the verb is preceded by the NM *nie* and '−' otherwise. For example, the negated verb *nie lubi* 'doesn't like', (2.75b) on p. 159, is specified as [NEG+], whereas its non-negated counterpart *lubi* 'likes', (2.75a), is [NEG−].

4.3.1 A Lexical Rule

A LR which may relate non-negated and negated verbs is schematically given in (4.42).[23]

(4.42)
$$\begin{bmatrix} \textit{word} & \\ \text{PHON } \boxed{1} & \\ \text{SS}|\text{LOC}|\text{CAT}|\text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{NEG} & - \end{bmatrix} \end{bmatrix} \mapsto \begin{bmatrix} \text{PHON } \langle \text{nie} \rangle \oplus \boxed{1} & \\ \text{SS}|\text{LOC}|\text{CAT}|\text{HEAD}|\text{NEG} & + \end{bmatrix}$$

The input of (4.42) (the left-hand side of '↦') is a non-negated ([NEG−]) verb. The output of the rule (the right-hand side) is the form of the input verb (encoded as token-identity of PHON values) preceded by the NM *nie*. Since the output verb is overtly negated, its NEG value is different as well, i.e., [NEG+]. The recursive application of (4.42), i.e., multiple NM-prefixation, is blocked by distinct NEG specifications of the input and output verbs. Formally, we assume here encoding of lexical rules using description-level lexical rules (DLR) proposed for HPSG in Meurers (1999). Therefore, lexical rules, as in (4.42), relate objects of the sort *word* rather than their lexical entries (descriptions of *word* objects).

The lexical rule in (4.42) provides correct forms for regularly negated verbs, e.g., (2.75). Since verbs with *nie* as part of their morphological base, e.g., *niepokoi* 'worries' in (4.17b), are regularly negated, cf. (4.19), the rule in (4.42) holds for these pairs as well. Irregular forms, however, cannot be obtained via this LR.

[23] We adopt the convention that only these attributes whose values are affected by the lexical rule are explicitly specified on both sides of the rule.

4.3.2 Problems

As shown in §4.2.3, verbal negation in Polish is not fully productive. The rule in (4.42) accounts for regularly negated verbs but nothing prevents it from (incorrectly) applying also to exceptions.

First, the existential copula in (4.6) is not overtly negated, i.e., [NEG −], and it may serve as the input for (4.42). Note, however, that the negated form in (4.7b) and (4.9a) cannot be obtained via (4.42). The base form of the verb changes when the verb is negated and the PHON value of the input does not occur in the output.

Furthermore, verbs which cannot be negated, e.g., *proszę* '(I) ask' in (4.24a), or weak auxiliaries, cf. (4.25)–(4.26), may be subject to (4.42). These verbs are not negated and they are marked as [NEG−] in the lexicon. Therefore, they may be used as the input of the lexical rule and their negated forms would be expected to exist, contrary to facts, see (4.24b), (4.25b), (4.26b).

Finally, lexically negated verbs such as *nie cierpieć* 'to hate/detest' in (4.17a) are not obtained via (4.42). This is less of a problem since they can be idiosyncratically marked as [NEG+] in the lexicon. Such a specification would make them exempt from the lexical rule in (4.42) and, as desired, only negated forms would be present in the lexicon.

4.3.3 A Solution

In order to maintain a LR approach, exceptional lexical items have to become 'invisible' for the lexical rule. One possible way to achieve this effect is to introduce a (rule-specific) attribute to mark irregular forms and modify the lexical rule accordingly. This would prevent the application of the LR to irregular verbs (cf. Koenig and Jurafsky (1994) and Riehemann (1998) for a discussion of the treatment of exceptions in LRs approaches). We briefly consider such a modification below.

Let us use the boolean NEG-LEX-RULE attribute to indicate whether a lexical item is subject to the LR (marked as '+') or not ('−'). The lexical rule in (4.42) can be modified as follows:[24]

(4.43)
$$\begin{bmatrix} \textit{word} \\ \text{PHON } \boxed{1} \\ \text{SS}|\text{LOC}|\text{CAT}|\text{HEAD} \begin{bmatrix} \textit{verb} \\ \text{NEG } - \end{bmatrix} \\ \text{NEG-LEX-RULE } + \end{bmatrix} \mapsto \begin{bmatrix} \text{PHON } \mathrm{f}_{nie}(\boxed{1}) \\ \text{SS}|\text{LOC}|\text{CAT}|\text{HEAD}|\text{NEG } + \end{bmatrix}$$

In this setup, verbs such as *nie cierpieć* 'to detest' in (4.17a) would be marked as [NEG-LEX-RULE −] in the lexicon. They are positive exceptions to (4.43): these verbs have a negated form which is not obtained in a regular way. Also the verb *proszę* '(I) ask' in (4.24a) or weak auxiliaries in (4.25a) and (4.26a) would be marked as [NEG-LEX-RULE −]. This would correctly block application of (4.43), and since the verbs are also marked [NEG−], the NM will never be present. In order to obtain the correct negated forms of the existential copula *być* 'to be', e.g., (4.7b), (4.9a), the PHON value of the output in (4.42) is replaced, as common in LRs approaches, by a more complex function f_{nie} in (4.43). The value of this function is the concatenation of *nie* and the phonology of the input verb (as in (4.42)) unless this verb is a

[24]This modification was proposed by Jesse Tseng (p.c.) in the discussion of preliminary versions of my paper Kupść (2000) these sections are based on.

form of the existential/locative copula *być*. In the latter case, the prefix *nie* is attached to the verbal stem idiosyncratically specified by the function.

Note that the LR in (4.43) changes the NEG value of the input but it does not affect [NEG-LEX-RULE +] marking: the value of this attribute is preserved in the output verb. If the value of NEG-LEX-RULE also changed in the output, such a specification would be redundant for regularly negated verbs. All non-negated ([NEG−]) forms of regularly negated verbs have to be [NEG-LEX-RULE +] to become subject to (4.43). If regularly negated forms ([NEG+]), obtained via (4.43), are additionally marked [NEG-LEX-RULE −], there would be two factors that prevent the recursive application of the LR: the morphosyntactic NEG attribute and the rule-specific NEG-LEX-RULE attribute. Changing only the NEG value in (4.43) avoids this redundancy and is sufficient to block the recursive application of the LR.

Although (4.43) does not have the deficiencies of (4.42), it still is not unproblematic.

Once NEG-LEX-RULE is introduced, one has to specify non-negated verbs which are subject to (4.43), i.e., marked [NEG-LEX-RULE+]. Since verbs which can be regularly negated do not form a uniform class, there seems to be no easy way in general to identify such verbs. This means that, in fact, the value of NEG-LEX-RULE has to be individually specified on each verb. Therefore, the relevance and usefulness of the lexical rule becomes questionable.

An alternative solution is to mark exceptional verbs as [NEG-LEX-RULE−] whereas regular [NEG−] verbs would have the value of the NEG-LEX-RULE attribute unresolved in the lexicon. Regular verbs would satisfy the description of the input in (4.43) and negated forms would be obtained. This solution, however, leads to doubling lexical items for regularly negated verbs: each regular [NEG−] verb would be [NEG-LEX-RULE+] but it would also have a [NEG-LEX-RULE−] counterpart.

Another, more general problem with lexical rules such as (4.43) is related to the acquisition of lexical exceptions. Learnability considerations provide an objection to lexical rules which do not apply to certain listed exceptions. As argued in Fodor (1992), such exceptions are not learnable.

The issues considered above discourage the use of a LR to Polish morphological negation.

4.4 Lexical Constraints

The analysis proposed in this section crucially relies on underspecification in lexical entries. Values of specific attributes are resolved via lexical constraints rather than lexical rules. Such an approach allows us to provide a general treatment of NM-prefixation both in regular and irregular negated verbs.

4.4.1 A General Idea

As stated in §4.3, the presence of the NM in a verb is signalled by the value of the morphosyntactic NEG attribute. At present, we assume that verbs which can be regularly negated, e.g., *lubi* 'likes' in (2.75), but also *niepokoi* 'worries' or *niewolili* '(they) kept imprisoned' in (4.17b–c), have the NEG value unspecified in the lexicon.[25] Hence, they can be either [NEG+] or [NEG−], i.e., negated or not. In §4.4.2, we propose an analysis which accounts for the morphological spell-out of the NEG attribute.

[25] An account along these lines has also been suggested in Borsley (1999).

Let us turn to exceptions. As above, verbs which cannot be negated, e.g., *proszę* '(I) ask' in (4.24a) or weak auxiliaries, (4.25)–(4.26), are lexically marked as [NEG−]. Therefore, the lack of their negated forms is accounted for. Verbs such as *nie cierpieć* 'to detest' or *nienawidzieć* 'to hate' in (4.17a), are lexically specified as [NEG+]. This correctly predicts that they have only a negative form.

The existential/locative copula *być* 'be' can be negated but it has a suppletive negative form. We leave past and future tense forms unspecified for NEG in the lexicon. We require, however, that in the present tense the NM is attached to a form of the verb 'have' rather than 'be'. A detailed analysis of the copula *być* is given below in §4.4.2.

In this approach, only exceptional forms are idiosyncratically marked. No special marking is required for regularly negated forms. Since they are left unspecified, both values of NEG are possible. The current solution also avoids redundant specification or duplication of regularly negated verbs.

4.4.2 A Morphosyntactic Analysis

In §4.2 we have shown that the NM is a verbal prefix. The question remains, however, what kind of a prefix it is. Since verbal negation might have no direct impact on semantics, Przepiórkowski and Kupść (1999), and does not change the syntactic category of the verb, we do not treat the NM as a derivational affix. Negated verbs are also unlikely to be compounds since the NM cannot be an independent item, see §4.2.6. In terms of traditional morphology, the most plausible category for the NM would be an inflectional affix. However, unlike many inflectional affixes, the morphological form of the NM never changes or affects the morphological shape of the verb. Of course, NM-prefixation could be treated as a very regular inflection. It should be noted that even if inflection of the non-negated verb is irregular, it is independent of the presence or absence of the NM. Thus, the attachment of the NM is 'invisible' for inflection of the verb proper. Since morphology as such is not the main topic of our concern here, we will adopt a solution which puts the NM at the syntax-morphology interface (as 'external' verbal inflection) rather than in morphology proper. NM-prefixation is immune both to the internal morphological structure of a verb and syntactic operations, and we treat verbal *nie*-prefixation similarly to morphological cliticization in Romance, cf. Miller and Sag (1997) and Abeillé *et al.* (1998) (French) and Monachesi (1999) (Italian). We assume that the NM is attached to an inflected verb and it is a spelled-out of the morphosyntactic NEG attribute at the syntax-morphology interface, see Richter and Sailer (1999b).

Signature In order to account for the morphosyntax of Polish verbal negation, we have to enrich the signature introduced in §1.3.1. The modification presented in (4.44) is a simplified version of the signature assumed for morphology in Reinhard (2000) and Richter and Sailer (1999b).[26]

[26] Since phonology is not our main concern here, the value of PHON is oversimplified and represented as a list of *phon_string* objects. See Höhle (1999) for an elaborate architecture for phonology in HPSG.

(4.44)

$$\begin{bmatrix} \textit{sign} \\ \text{PHON } \textit{list(phon\_string)} \\ \text{SYNSEM } \textit{synsem} \end{bmatrix}$$

$$\begin{bmatrix} \textit{word} \\ \text{MORPH } \textit{morph} \end{bmatrix} \qquad \textit{morph} \qquad \textit{phrase}$$

As (4.44) shows, the MORPH attribute is appropriate to the sort *word*. The sort *morph*, a subsort of *sign*, serves to encode morphological structure of words. The sort *morph* is split into *simple-morph* (for basic non-recursive morphological units such as inflectional or derivational affixes and stems) and *complex-morph* (for more complex morphological entities such as compounds, inflected or derived forms), cf. (4.45).

(4.45)

$$\textit{morph}$$

$$\textit{simple-morph} \qquad \textit{complex-morph}$$

Therefore, both morphological affixes and stems are *sign*s on this approach. They are combined into larger units via morphological principles similar to those used in the syntax proper, see Reinhard (2000) for details.

The HPSG lexicon can be defined as a constraint on a minimal non-recursive sort. In Pollard and Sag (1994), this sort is *word*. Any *word* licensed by the grammar must satisfy one of the lexical descriptions, i.e., lexical entries, see §1.3.1. In the current extension of the signature, proposed in (4.44)–(4.45), the sort *word* is no longer a non-recursive minimal subsort of *sign*, see Höhle (1999). We follow Richter and Sailer (1999b) and assume that, in the present architecture, the lexicon consists of descriptions of non-recursive *simple-morph*s, which in turn serve as word-formation units.

A precise account of Polish morphology is far beyond the scope of this study. Below we present only basic intuitions behind an analysis of verbal morphology and we concentrate on the syntax-morphology interface, where the NM is attached.

Regularly Negated Verbs As stated in §4.4.1, verbs which can be regularly negated have the value of the NEG attribute unresolved in the lexicon. In the current architecture, the NEG value is set on the verbal stem, i.e., on the corresponding *simple-morph*. The relevant part of the lexical entry for the stem *lub-* of the verb *lubi* 'likes' in (2.75) is given in (4.46). (The description of the stem of the regularly negated verb *niepokoi* 'worries' in (4.17b) would be analogous.)

(4.46)
$$\begin{bmatrix} \textit{simple-morph} \\ \text{PHON } \langle \text{lub} \rangle \\ \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD } \begin{bmatrix} \textit{verb} \\ \text{NEG } \textit{bool} \end{bmatrix} \end{bmatrix}$$

We use the syntax-morphology interface, similarly to Richter and Sailer (1999b), to encode the spell-out of the NM *nie* at the word level. The relevant constraint is given in (4.47).

(4.47) $\textit{word} \rightarrow$

$$\begin{bmatrix}\textsc{phon}\ \langle\#\rangle \oplus \langle\ \textit{nie}\ \rangle \oplus \boxed{1} \oplus \langle\#\rangle \\ \textsc{morph}\ \begin{bmatrix}\textsc{phon}\ \boxed{1} \\ \textsc{synsem}\ \boxed{2}\left[\textsc{loc}|\textsc{cat}|\textsc{head}|\textsc{neg}\ +\right]\end{bmatrix} \\ \textsc{synsem}\ \boxed{2}\end{bmatrix} \vee$$
$$\begin{bmatrix}\textsc{phon}\ \langle\#\rangle \oplus \boxed{3} \oplus \langle\#\rangle \\ \textsc{morph}\ \begin{bmatrix}\textsc{phon}\ \boxed{3} \\ \textsc{synsem}\ \boxed{4}\left[\textsc{loc}|\textsc{cat}|\textsc{head}\ \neg[\textsc{neg}+]\right]\end{bmatrix} \\ \textsc{synsem}\ \boxed{4}\end{bmatrix}$$

The principle in (4.47) relates *morph*s to fully-fledged syntactic *word*s. We assume that the value of the MORPH attribute in (4.47) is an object of the sort *morph* which does not have any open morphological requirements.[27] This ensures that the NM is attached to an inflected verb rather than to an uninflected stem, (4.46). The first disjunct of the consequent accounts for the morphological spell-out of [NEG+] as the prefix *nie* (the NM precedes the phonology of the verb). The second disjunct is satisfied by those *word*s which do not have a [NEG+] specification, in particular they might have no NEG attribute at all. Remember that we do not treat lexical negation, (4.2a), on a par with verbal negation; NEG accounts only for the spell out of the NM. Since the NM is inserted at the syntax-morphology interface, *morph*s of [NEG+] verbs do not contain the NM. Therefore, multiple NM-prefixation does not result. The principle indicates also where phonological word boundaries (indicated as '#' in (4.47)) are inserted. Thus, a negated verb is treated as a syntactic (and prosodic, see (4.4a)) word and the NM is invisible for the syntax proper.[28]

Note also that the NM itself does not contribute to the semantics of the verb: the SYNSEM value of the negated verb comes from its morphological base, its MORPH value, which does not contain the NM. This is a desirable effect since, as argued in Przepiórkowski and Kupść (1999), the morphosemantic impact of the Polish NM is resolved at the *phrase* (clause) rather than *word* level. Since NEG is appropriate to a subsort of *head*, it is passed from *word* to the phrasal level via the syntactic Head Feature Principle, see §1.3.1.

Exceptions Verbs which cannot be morphologically negated, e.g., *proszę* '(I) ask' in (4.24a), are specified as [NEG−] in the lexicon, cf. (4.48). Therefore, the first disjunct of the constraint in (4.47) cannot describe the inflected form obtained from (4.48) and the verb correctly will not be negated. (Descriptions of *word* objects corresponding to weak auxiliaries are given in §2.5.6.)

(4.48) $$\begin{bmatrix}\textit{simple-morph} \\ \textsc{phon}\ \langle\text{prosz}\rangle \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}\ \begin{bmatrix}\textit{verb} \\ \textsc{neg}\ -\end{bmatrix}\end{bmatrix}$$

We assume that verbs such as *nie cierpię* '(I) detest' in (4.17a) always occur with the NM and are lexically specified as [NEG+]. *Nie* is not part of their stem and it is added to inflected forms via the general constraint proposed in (4.47). The only difference is that, unlike for regularly negated verbs, the stem of *nie cierpieć* 'to detest', i.e., *cierp-*, is always (and only) [NEG+], cf. (4.49).[29]

[27]In Reinhard's (2000) terms, these *morph*s have empty morphological requirements, i.e., [M-SPEC ⟨⟩] and [M-SUBCAT ⟨⟩].

[28]Remember that we do not consider *nie* in (4.38) to be the NM.

[29]The regularly negated *cierpieć* 'to suffer' mentioned in fn. 13 has a different meaning and will have a separate lexical entry with the NEG value left unspecified.

(4.49)
$$\begin{bmatrix} \textit{simple-morph} & \\ \textsc{phon}\ \langle \text{cierp} \rangle & \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head} & \begin{bmatrix}\textit{verb}\\ \textsc{neg}\ +\end{bmatrix}\end{bmatrix}$$

Since the stem in (4.49) is [NEG +], the verb will, correctly, never occur without the NM in syntax. This is ensured by the first disjunct of (4.47).

Finally, we have to account for negation of the existential/locative copula *być* 'to be/exist', cf. (4.6) and (4.7b). Inflection of this verb is irregular and all stem forms have to be listed in the lexicon. In (4.50), we give a partial description of its stem forms for different tenses.

(4.50)
$$\begin{gathered}\begin{bmatrix} \textit{simple-morph} & \\ \textsc{synsem}|\textsc{loc} & \begin{bmatrix}\textsc{cont}\ \textit{exist-be}\\ \textsc{cat}|\textsc{head}\ \textit{verb}\end{bmatrix}\end{bmatrix} \wedge \\ \left(\begin{bmatrix} \textsc{phon}\ \langle \text{są} \rangle & \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head} & \begin{bmatrix}\textsc{tense}\ \textit{pres}\\ \textsc{neg}\ -\end{bmatrix}\end{bmatrix} \vee \ldots \right. \\ \begin{bmatrix} \textsc{phon}\ \langle \text{był} \rangle \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}|\textsc{tense}\ \textit{past}\end{bmatrix} \vee \ldots \\ \left. \begin{bmatrix} \textsc{phon}\ \langle \text{będz} \rangle \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}|\textsc{tense}\ \textit{fut}\end{bmatrix} \vee \ldots \right)\end{gathered}$$

The verb can be negated and, apart from the present tense stems, the NEG value is unspecified in (4.50). Since the semantic impact of the NM is not resolved at the lexical level, the CONT value in (4.50) comprises semantics of both negated and non-negated forms. In the present tense the NM is added to the form of the verb *mieć* 'to have' rather than 'be'. The lexical entry in (4.51) (one of the disjuncts in (4.50)) ensures that the negated verb in the present tense gets the appropriate form.

(4.51)
$$\begin{bmatrix} \textit{simple-morph} & \\ \textsc{phon}\ \langle \text{ma} \rangle & \\ \textsc{synsem}|\textsc{loc} & \begin{bmatrix} \textsc{cat}|\textsc{head} & \begin{bmatrix}\textit{verb}\\ \textsc{neg}\ +\\ \textsc{tense}\ \textit{pres}\end{bmatrix} \\ \textsc{cont}\ \textit{exist-be} & \end{bmatrix}\end{bmatrix}$$

We also have to account for the fact that the nominative subject of *być* 'to be' is a genitive NP when the verb is negated, see (4.7b) and (4.9a). As observed in Dziwirek (1994, p. 154) and Przepiórkowski (1999, pp. 105ff), it is rather unlikely that the genitive NP should be still considered the syntactic subject of the negated verb. Unlike the nominative NP in (4.52a), NP[gen] cannot bind subject-oriented anaphoric pronouns in (4.52b) and a personal pronoun has to be used instead. (See Przepiórkowski (1999, pp. 106ff) for other arguments.)

(4.52) a. Jan$_i$ jest w swoim$_i$ / jego$_{*i}$ pokoju.
John$_{nom}$ is in self's his room
'John is in his (own) room.'

b. Jana$_i$ **nie** ma w jego$_i$ / *swoim pokoju.
John$_{gen}$ NM has in his self's room
'John is not in his room.'

We follow Czuba and Przepiórkowski (1995) and Przepiórkowski (1999) and assume that NP[gen] in (4.52b) is a complement rather than the subject of *nie ma*. In fact, we assume that all negated forms of the existential/locative *być* 'to be' do not have the syntactic subject and NP[gen] is a complement in (4.7b), (4.9a), (4.10a) and (4.52b). The constraint in (4.53) captures this fact.

(4.53) $$\begin{bmatrix} \textit{simple-morph} \\ \textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head}|\textsc{neg}\ + \\ \textsc{cont}\ \textit{exist-be} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{subj}\ \langle\rangle \\ \textsc{comps}\ \langle \text{NP}[\textit{gen}] \rangle \end{bmatrix} \end{bmatrix}$$

The third person neuter singular form of the negated verb in (4.7b), (4.9a) and (4.10a), follows from the general principle of subject-verb agreement in Polish proposed in Czuba and Przepiórkowski (1995). In Polish, if the subject of a finite verb is not a nominative NP, no morphological agreement holds and the verb occurs in the third neuter singular form. This is precisely the case in (4.53): the subject is empty and the verb is finite, thus, the third person singular neuter form of the verb is predicted. No additional specification on negated forms of *być* is necessary to obtain the correct verb form.

Prefixation of Negated Forms As we have mentioned in §4.2.3, there are two perfective verbs which are obtained via prefixation of negated verbs, i.e., *znienawidzić* 'to hate (perfective)' in (4.22a) and (colloquial) *znielubić* 'to dislike' in (4.23b). Since we do not treat the NM as a real morpheme, this fact seems problematic for our approach. Since the NM is added at the syntax-morphology interface, the negated verb is a *word* but it has no corresponding *morph*. Therefore, there is no *morph* the perfective prefix *z-* could be attached to (unlike in the case of *zniewolić* 'to imprison' (perfective) where *nie* belongs to the morphological base of the verb and a morph *niewolić* exists).

Before proposing a solution to this puzzle, let us note that *z*-prefixed forms have different properties than the plain negated verbs. As discussed in §4.2.2 and §4.2.3, both *nie lubić* 'not like' and *nienawidzić* 'to hate' trigger GoN. The prefixed forms, however, do not have this property: the NP complement appears in accusative, see (4.22a) and (4.23b), respectively. Moreover, these verbs can be overtly negated, (4.54), i.e., they behave as non-negated verbs do.

(4.54) Jan **nie** **znie**nawidził / ?**nie** **znie**lubił jeszcze Marysi.
John NM z-hated / NM z-disliked yet Mary$_{gen}$
'John has not hated / disliked Mary yet.'

Since the prefixation of negated forms is an exception rather than a rule, see (4.22) and (4.23), we treat these two verbs as idiosyncratic forms and apply the analysis of idioms developed in Sailer (2000).

In order to formally distinguish idioms from regularly obtained linguistic objects Sailer (2000) introduces the attribute COLL (Context of Lexical Licensing) appropriate to the sort *sign*. For internally regular non-idiomatic linguistic expressions, its value is the empty list. All principles of the grammar which account for regular combinations of signs are modified in such a way that they concern only regular signs ([COLL *e_list*]). Idiomatic signs which are internally irregular have a non-empty COLL list which makes them exempt from principles satisfied by regular signs. Irregular signs are specified in the grammar by the Lexicon Principle, Sailer (2000, p. 417), which enumerates all internally irregular linguistic expressions (IE$_i$) and lexical entries (L$_i$), see (4.55).

(4.55) $\begin{bmatrix} \textit{sign} \\ \textsc{coll}\ \ \textit{ne\_list} \end{bmatrix} \rightarrow (\mathrm{L}_1 \vee \ldots \vee \mathrm{L}_m \vee \mathrm{IE}_1 \vee \ldots \vee \mathrm{IE}_n)$

Although Sailer (2000) does not consider morphology and *sign* in his architecture has only two subsorts, *word* and *phrase*, irregular combinations are not uncommon in morphology and it is natural to assume that COLL is present also on *morph*s.

Having adopted these extensions, we treat *znienawidzić* 'to hate' (perfective) and *znielubić* 'to dislike' (perfective) as irregular *complex-morph*s: their COLL list is non-empty and they are some IE_i disjuncts in the consequent of (4.55) (in contrast to other regularly *z*-prefixed perfective verb forms). The internal structure of the 'saturated' *complex-morph* denoting the verb *znienawidzić* is schematically given in (4.56). (We assume here that the affix is a morphological head of the complex morph.)

(4.56)

$$\begin{bmatrix} \textit{complex-morph} \\ \textsc{phon}\ \langle \text{znienawidzić} \rangle \\ \textsc{coll}\ \ \textit{ne\_list} \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}|\textsc{neg}\ \ \textit{bool} \end{bmatrix}$$

$$\overset{\textbf{H}}{\begin{bmatrix} \textit{simple-morph} \\ \textsc{phon}\ \langle \text{z} \rangle \end{bmatrix}} \qquad \overset{\textbf{NH}}{\begin{bmatrix} \textit{word} \\ \textsc{phon}\ \langle \text{nienawidzić} \rangle \\ \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}|\textsc{neg}\ + \end{bmatrix}}$$

There are few comments to be made with respect to (4.56). First, note that the resulting *complex-morph* has the NEG value unspecified and the verb can be regularly negated, see (4.54). Second, since *znienawidzić* and *znielubić* are treated as exceptions, prefixation of negated verbs is blocked in the general case and the ungrammatical verbs as in (4.22b) and (4.23a) will not result. Finally, note that although *znienawidzić* is an irregular *complex-morph*, its component *nienawidzić* is a verb obtained from the [NEG+] stem *nawidz-* by the principle in (4.47) responsible for the spell-out of the NM in general.

Before completing this section, let us consider some possible alternatives. The most obvious alternative is to treat *znienawidz* as a new [NEG *bool*] stem, in addition to [NEG +] *nawidz*. In this case, however, we loose the intuition that *znienawidzić* is closely morphologically related to *nienawidzić*.

Assuming *nienawidz* to be a stem is also problematic. In order to prevent its NM-prefixation it should be [NEG −] (recall that (4.20a) exemplifies constituent negation rather than sentential negation). Although *nienawidz* would be a *morph* and could be the base for *z*-prefixation, the verb *nienawidzić*, also obtained from this stem, would be [NEG−] and could not trigger GoN, contrary to facts, see (4.17a) and (4.18).

The analysis of prefixed negated verbs as exceptions easily overcomes these problems.

4.5 Comparisons

In this section we compare our analysis with other HPSG approaches to morphological cliticization proposed recently for Romance pronominal clitics.

4.5.1 Miller and Sag (1997)

On the basis of arguments presented in Miller (1992a), French pronominal clitics are treated as morphological affixes attached to inflected verbs, see Miller and Sag (1997) and Abeillé *et al.* (1998). All inflected words are cross-classified into three dimensions: (1) CLITIC-REALIZATION, i.e., either *plain-word* (*pl-wd*) or *cliticized-word* (*cl-wd*), (2) INFLECTION specifying different inflection types, e.g., *3pl-pres-ind-vb*, *1sg-past-ind-vb*, and (3) LEXEME which specifies properties common to a family of inflected forms, e.g., the form of the stem, syntactic category, argument structure or meaning. The lexicon consists of *lexemes*, i.e., entities which belong to the LEXEME class. The morphological structure of an inflected word is encoded as the value of three attributes: STEM, I-FORM and FORM. In the case of plain words, the form of the word, i.e., the FORM value, is identical to its inflectional form, i.e., I-FORM, see (4.57).

(4.57) $\quad pl\text{-}wd \rightarrow \begin{bmatrix}\text{MORPH} \begin{bmatrix}\text{FORM } \boxed{1}\\ \text{I-FORM } \boxed{1}\end{bmatrix}\end{bmatrix}$

The correct FORM of the word is obtained by the morphological function which relates the STEM value of the lexeme to the appropriate inflected form, see Abeillé *et al.* (1998, p. 8).

On the other hand, words which have arguments realized as clitics, i.e., *cl-wd*, have the clitics added to the phonology of their inflectional form: their FORM value is a function of the I-FORM value and an appropriate combination of clitics which are realized as arguments.

Such an approach to morphological cliticization can be quite successfully applied to Polish verbal negation. The NEG value is specified for the whole inflectional family of a verb and it should be introduced for each verbal lexeme. For regularly negated verbs, the NEG value is left unspecified, i.e., a verb can be either negated or not, whereas verbs which cannot be regularly negated would have NEG set directly in their lexemes. More problematic is the existential/locative copula *być* 'to be' since in the present tense the verb has a suppletive negative form. The correct forms of the stem can be obtained, however, if we adopt further cross-classification of irregular lexemes, as in Koenig and Jurafsky (1994).

Then, the *pl-wd* type would correspond to non-negated verbs (and other non-cliticized words) whereas *cl-wd* would account for negated verbs. The relevant constraints are given in (4.58) and (4.59), respectively.

(4.58) $\quad pl\text{-}wd \rightarrow \begin{bmatrix}\text{MORPH} & \begin{bmatrix}\text{FORM } \boxed{1}\\ \text{I-FORM } \boxed{1}\end{bmatrix}\\ \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD} & \neg\left[\text{NEG}+\right]\end{bmatrix}$

(4.59) $\quad cl\text{-}wd \rightarrow \begin{bmatrix}\text{MORPH} & \begin{bmatrix}\text{FORM } \langle\text{nie}\rangle\oplus\boxed{1}\\ \text{I-FORM } \boxed{1}\end{bmatrix}\\ \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD}|\text{NEG} & +\end{bmatrix}$

The analysis of NM-prefixation we presented in §4.4 is in principle compatible with this approach. The constraints in (4.58) and (4.59) are analogs of our (4.47). Our account, however, is more parsimonious, e.g., we do not have to introduce distinct types for cliticized and plain words. The spell-out of the NM is encoded as a general constraint on *word*s and no subtypes of *word* are necessary.

The crucial difference is, however, the treatment of exceptional prefixed negated verbs, e.g., *znienawidzić* 'to hate' (perfective). In the analysis à la Miller and Sag (1997), such a verb would be treated as an independent lexeme (note a different subcategorization frame of the perfective, (4.23a), and the imperfective verbs, (4.17a)). Thus, the intuition that the perfective verb is morphologically related to the negated verb would be lost.

4.5.2 Monachesi (1999)

Italian pronominal clitics are also analysed as morphological affixes, cf. Monachesi (1995, 1999). As in Miller and Sag (1997), Italian cliticization is treated similarly to inflection. The difference between cliticization and inflection is that pronominal clitics take an inflected *word* as the STEM value whereas inflectional affixes are attached to a bare verbal *stem*, Monachesi (1999). Unlike in Miller and Sag (1997), however, no morphological function is employed and realization of clitics is encoded as the value of the AFFIX attribute. If Polish verbal negation is put into this setup, the NEG value would be introduced for verbal *stem*s. Thus, regularly negated verbs would have the NEG unspecified while exceptions would be marked directly in the lexicon. The suppletive negative form of *być* ‘to be’ would be correctly obtained if its stems are given analogously to (4.50) and (4.51).

A schematic analysis of the regularly NM-prefixed verb *nie lubi* ‘(s)he does not like’ is given in (4.60).

(4.60)
$$\begin{bmatrix} \textit{word} \\ \text{PHON } \boxed{2} \oplus \boxed{1} \\ \text{MORPH} \begin{bmatrix} \text{STEM} \begin{bmatrix} \textit{word} \\ \text{PHON } \boxed{1}\langle\text{lubi}\rangle \\ \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{HEAD}|\text{NEG } + \end{bmatrix} \\ \text{AFFIX} \begin{bmatrix}\text{PHON } \boxed{2}\langle\text{nie}\rangle\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Monachesi’s (1999) framework allows one to express certain morphological generalizations. For example, a general constraint can capture the fact that in prefixed forms, the AFFIX|PHON value precedes the STEM|PHON value. Therefore, the PHON value of the negated verb would follow from this general principle and needn’t be specified explicitly in (4.60).

Note that (4.60) is recursive: the value of STEM is an object of the sort *word*, just like objects in the denotation of the entire description. This, in fact, allows us to obtain the verb *znienawidzić*. If the inflected word *nienawidzić* is used as the STEM and the perfective prefix *z* as the AFFIX value, the prefixed verb form would be obtained. At first glace this seems to be a desirable result. Remember, however, that prefixation of negated verbs is an exception rather than a rule. Thus, it is not clear how to restrict such a mechanism just to a few exceptions and prevent overgeneration in the general case, see the ungrammatical example in (4.22). Moreover, since a negated verb can be used as the STEM value, nothing prevents it from serving again as a host for the NM and multiple NM-prefixation (incorrectly) is not blocked. Unfortunately, apparent virtues of this approach turn out to be its weaknesses.

4.6 Conclusions

In this chapter we have provided a morphosyntactic account of Polish verbal negation. We have discussed two possible ways of formalizing the analysis in the framework of HPSG. We have considered a more ‘standard’ approach based on a lexical rule. In such an analysis, however, exceptions cannot be easily handled. We provide a different account based on underspecification. The advantage of this solution is that only irregular forms have to be idiosyncratically marked whereas no specific marking is required for regularly negated verbs. The constraint-based approach avoids the redundancy of the lexical rule and overcomes more general problems (learnability) of the latter approach. We deal with morphosyntactic aspects of verbal negation in a way similar to Romance cliticization. The NM is added to an inflected verb at the syntax-morphology interface. The NM is never part of a morphological base which allows us to uniformly capture the spell-out of the NM both in regularly and irregularly negated verbs. We account for perfective prefixation of negated verbs treating them as exceptions and applying the formal analysis of idioms proposed in Sailer (2000). Such an account allows us to retain links with the morphological components and simultaneously encode the idiosyncratic properties of relevant words.

Chapter 5

Implementation

The aim of this chapter is twofold: first, to verify the coherence of (a fragment of) the formal linguistic theory developed for Polish clitics in previous chapters and second, to test ALE as a platform for encoding HPSG grammars.

Although the implementation presented in this chapter is not directly part of the ongoing project on developing an ALE implementation of a large HPSG grammar of Polish carried out at the Institute of Computer Science, Polish Academy of Sciences, Warsaw, cf. Bolc *et al.* (2000), we have chosen the same computational platform to facilitate the incorporation of our grammar to this project.

There are several systems available for implementing HPSG-style grammars, see e.g., Backofen *et al.* (1996) and Bolc *et al.* (1996) for an overview of some of the systems. ALE has been chosen as a platform for the implementation in the project for the following reasons. First, ALE provides an implementation of a version of a typed feature structure logic which can be used for HPSG.[1] Second, certain aspects of HPSG can be encoded directly in ALE, e.g., a signature with a multi-inheritance type hierarchy, type constraints. Third, ALE has relatively good termination properties and parsing efficiency, in contrast to, e.g., TFS, Emele (1993, 1994), or ConTroll cf. Götz *et al.* (1997). From this perspective, the choice of ALE is a reasonable trade-off between the theory and practice. Other reasons for choosing ALE were mostly technical, e.g., the availability of the system or the software it requires.

After a brief presentation of ALE, we provide a description of the implemented grammar. The code of the grammar and a set of test sentences can be found in Appendix A.

5.1 ALE

The Attribute Logic Engine (ALE, Carpenter and Penn, 1999), created at Carnegie Mellon University, is a definite clause logic programming system integrated with a phrase structure parser and semantic head-driven generator. Terms in ALE grammars and in logic programs are typed feature structures as defined in Carpenter (1992), a typed version of Kasper-Rounds logic, cf. Kasper and Rounds (1986).

In this section, we briefly present these aspects of ALE which are necessary for understanding the implementation of the grammar presented in §5.2. (See Carpenter and Penn (1999)

[1]See Meurers (1994) for a discussion of differences between the architecture of HPSG in Pollard and Sag (1994) and ALE in Carpenter and Penn (1995).

for a full description of the system; for comparisons of earlier versions of ALE (cf. Carpenter and Penn (1995)) with other systems see Backofen *et al.* (1996) and Bolc *et al.* (1996).)

5.1.1 Types

ALE is a language with strong typing, which means that every object of this language (a feature structure) must have a type. Types are arranged in an inheritance hierarchy declared by the user. The transitive closure of the subtyping relation should form a *partial order* (the subtyping relation must be transitive and anti-symmetric) and be *bounded complete* (every pair of types which have a common subtype must have a unique most general common subtype). There is a special pre-defined type `bot`, which is the unique most general type: every type in the hierarchy is a subtype of `bot`. A type specification consists of the type name, followed by the keyword `sub`, and a (possibly empty) list of its subtypes. An example of the type specification for the type `person` is given in (5.1).

(5.1)
```
person sub [first, second, third].
  first sub [].
  second sub [].
  third sub [].
```

ALE is based on an open world interpretation of the type hierarchy, i.e., there may be objects in the denotation of a non-minimal type which are not in the denotation of any of the minimal (maximally specific) subtypes of this type.

5.1.2 Feature Structure Descriptions

In ALE, descriptions of feature structures can consist of type names (Prolog constants), (Prolog) variables, selection (<*feature*> `:` <*description*>), conjunction (<*desc*> `,` <*desc*>) or disjunction (<*desc*> `;` <*desc*>) of descriptions.[2] The conjunction of two descriptions is interpreted as unification of feature structures. ALE does not support general negation of descriptions.

Feature structures in ALE are *totally well typed*, Carpenter and Penn (1999, p. 14). This condition of feature appropriateness forces each type to specify for which features it can be defined and what values such features can take. Appropriateness specifications are inherited by a type from its supertypes. The specification of the type `sign_list` is given in (5.2).

(5.2)
```
sign_list    sub   [e_list, ne_sign_list].
ne_sign_list sub   []
             intro [hd: sign,
                    tl: sign_list].
```

The type declaration in (5.2) says that a list of signs can be either empty (`e_list`) or non-empty (`ne_sign_list`). There are two features appropriate to the type `ne_sign_list` (introduced by the `intro` operator): the head (`hd`) and the tail (`tl`) of the list. Every feature structure of the type `ne_sign_list` must have these features and their values must be compatible with the declaration. The value of `hd` must be of the type `sign` whereas the `tl` value must be of the type `sign_list`.

[2]Additionally, inequations (`=/=` <*desc*>) or path equations (<*path*> `==` <*path*>) are available, cf. Carpenter and Penn (1999, ch. 3.8), but we do not employ them in our implementation.

The appropriateness conditions in ALE cannot be cyclic, i.e., the value of a feature appropriate to a given type cannot be of the same or more specific type. This means that the specification given in (5.3) is not allowed; this specification would lead to non-terminating inference as there is no most general cyclic finite feature structure.

(5.3)
```
person sub [male,female]
  intro [father:male,
         mother:female].
    male sub [].
    female sub [].
```

The acyclicity condition on the signature does not rule out recursive structures such as lists, cf. (5.2). In (5.2), the `tl` value is not introduced for the type `sign_list` but for its subtype `ne_sign_list`.

The second condition on appropriateness, the feature introduction condition, states that every feature has to be introduced at a unique most general type (bounded completeness ensures that the set of types a given feature is appropriate to must have a most general element). The following type declaration is invalid in ALE, cf. Carpenter and Penn (1999, p. 15).

(5.4)
```
a sub [b,c,d].
  b sub []
    intro [f:w,
           g:x].
  c sub []
    intro [f:y,
           h:z].
  d sub [].
```

The problem with this declaration is that the feature `f` is appropriate to `b` and `c` but there is no most general type to which it is appropriate. The types `b` and `c` have a common supertype (namely `a`) but `f` is not appropriate to this type. An obvious solution is to introduce an intermediate type `e` (an immediate supertype of `b` and `c`) for which `f` is introduced. This modification of (5.4), consistent with the feature introduction condition, is given in (5.5) (see Carpenter and Penn (1999, p. 16)).

(5.5)
```
a sub [d,e].
  e sub [b,c]
    intro [f:bot].
  b sub []
    intro [f:w,
           g:x].
  c sub []
    intro [f:y,
           h:z].
  d sub [].
```

As (5.5) illustrates, the value of a feature introduced for the type can be more general than its value specified for the subtypes: the value of `f` introduced at `e` is a supertype of the values declared for `b` and `c` (subtypes of `e`).

5.1.3 Type Constraints

In ALE, it is possible to specify constraints on types expressing restrictions on feature structures of the type. Constraints are introduced by the `cons` operator and may consist of arbitrary descriptions. Type constraints may include definite clause attachments introduced by the operator `goal`.

The fact that every headed phrase in HPSG must satisfy the Head Feature Principle can be encoded as a relational constraint in (5.6).

(5.6)
```
phrase cons P goal hfp(P).
```

In the example, the constraint is expressed as a variable `P` which matches every feature structure of the type `phrase`. This feature structure is used as an argument of the predicate `hfp`.

5.1.4 Macros

Abbreviations for ALE descriptions can be defined by means of macros. They cannot contain calls to definite clauses and cannot be recursive but they can be parameterized. Macro definitions have the following form:

(5.7) *<macro_name>*(*<parameters>*) `macro` *<description>*.

An example of a macro which describes a synsem object of a noun is given in (5.8).

(5.8)
```
noun(Case) macro
  loc:cat:(head: case: Case,
           subj: [],
           comps: []).
```

The case value of the noun is used as a parameter in the macro `noun`.

The macro call begins with the '`@`' operator. The macros allowing for the traditional Prolog list notation are built-in and are called without `@`, e.g., `[one, two]` or `[Head|Tail]`.

5.1.5 Lexical Entries

Lexical entries in ALE have the following form:

(5.9) *<word>* `--->` *<description>* .

Macros are especially useful to compactly express large lexical descriptions and keep the specification consistent across lexical entries.

A lexical entry for the word `john`, which employs the macro in (5.8), is given in (5.10).

(5.10)
```
john --->
   word, synsem: @noun(nom).
```

The lexical entry for `john` in (5.10) is equivalent to the expanded form in (5.11).

(5.11)
```
john --->
   word,
   synsem:loc:cat:(head: case: nom,
                   subj: [],
                   comps: []).
```

ALE provides a special notation to encode lexical rules. Since we do not use this mechanism in the grammar described in §5.2, it is not discussed here (but see Carpenter and Penn (1999§5.3)).

5.1.6 Grammar Rules

Grammar rules in ALE are similar to Prolog DCG phrase structure rules with procedural annotations allowed. In contrast to DCGs, categories used in ALE rules are descriptions of features structures. Daughter categories of the rule are always evaluated in the order they are specified, from left to right.

Each rule consists of a mother category, at least one daughter category (introduced by the `cat>` operator) and optional procedural attachments expressed as ALE definite clauses (after the `goal>` operator). The operator `cats>` represents a list of daughters of unspecified length.

A possible ALE emulation of the traditional HPSG's SCHEMA2 is given in (5.12).[3]

(5.12)
```
schema2 rule
  (phrase,
   synsem:loc:cat:subj: Subj,
   head_dtr: HeadDtr,
   comp_dtrs: Comps)
===>
 cat> (HeadDtr,
       word,
       synsem:loc:cat:(subj: Subj,
                       comps: WComps)),
 goal> synsem_to_phrases(WComps, Comps),
 cats> Comps.
```

The mother category of the rule in (5.12) is a phrase whose head and complement daughters (`head_dtr` and `comp_dtrs`, respectively) are two daughter categories specified in the body of the rule (the '`===>`' operator separates the head and the body of the rule). The procedural attachment `synsem_to_phrases` takes the list of synsems (`WComps`) and maps it to a list of phrases (`Comps`). A definition of an analogous predicate used in our grammar is given in §5.2.4.

5.1.7 Definite Clauses

The syntax of ALE's definite clauses slightly differs from that used in Prolog. Every clause in ALE is required to have a body, which might be simply the goal `true`, introduced by the operator `if`.

ALE definite clauses which define the `append` predicate are given in (5.13).

(5.13)
```
append([], X, X) if true.

append([H|R], L, [H|L1]) if
  append(R, L, L1).
```

[3]Grammar rules in ALE are order-sensitive, unlike immediate dominance schemata in HPSG. We elaborate on this difference in §5.2.

The definite clause programs allow conjunction, disjunction, negation and cut specified with Prolog syntax.

A sequence of goals in ALE is evaluated depth-first, from left to right according to the order of clauses in the program.

5.1.8 Computational Aspects

ALE is implemented in Prolog and a SICStus, SWI Prolog or Quintus Prolog compiler is required. For our grammar, version 3.2.beta of ALE runing with SICStus 3.8.3 was used.

For parsing, ALE uses a bottom-up active chart parser Carpenter and Penn (1999, p. 55) built-in to the system: grammar rules are evaluated from left to right with the mother description coming last. The edges in the chart are filled in from right to left. The parser moves through the string from right to left, one word at a time. For example, in the string `john loves mary`, the lexical entry for `mary` is checked first and entered to the chart.

Rules can incorporate definite clause goals before, between or after category specifications. The goals are evaluated when they are found. For example, the goal in (5.12) is evaluated after the first category is found but before the second one is.

The parser terminates after all of the inactive edges derivable from lexical entries and grammar rules have been found. If an infinite number of derivations can be produced by the grammar, the parser will not terminate.

ALE allows for incremental compilation of the type hierarchy, type constraints, the attribute-value logic and definite clauses. ALE provides tools for debugging but there is no sophisticated graphical user interface (only an Emacs-based interface is available). Hdrug, cf. van Noord (1999), can be adapted as a front end for ALE.[4]

5.2 Implementation of Clitic Climbing in ALE

In this section we describe the ALE implementation of Clitic Climbing (CC) in verb clusters (VCs) presented in Chapter 2.

Let us recall the main features of the HPSG analysis.

- Polish pronominal clitics, as well as the reflexive marker (RM) *się*, are syntactic items;
- they are distinguished from other subcategorized arguments of a verb via the distinct *synsem* value;
- verbs which may occur with the RM are idiosyncratically marked as such in the lexicon (in contrast to pronominal clitics);
- the relative order of clitics (as well as other NPs) is not fixed;
- CC is encoded in lexical entries of the verbs which subcategorize for a bare infinitive complement;
- CC in Polish is optional, clitics do not have to be realized on the same verb and they can climb in any order;

[4]For earlier releases of ALE, also Pleuk, cf. Calder and Humphreys (1995), was available as a graphical user interface.

- the realisation of clitics is licensed by syntactic principles and immediate dominance schemata slightly modified with respect to those in Pollard and Sag (1994):
  - the immediate dominance schema which combines complements with the lexical head of the phrase allows any number of its complements to be attached to the head;
  - the Valence Principle allows complements to be realised in any order;

The following subsections present an ALE encoding of the above theoretical issues. Throughout this section, the terms *sort* and ATTRIBUTE are used to refer to HPSG-theoretical concepts whereas the terms `type` and `feature` are reserved for their ALE equivalents.

5.2.1 Signature

In order to distinguish clitics from non-clitics we have split the sort *synsem* into subsorts, see (2.39) repeated as (5.14).

(5.14)

synsem

non-canonical *canonical*

non-clitic *clitic*

The translation of this sort hierarchy into ALE's type specification is given in (5.15).[5]

(5.15)
```
synsem_t sub [clitic_t, non_clitic_t]
  intro [loc: local_t].

  clitic_t sub [].
  non_clitic_t sub [].
```

Since we do not consider subsorts of *non-canonical*, for simplicity we ignore the distinction of *non-canonical* vs. *canonical* subsorts. In (5.15), `clitic_t` and `non_clitic_t` are immediate subtypes of `synsem_t`. The feature `loc` is appropriate to the type `synsem_t` and its subtypes. Therefore, every ALE feature structure of type `clitic_t` or `non_clitic_t` will appear with this feature as well. The value of `loc` is specified to be `local_t`.

5.2.2 Type Constraints

Implicational HPSG constraints with a sort antecedent can be expressed in ALE as type constraints with optional procedural attachments, see §5.1.3.

The implicational HPSG constraint given in (5.16) is represented in ALE via the relational constraint in (5.17).[6]

(5.16) $\textit{phrase} \rightarrow$

$$\begin{bmatrix} \textsc{synsem} & \textit{non-clitic} \\ \textsc{dtrs}|\textsc{head-dtr}|\textsc{synsem} & \textit{non-clitic} \end{bmatrix} \wedge$$

[5] For convenience, we use the '`_t`' to distinguish names of types from occasionally homophonous names of features, e.g., the feature `synsem` vs. the type `synsem_t`.

[6] We restrict ourselves only to headed phrases and we encode the constraint in (5.17) as valid for all *phrase*s.

$$
\begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}\ \boxed{1} \\ \textsc{dtrs}|\textsc{head-dtr}|\textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{head}\ \boxed{1} \end{bmatrix} \wedge
$$

$$
\left(\begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{marking}\ \boxed{2} \\ \textsc{dtrs}\ \begin{bmatrix} \textit{head-comp-str} \vee \textit{head-adjunct-str} \\ \textsc{head-dtr}|\textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{marking}\ \boxed{2} \end{bmatrix} \end{bmatrix} \vee \right.
$$

$$
\left. \begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{marking}\ \boxed{3} \\ \textsc{dtrs}\ \begin{bmatrix} \textit{head-mark-str} \\ \textsc{marker-dtr}|\textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{marking}\ \boxed{3} \end{bmatrix} \end{bmatrix} \right) \wedge \ldots
$$

(5.17)
```
phrase_t cons P goal (non_clitic_phrases(P),
                      hfp(P),
                      marking_principle(P)).
```

(5.18)
```
non_clitic_phrases((synsem: non_clitic_t,
                    dtrs:hd_dtr:synsem: non_clitic_t)) if true.

hfp((synsem:loc:cat:head:H,
     dtrs:hd_dtr:synsem:loc:cat:head:H)) if true.

marking_principle((dtrs:(hd_mk_t,
                         mk_dtr:synsem:loc:cat:marking:X),
                   synsem:loc:cat:marking:X ;
                   dtrs:((hd_comps_t;
                          hd_adj_t),
                          hd_dtr:synsem:loc:cat:marking:X),
                   synsem:loc:cat:marking:X
                  )) if
   true.
```

The constraint in (5.17) enforces any feature structure of the type `phrase_t` to satisfy the conjunction of predicates enumerated after the `goal` operator. Definite clauses, which define these predicates, are given in (5.18).

The predicate `non_clitic_phrases/1` states that every (headed) phrase has a *non-clitic* as the value of its head daughter's SYNSEM.[7] As discussed in §2.1.3 and §3.1.3, pronominal clitics and the RM cannot be modified or conjoined, unlike other NPs. As we will see shortly, we make clitics illegitimate for schemata which 'project' a lexical head to a phrase. Due to these restrictions, pronominal clitics will always appear in syntactic structures in the form taken directly from the lexicon, i.e., as *word*s.

The predicates `hfp/1` and `marking_principle/1` encode, respectively, the Head Feature Principle and the Marking Principle of Pollard and Sag (1994). The definite clause which specifies the `marking_principle/1` predicate uses (twice) disjunction in the description of its argument. The definition of this predicate is logically equivalent to three (trivially satisfied) definite clauses introduced separately for `hd_mk_t`, `hd_comps_t` and `hd_adj_t`. Using disjunctions in descriptions, however, is more efficient than using multiple clauses in ALE, cf. Carpenter and Penn (1999, p. 41).

[7]The arity of a predicate is indicated in the Prolog manner, e.g., `non_clitic_phrases/1` is a one-place predicate.

5.2.3 Lexical Entries

In our grammar, only a toy lexicon of fully inflected word forms has been used.[8]

The lexical entries for the pronominal clitic *go* 'him' and the reflexive marker (RM) *się* 'self', which employ the macros in (5.19), are exemplified in (5.20).

(5.19)
```
nps(Case, Index) macro
  (synsem_t,
  loc:(cat:(head: case: Case,
            subj: [],
            comps: [],
            marking: unmarked_t),
       cont:index: Index)).
thsg(Gend) macro
  person: third_t,
  number: sing_t,
  gender: Gend.
rm macro
  clitic_t,
  loc:(cat:(head: case: (acc_t; gen_t),
            subj: [],
            comps: [],
            marking: unmarked_t),
       cont: ana_t).
```

(5.20)
```
go ---> word_t, synsem: (clitic_t, @nps((acc_t; gen_t), @thsg(masc_t))).

sie ---> word_t, synsem: @rm.
```

The lexical entry for the reflexiva tantum verb *boi się* 'is afraid' which subcategorizes for the RM *się* and an NP[gen] is given in (5.22); the macros are specified in (5.21).

(5.21)
```
noun(Synsem, Case) macro
 (Synsem,
  loc:cat:(head: case: Case,
           subj: [],
           comps: [],
           marking: unmarked_t)).
non_rm(Case) macro
   (@npnc(Case, _);
   (@noun(clitic_t, Case),
    loc:cont: ppro_t)).
```

(5.22)
```
boi --->  word_t,
          synsem:loc:cat:(head:vform:fin_t,
                          subj: [@noun(non_clitic_t, nom_t)],
                          comps: [@rm, @non_rm(gen_t)],
                          marking: unmarked_t).
```

[8]See, however, Czuba (1995a) for an implementation of a lexeme-based lexicon of Polish in DATR (cf. Evans and Gazdar (1989, 1990)), and Czuba (1995b) for its integration with ALE.

As shown in §3.5.1, a verb may occur with at most one RM. In our theory, this has been forced by two implicational constraints, cf. (3.103) and (3.104), which restricted the COMPS list to contain at most one RM.

(3.103) $$word \rightarrow \left[\text{SYNSEM}|\text{LOCAL}|\text{CATEGORY}\ \left[\begin{array}{l} \textit{category} \\ \text{VALENCE}\ \left[\begin{array}{l} \textit{valence} \\ \text{SUBJ}\ \boxed{1} \\ \text{COMPS}\ \boxed{2} \end{array}\right] \\ \text{ARG-ST}\ \boxed{1} \oplus \boxed{2'} \end{array}\right]\right] \wedge \texttt{arg2comps}(\boxed{2}, \boxed{2'})$$

(3.104) $$\forall \boxed{1}\ \forall \boxed{2}\ \Big(\big([\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{SUBJ}\ \boxed{1}] \wedge \texttt{member}(\boxed{2}, \boxed{1}) \big) \rightarrow \neg(\boxed{2}\text{RM}) \Big)$$

Since in ALE constraints can be imposed on types, cf. §5.1.3, but not on descriptions of feature structures, principle (3.104) cannot be expressed directly in ALE. Instead, we exclude the multiple occurrence of the RM in every relevant lexical entry. In (5.22), the NP[gen] complement is specified (using the macro `non_rm`) to be a non-RM element.

As discussed in §2.5.1, not all verbs which form a VC accept CC. The possibility of CC is encoded in lexical entries, cf. §2.5.3. For VC-verbs which allow CC, we assume that the main verb inherits all unrealized complements of its infinitive argument. The lexical entry for the subject control verb *może* 'may' (which permits CC, see §2.5.2), is given in (5.24).

(5.23)
```
inf_comp(Index, List) macro
  non_clitic_t,
  loc:cat:(head:vform:inf_t,
           subj: [@nps(_, Index)],
           comps: List,
           marking:unmarked_t).
```

(5.24)
```
moze ---> word_t,
          synsem:loc:cat:(head:vform:fin_t,
                          subj:[@nps(nom_t, X)],
                          comps:[@inf_comp(X, List) | List],
                          marking:unmarked_t).
```

On the other hand, VC-verbs which do not allow for CC, can inherit only non-clitics. This fact is expressed directly in lexical entries of the verbs, see (5.25) for the verb *przeszkodził* '(he) disturbed' which does not accept CC; an HPSG encoding is given in (2.82) on p. 66.

(5.25)
```
przeszkodzil ---> word_t,
      synsem:loc:cat:(head:vform:fin_t,
                      subj:[@noun(nom_t)],
                      comps:([@nps(dat_t, X),
                              @inf_comp(X, (List, non_clitic_list)) | List]),
                      marking:unmarked_t).
```

Note that, unlike in (5.22), the length of the COMPS lists in (5.24) and (5.25) is not specified. This is because of the variable `List` which may be of an arbitrary length. Such lexical descriptions are satisfied by an infinite number of feature structures and, if no restrictions are given, the parser would not terminate, see §5.1.8. This issue will be discussed in more detail in the next subsection (see also Carpenter and Penn (1999, p. 54)).

5.2.4 Phrase Structure Rules

So far, the ALE implementation was able to follow the HPSG theory of CC in VCs developed in §2.5.3 rather closely. The crucial difference, however, appears in the encoding of the phrase structure. In ALE, grammar rules are order sensitive, unlike immediate dominance schemata in HPSG. ALE does not have a linear component yet and the linear order has to be encoded differently than in HPSG.[9] Thanks to procedural annotations, however, ALE can imitate ID schemata.

In order to account for CC, we have modified some of the ID schemata of Pollard and Sag (1994), see (2.87) repeated below.

(2.87) $$\textsc{Schema2'} \equiv \begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat}|\textsc{val}|\textsc{comps} & \textit{list(synsem)} \\ \textsc{dtrs} & \begin{bmatrix} \textit{head-comp-struc} & \\ \textsc{head-dtr} & \textit{word} \\ \textsc{comp-dtrs} & \textit{list(phrase)} \end{bmatrix} \end{bmatrix}$$

The ALE rule which combines the lexical head with its complements (as in (2.87)) and additionally captures postverbal linear positions of clitics is given in (5.26).

(5.26)
```
(M, phrase_t,
    dtrs: (hd_comps_t,
           hd_dtr: HDtr,
           non_hd_dtrs: Signs))
 ===>
 cat> (HDtr, word_t,
        synsem:loc:cat:comps:Comps),
 goal> (at_most_three(Comps),
        shuffle((SubComps, ne_list), Rest, Comps),
        split(SubComps),
        synsems_to_signs(SubComps,Signs)),
 cats> Signs,
 goal> (comps(M, Rest),
        subj_val(M, [], HDtr)).
```

The body of the rule consists of two daughter categories. The first one, indicated by the variable `HDtr`, represents the lexical head of the phrase. The second one, `Signs`, is a list of an arbitrary length (it follows the operator `cats>`, see §5.1.6). The procedural attachments, present after each daughter category, are discussed below.

Let us have a closer look at the first sequence of goals. Definitions of the corresponding definite clauses are given in (5.27).

(5.27)
```
at_most_three((tl: ([];
                    (tl: ([];
                     tl:[]))) )) if true.

shuffle([X|Xs], R, Ys) if
  select(X, Ys, Ys1), shuffle(Xs, R, Ys1).
```

[9] The ongoing project on a prototype of a linearization parser integrated with ALE is reported in Fouvry and Meurers (2000). At the moment of writing this thesis, however, the parser is still in the experimental stage.

```
shuffle([], Ys, Ys) if true.

select(Z, [Z|Xs], Xs) if true.
select(Z, [Y|Ys], [Y|Zs]) if
  select(Z, Ys, Zs).

split(L) if
  append(clitic_list, non_clitic_list, L).

append([], X, X) if true.
append([H|R],L,[H|L1]) if append(R,L,L1).

synsems_to_signs([],[]) if  true.
synsems_to_signs([Syns|SynsList],[(sign_t, synsem:Syns)|SignList]) if
  synsems_to_signs(SynsList,SignList).
```

The predicate `at_most_three/1` ensures that `Comps` is a finite list. As we have remarked above, verbs which inherit arguments may have COMPS of an arbitrary length (see (5.24)). In order to allow the program to terminate, the length of this list must be limited (and is restricted here to three). Note that the empty list is not included in the definition of `at_most_three/1`. Therefore, only words with a non-empty COMPS value will be subject to the `schema2` rule. Words which have the empty COMPS value (e.g., clitics, proper names, intransitive verbs, etc.) will not be projected to phrases via this rule and their dual analysis (i.e., as either *word*s or as *phrase*s) will be disambiguated to be *word*s. For this reason, we do not require non-head daughters in (5.26) to be *phrase*s (as in (2.87)) but we assume that they can be objects denoted by any subsort of *sign*, as specified by the `synsems_to_signs/2` predicate.

ALE grammar rules are order-sensitive, see §5.1.6, and categories are realized in the order they appear in the rule. According to (5.26), the head daughter (`HDtr`) will precede the list of non-head daughters (`Signs`). Therefore, we account only for these orders within a VP where the verb is followed by its complements (for a general outline of Polish linear order see §2.4.2). In Polish, there are no restrictions on the relative order of postverbal non-clitic complements but a postverbal clitic can be separated from the verb only by another clitic, see (2.28) and §2.2.5.

(2.28) a. Jan wytłumaczył mu go cierpliwie.
John explained him$_{cl}$ him$_{cl}$ patiently
'John patiently explained it to him.'

b. * Jan wytłumaczył jemu go cierpliwie.
John explained him him$_{cl}$ patiently
'John patiently explained it to him.'

c. * Jan wytłumaczył [nowemu koledze]$_{NP}$ go cierpliwie.
John explained new colleague him$_{cl}$ carefully
'John patiently explained it to a new colleague.'

As argued in §2.5.2, CC is syntactic, i.e., clitics are raised in syntax to a governing verb in VCs. Moreover, CC in VCs is optional and clitics may raise in any order and quantity. In order to account for these facts, we employ the predicate `shuffle/3`. This predicate takes an arbitrary ordered subset from the `Comps` list in (5.26): all possible permutations of all lists which consist of elements of `Comps`' subsets are obtained. The resulting `SubComps` list contains these arguments which are to be realised whereas `Rest` is a list of remaining arguments. In order to reduce the number of solutions in case of CC, we additionally require `SubComps` to be a non-empty list: at least one complement must be realised to form a phrase with its lexical head (`at_most_three/1` ensures that the COMPS list is non-empty). The correct order of postverbal clitic vs. non-clitic arguments is obtained via the predicate `split/1` which splits `SubComps` in such a way that postverbal clitic arguments precede non-clitics. As raised clitics are present on the 'higher' COMPS list, cf. (5.24), the same restriction holds for linear positions of clitics in VCs as well: only immediately postverbal linear positions of clitics are taken into account in our ALE implementation. Note that, unlike in §2.4.3 and §2.5.4, we do not propose here a general account of the linear order of clitics. Nevertheless, the implemented grammar captures the most crucial restriction: immediately postverbal positions of clitics.

Finally, the predicate `synsems_to_signs/2` maps *synsem*s (elements of the `SubComps` list) to corresponding *sign*s (in §1.3.1, a similar relation is used in the Valence Principle). As mentioned above, *synsem*s do not necessarily correspond to *phrase*s: clitics as well as words which do not have complements are not projected to phrases.

The two predicates attached after the second `goal>` in (5.26) have the effect of the Valence Principle: they ensure that realised arguments of the head daughter are not present on the SUBJ and COMPS lists of the mother.[10]

(5.28)
```
comps((synsem:loc:cat:comps: M), M) if true.

subj_val((synsem:loc:cat:subj: M), S, (synsem:loc:cat:subj: H)) if
  append(S, M, H).
```

Since not all complements must be realized, the list `Comps` does not have to equal `SubComps`. The list of arguments which remain unrealised, `Rest`, is computed by the predicate `shuffle/3` in (5.26). We do not compute this value again and use it directly as the second argument of `comps/2` (the COMPS value of the mother) in (5.26).

In §2.5.2, we also slightly modified the ID SCHEMA1 of Pollard and Sag (1994) which combines the head daughter with the subject. Since we allow phrases to have unrealised complements and not all words are projected to phrases, we additionally assume here that both daughters in this schema are *sign*s (rather than just *phrase*s as in Pollard and Sag (1994) and in §2.5.2). We require, however, that the COMPS list of the head daughter is empty. The modified SCHEMA1 and its ALE equivalent are given in (5.29) and (5.30), respectively.

(5.29) $$\text{SCHEMA1}'' \equiv \begin{bmatrix} \text{SYNSEM}|\text{LOC}|\text{CAT}|\text{VAL}|\text{SUBJ}\ \langle\rangle \\ \text{DTRS}\ \begin{bmatrix} \textit{head-subj-struc} \\ \text{HEAD-DTR}|\text{SYNSEM}|\text{LOC}|\text{CAT}|\text{VAL}|\text{COMPS}\ \langle\rangle \\ \text{SUBJ-DTR}\ \langle \textit{sign} \rangle \end{bmatrix} \end{bmatrix}$$

(5.30)
```
schema1 rule
(M,
 synsem:loc:cat:comps: (X, []),
 dtrs: (hd_comps_t,
```

[10]For a discussion of some issues concerning encoding the Valence Principle in ALE, see Meurers (1994, p. 48).

```
        hd_dtr: HDtr,
non_hd_dtrs: [NHdtr]))
 ===>
 cat> (NHdtr, synsem:SubjSynsem),
 cat> (HDtr, synsem:loc:cat:comps: (X, [])),
 goal> (subj_val(M, [SubjSynsem], HDtr)).
```

Since `schema1` does not realize any complements of the head daughter, the COMPS value of the head daughter is passed up directly to the mother (the variable `X` ensures token identity of COMPS values of the mother and the head daughter).

5.2.5 Example

A parse of the sentence *Janek może mu dać go* 'John may give it to him' illustrating CC with clitics' split is given below.

```
| ?- test(11).

STRING:
0 janek 1 moze 2 mu 3 dac 4 go 5

CATEGORY:

phrase_t
DTRS hd_comps_t
     HD_DTR phrase_t
            DTRS hd_comps_t
                 HD_DTR word_t
                        SYNSEM non_clitic_t
                               LOC local_t
                                   CAT category_t
                                       COMPS ne_synsem_list
                                             HD [8] non_clitic_t
                                                LOC local_t
                                                    CAT category_t
                                                        COMPS [0] ne_synsem_list
                                                              HD [2] clitic_t
                                                                 LOC local_t
                                                                     CAT category_t
                                                                         COMPS e_list
                                                                         HEAD noun_t
                                                                              CASE dat_t
                                                                              MOD none_t
                                                                         MARKING unmarked_t
                                                                         SUBJ e_list
                                                                     CONT nom_obj_t
                                                                          INDEX index_t
                                                                                GENDER masc_t
                                                                                NUMBER sing_t
                                                                                PERSON third_t
                                                              TL [9] e_list
                                                        HEAD [3] verb_t
                                                             MOD none_t
                                                             VFORM inf_t
                                                        MARKING [4] unmarked_t
```

```
                                            SUBJ [5] ne_synsem_list
                                                 HD synsem_t
                                                    LOC local_t
                                                        CAT category_t
                                                            COMPS e_list
                                                            HEAD noun_t
                                                                 CASE case_t
                                                                 MOD none_t
                                                            MARKING unmarked_t
                                                            SUBJ e_list
                                                        CONT nom_obj_t
                                                             INDEX [1] index_t
                                                                   GENDER gend_t
                                                                   NUMBER numb_t
                                                                   PERSON pers_t
                                                 TL e_list
                                        CONT [6] content_t
                                 TL [0]
                           HEAD [10] verb_t
                                MOD none_t
                                VFORM fin_t
                           MARKING [11] unmarked_t
                           SUBJ [12] ne_synsem_list
                                HD [14] synsem_t
                                   LOC local_t
                                       CAT category_t
                                           COMPS e_list
                                           HEAD noun_t
                                                CASE nom_t
                                                MOD none_t
                                           MARKING unmarked_t
                                           SUBJ e_list
                                       CONT nom_obj_t
                                            INDEX [1]
                                TL [15] e_list
                       CONT [13] content_t
NON_HD_DTRS ne_sign_list
            HD word_t
               SYNSEM [2]
            TL ne_sign_list
               HD phrase_t
                  DTRS hd_comps_t
                       HD_DTR word_t
                              SYNSEM non_clitic_t
                                     LOC local_t
                                         CAT category_t
                                             COMPS ne_synsem_list
                                                   HD [7] clitic_t
                                                      LOC local_t
                                                          CAT category_t
                                                              COMPS e_list
                                                              HEAD noun_t
                                                                   CASE acc_t
                                                                   MOD none_t
                                                              MARKING unmarked_t
                                                              SUBJ e_list
                                                          CONT nom_obj_t
```

```
                                                                                                INDEX index_t
                                                                                                      GENDER masc_t
                                                                                                      NUMBER sing_t
                                                                                                      PERSON third_t
                                                                                    TL [0]
                                                                              HEAD [3]
                                                                              MARKING [4]
                                                                              SUBJ [5]
                                                                          CONT [6]
                                                        NON_HD_DTRS ne_sign_list
                                                                    HD word_t
                                                                       SYNSEM [7]
                                                                    TL e_list
                                                   SYNSEM [8]
                                                TL e_list
                            SYNSEM non_clitic_t
                                   LOC local_t
                                       CAT category_t
                                           COMPS [9]
                                           HEAD [10]
                                           MARKING [11]
                                           SUBJ [12]
                                       CONT [13]
                     NON_HD_DTRS ne_sign_list
                                 HD word_t
                                    SYNSEM [14]
                                 TL e_list
                SYNSEM non_clitic_t
                       LOC local_t
                           CAT category_t
                               COMPS [9]
                               HEAD [10]
                               MARKING [11]
                               SUBJ [15]
                           CONT [13]


ANOTHER?  y.

no
| ?-
```

5.2.6 Conclusions

In this section, we outlined an ALE implementation of the core HPSG theory developed for Polish clitics in Chapter 2. As in the formal theory, the grammar employs a small number of syntactic schemata whereas the main stress of the account is put on the lexicon. Special care has been taken to reduce ambiguity of the obtained computational but also theoretical solutions. From this perspective, the implementation was not only a test of the theory but also improved its precision. Despite of the stiff phrase structure backbone of ALE, a variety of postverbal positions of clitic and non-clitic complements have been successfully implemented. Due to the differences between the system and the HPSG formalism, some aspects of the HPSG theory could not be encoded without modifications.

To conclude, the implemented grammar quite closely follows the theoretical approach, which both confirms the coherence of our theory and the usefulness of ALE as a tool for implementing HPSG-like grammars.

Chapter 6

Conclusions

In this dissertation we have presented a theoretical HPSG account of certain linguistic phenomena associated with Polish clitics (unstressed elements). The core of the formal analysis has been encoded in ALE, a computational platform for implementing HPSG grammars.

We are mostly concerned with Polish pronominal clitics, the reflexive clitic *się* 'self' (the reflexive marker, RM), and the sentential negative marker (henceforth, the NM) *nie* 'not'. Auxiliary clitics are not the main topic of our investigation but their behaviour is discussed in the context of other clitics.

As a thorough examination of properties of Polish clitics shows they do not form a uniform grammatical class. As we argue, Polish pronominal clitics and the RM are syntactic items whereas the NM *nie* 'not' is a verbal prefix. After Borsley (1999), we treat auxiliary clitics as syntactic items.

Since Polish pronominal clitics and the RM are syntactic clitics, they share a number of properties. First, they are deficient nominal forms (in the sense of Cardinaletti and Starke (1994)): they cannot be coordinated or modified and their distribution is more restricted than other NPs. Then, linear positions of pronominal clitics and the RM are the same: if postverbal, they can be separated from the preceding verb only by other clitics. Furthermore, clitic climbing (CC) is equally possible for these clitics: in certain verbal environments, a clitic is either realised as a syntactic argument of a local verb or as an argument of a syntactically higher verb. This phenomenon is always optional in Polish and if there are several clitics, they can be realised as arguments of different verbs in the complex.

Pronominal clitics and the RM also differ in several respects. Pronominal clitics are in principle interchangeable with full NPs which are semantic and syntactic arguments of a verb. The RM, however, is mutlifunctional and its substitution with another NP is not always possible. The RM is often the unique realisation of a syntactic requirement of the verb and, as we argue, the RM is not assigned a semantic role. The latter observation indicates also that the presence of the RM, unlike pronominal clitics, is tightly related to lexical semantics of verbs. Furthermore, in contrast to pronominal clitics, the RM exhibits the so-called haplology: in some contexts, if there are several predicates which independently require the RM, the realisation of a single RM is sufficient if other RMs are omitted. Haplology of the Polish RM is particularly interesting since the RM is a syntactic item and the usual explanation of haplology as a morphological or phonological phenomenon cannot be adopted for the RM *się*.

We address the behaviour of clitics in all syntactic environments they occur: simple sentences, verbal complexes, tense auxiliary constructions, coordination but also gerund and participle phrases. The syntactic HPSG analysis of pronominal clitics and the RM is essentially the same. Clitics are represented as syntactic items subcategorized for by the verb on a par with other NPs. The distinction between clitics and other syntactic items is encoded in the *synsem* component of a sign. This allows clitics to retain usual characteristics of NPs: inflection for case or person, gender, number specification. The realisation of Polish pronominal clitics and the RM is syntactic and they obey general syntactic principles of the language. In order to capture Polish data, we slightly modify HPSG principles proposed in Pollard and Sag (1994). In particular, we license phrases with unrealised complements and provide a version of the Valence Principle which can realise complements in any order and quantity. These modifications are necessary to account not only for the behaviour of Polish clitics but also for other phenomena in the language, see Przepiórkowski (2000a). Since clitic climbing is possible only in certain verbal environments, we encode this phenomenon lexically adopting a variant of argument composition à la Hinrichs and Nakazawa (1989). Verbs which (dis)allow CC are lexically specified whether they can(not) inherit unrealised complements of their verbal complements. As the subcategorization frame of verbs which accept CC is not further constrained, argument raising is optional which accounts for the optionality of CC.

The RM, unlike pronominal clitics, is present among syntactic requirements of a verb directly in the lexicon which is triggered either by lexical semantics of the verb or a lexical rule for impersonal forms. Haplology is encoded as a syntactic non-realisation of the RM. In case of VCs, it is encoded lexically as a restriction on syntactic raising of the RM to reflexivized verbs: in contrast to other arguments, an RM cannot be raised to a reflexivized verb. This allows us to account for distinct behaviour of haplology and CC in case of negation: haplology is possible irrespectively of negation, unlike CC.

We provide an account of linear positions of syntactic clitics using order domains of Reape (1992) and Kathol (1995). Due to the flexibility of Polish word order, we adopt a mechanism which orders not only constituents but also allows them to be interleaved. We provide an analysis which uniformly captures the distribution of nominal clitics in simple clauses, verbal complexes, participle phrases and tense auxiliary constructions. An additional restriction is necessary to fully capture the order of elements (and not only clitics) in gerund phrases. Two separate linear constraints regulate the relative order of auxiliary clitics with respect to the participial complement and other syntactic clitics.

The NM is a verbal prefix and its behaviour sharply contrast with that of the clitics discussed above. The affix status of the NM is strongly supported by idiosyncratic behaviour of certain verbs with respect to negation, the strict adjacency of the NM to a verb as well as prosody and coordination facts. We provide a morphosyntactic analysis of the NM prefixation. We propose a constraint-based approach which overcomes problems with a more 'standard' account based on a lexical rule. We use the syntax-morphology interface as a locus of the NM spell-out which provides a uniform account of the NM prefixation both in regularly and irregularly negated verbs.

Finally, the core of HPSG theory developed for syntactic clitics has been implemented in ALE. As in the HPSG theory, the ALE grammar employs a small number of syntactic schemata whereas the main stress of the account is put on the lexicon. Despite of the stiff phrase structure backbone of ALE, the variety of postverbal word order, including restrictions on postverbal clitic positions, have been successfully implemented. The differences between the computational platform and the HPSG formalism disallow encoding of the HPSG theory without modifications. Nevertheless, the implemented grammar closely follows the theoretical approach, which confirms the coherence of our theory.

Appendix A

Clitic Climbing in ALE

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CC in Polish VCs
% author: AK
% version 25.06.2000
%
% arguments can be realized in any order and in any number
% due to the phrase structure rule component of ALE
% the lexical head must always precede its complements, e.g.,
% the verb precedes its complements: OK lubi marysie but * marysie lubi
% postverbal clitics precede non-clitic arguments
% disambiguated wrt to the order of clitics
%
% the grammar tested on the following examples:
%
% janek moze lubic go         % 1 solution
% janek go lubi               % 0 solutions: the head must precede complements
% janek dal mu kwiaty         % 1 solution
% kwiaty                      % 2 solutions: kwiaty is either NP[nom] or NP[acc]
% janek moze go lubic         % 1 sloution
% janek kazal mu lubic go     % 1 solution
% janek kazal mu go lubic     % 1 solution
% janek kazal marysi go lubic % 0 solutions: raised non-clitic
%                                     % 'marysi' precedes raised clitic 'go'
% janek kazal go marysi lubic         % 1: go is raised and it precedes marysi
%                                     % NP[dat] complement of 'kazal'
% janek powiedzial ze marysia lubi go  % 1
% janek obiecal mu lubic go           % 1: no CC; 'mu' is a
%                                     % NP[dat] argument of 'obiecal'
% janek obiecal mu go lubic           % 0: (OK, no raising to a subject
%                                     % control verb with NP[dat] complement
% janek moze mu dac go         % 1: clitic's split
% janek idzie                  %  1: N + V
% janek boi sie go             %  1
```

```
% janek moze sie go bac        %  1: the RM and ‘go’ are raised
% janek kazal mu go dac marysi %  1: ‘go’ is raised, ‘mu’ is NP[dat] complement of
%                              %  ‘kazal’, NP ‘marysi’ is not raised
% janek kazal mu sie bac go    %  1: the RM is raised
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%


%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Signature
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%

bot sub [sign_t, mod_synsem_t, list, local_t, hd_str_t, case_t, category_t,
         content_t, head_t, marking_t, marker_t, vform_t, pform_t, bool_t,
         index_t, pers_t, numb_t, gend_t].

sign_t sub [phrase_t, word_t]
  intro [synsem: synsem_t].

word_t sub [].
phrase_t sub []
    intro [dtrs: hd_str_t].

sign_t cons S goal (single_subj(S)).

phrase_t cons P goal (non_clitic_phrases(P),
                      hfp(P),
                      marking_principle(P),
                      sem_principle(P)).

hd_str_t sub [hd_mk_t, hd_adj_t, hd_comps_t]
  intro [hd_dtr: sign_t,
         non_hd_dtrs: sign_list].

hd_mk_t sub []
  intro [mk_dtr: word_t,
         non_hd_dtrs: e_list].

hd_adj_t sub []
  intro [adj_dtr: sign_t,
         non_hd_dtrs: e_list].

hd_comps_t sub [].
```

```
list sub [e_list, ne_list, sign_list, synsem_list, non_clitic_list,
          clitic_list].
  e_list sub [].

sign_list sub [e_list, ne_sign_list].
  ne_sign_list sub []
     intro [hd: sign_t,
            tl: sign_list].

ne_list sub [ne_sign_list, ne_synsem_list]
  intro [hd: bot,
         tl: list].

synsem_list sub [e_list, ne_synsem_list, non_clitic_list, clitic_list].
  ne_synsem_list sub [ne_clitic_list, ne_non_clitic_list]
    intro [hd: synsem_t,
           tl: synsem_list].

non_clitic_list sub [e_list, ne_non_clitic_list].
  ne_non_clitic_list sub []
    intro [hd: non_clitic_t,
           tl: non_clitic_list].

clitic_list sub [e_list, ne_clitic_list].
  ne_clitic_list sub []
    intro [hd: clitic_t,
           tl: clitic_list].

synsem_t sub [clitic_t, non_clitic_t] % non_canonical is not used
  intro [loc: local_t].

  clitic_t sub [].
  non_clitic_t sub [].

local_t sub []
  intro [cat: category_t,
         cont: content_t].

content_t sub [nom_obj_t, psoa_t].
  nom_obj_t sub [npro_t, pron_t]
      intro [index: index_t].

    npro_t sub [].
    pron_t sub [ppro_t, ana_t].
      ppro_t sub [].
      ana_t sub [refl_t, recp_t].
        refl_t sub [].
```

```
        recp_t sub [].
  psoa_t sub [].

index_t sub [ref_t, non_ref_t]
  intro [person: pers_t,
         number: numb_t,
         gender: gend_t].

pers_t sub [first_t, second_t, third_t].

numb_t sub [sing_t, plur_t].

gend_t sub [fem_t, masc_t, neut_t, pl_gend_t].

category_t sub []
  intro [head: head_t,
         subj: synsem_list,
         comps: synsem_list,
         marking: marking_t].

marking_t sub [marked_t, unmarked_t].

marked_t sub [ze_t, zeby_t].
  ze_t sub [].
  zeby_t sub [].

unmarked_t sub [].

head_t sub [subst_t, marker_t].

subst_t sub [noun_t, adj_t, verb_t, prep_t, adv_t]
  intro [mod: mod_synsem_t].

mod_synsem_t sub [synsem_t, none_t].
  none_t sub [].

noun_t sub []
  intro [case: case_t,
         mod: none_t].

case_t sub [nom_t,gen_t,dat_t,acc_t,loc_t,instr_t].
  nom_t sub [].
  gen_t sub [].
  dat_t sub [].
  acc_t sub [].
  loc_t sub [].
  instr_t sub [].
```

```
adj_t sub []
  intro  [mod: synsem_t,
          pred: bool_t].

  bool_t sub [yes_t, no_t].
    yes_t sub [].
    no_t sub [].

adv_t sub []
  intro [mod: synsem_t].

verb_t sub []
  intro [vform: vform_t,
         mod: none_t].

vform_t sub [inf_t, fin_t].
  inf_t sub [].
  fin_t sub [].

prep_t sub []
  intro [pform: pform_t].

pform_t sub [na_t, na_loc_t, w_t].
  na_loc_t sub [].
  na_t sub [].
  w_t sub [].

marker_t sub []
  intro [spec: synsem_t].


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Phrase Structure Rules
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


schema1 rule
(M,
 synsem:loc:cat:comps: (X, []),
 dtrs: (hd_comps_t,
        hd_dtr: HDtr,
non_hd_dtrs: [NHdtr]))
 ===>
 cat> (NHdtr, synsem:SubjSynsem),              % phrase_t in HPSG94
```

```
 cat> (HDtr, synsem:loc:cat:comps: (X,[])),   % phrase_t in HPSG94
 goal> (subj_val(M, [SubjSynsem], HDtr)).


schema2 rule
(M, phrase_t,
    dtrs: (hd_comps_t,
           hd_dtr: HDtr,
           non_hd_dtrs: Signs))
 ===>
 cat> (HDtr, word_t,
  synsem:loc:cat:comps:Comps),
 goal> (at_most_three(Comps),                     % restricts the length of the Comps list
        shuffle((SubComps, ne_list),Rest,Comps), % takes any ordered subset of Comps
        split(SubComps),                    % splits SubComps into clitics and non-clitics;
                                            % clitics must precede non-clitics
        synsems_to_signs(SubComps,Signs)),        % maps synsems (SubComps) to signs
 cats> Signs,
 goal> (comps(M, Rest),                           % gives the value of mother's COMPS
        subj_val(M, [], HDtr)).                   % gives the value of mother's SUBJ


schema4 rule
(M, phrase_t,
 dtrs: (hd_mk_t,
        mk_dtr: MDtr,
        hd_dtr: HDtr))
 ===>
 cat> (MDtr, word_t, synsem:loc:cat:(head:(marker_t,
                                            spec: Synsem))),
 cat> (HDtr, phrase_t, synsem:Synsem),
 goal> (comps_val(M, HDtr,[]),
        subj_val(M, [], HDtr)).


schema5 rule
(M,phrase_t,
 dtrs: (hd_adj_t,
        hd_dtr: HDtr,
        adj_dtr: AdjDtr))
 ===>
 cat> (AdjDtr, sign_t,                      % phrase_t in HPSG94
       synsem:loc:cat:head:mod: (Synsem, non_clitic_t)),
 cat> (HDtr, sign_t,                        % phrase_t in HPSG94
       synsem: Synsem),
 goal> (comps_val(M, HDtr,[]),
        subj_val(M, [], HDtr)).
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% definite clauses
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%


synsems_to_signs([],[]) if  true.
synsems_to_signs([Syns|SynsList],[(sign_t, synsem:Syns)|SignList]) if
  synsems_to_signs(SynsList,SignList).

append([], X, X) if true.
append([H|R],L,[H|L1]) if append(R,L,L1).

at_most_three((tl: ([];
                   (tl: ([];
                    tl:[]))) )) if true.

% shuffle/3 gets an arbitrary subset of Ys with R as the remaining list

shuffle([X|Xs], R, Ys) if
  select(X, Ys, Ys1), shuffle(Xs, R, Ys1).
shuffle([], Ys, Ys) if true.

% split/1 splits L into clitic and non-clitics lists;
% clitics precede non-clitics
split(L) if
  append(clitic_list, non_clitic_list, L).

% removes one element from a list; the usual Prolog select/3 predicate
select(Z, [Z|Xs], Xs) if true.
select(Z, [Y|Ys], [Y|Zs]) if
  select(Z, Ys, Zs).

% clitics are not projected to phrases (HD-DTR|SYNSEM is always a non-clitic) and
% phrases have non-clitic synsem
%
non_clitic_phrases((phrase_t, synsem: non_clitic_t,
                    dtrs:hd_dtr:synsem: non_clitic_t)) if true.

% Head Feature Principle

hfp((synsem:loc:cat:head:H, dtrs:hd_dtr:synsem:loc:cat:head:H)) if true.

comps((synsem:loc:cat:comps: M), M) if true.
```

```
comps_val((synsem:loc:cat:comps: M), (synsem:loc:cat:comps: H), X) if
  append(X, M, H).

subj_val((synsem:loc:cat:subj: M), S, (synsem:loc:cat:subj: H)) if
  append(S, M, H).

single_subj((synsem:loc:cat:subj: ([]; tl:[]))) if true.

marking_principle((dtrs:(hd_mk_t,
                         mk_dtr:synsem:loc:cat:marking:X),
                   synsem:loc:cat:marking:X ;
                   dtrs:((hd_comps_t;
                          hd_adj_t),
                          hd_dtr:synsem:loc:cat:marking:X),
                   synsem:loc:cat:marking:X
                  )) if
   true.

sem_principle((dtrs:(hd_adj_t,
                     adj_dtr:synsem:loc:cont:X),
               synsem:loc:cont:X ;
               dtrs:((hd_comps_t;
                      hd_mk_t),
                      hd_dtr:synsem:loc:cont:X),
               synsem:loc:cont:X
                  )) if
   true.


%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% macros
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%


noun(Case) macro
  loc:cat:(head: case: Case,
   subj: [],
           comps: [],
           marking: unmarked_t).

noun(Synsem, Case) macro
 (Synsem,
  loc:cat:(head: case: Case,
   subj: [],
```

```
          comps: [],
          marking: unmarked_t)).

nps(Case, Index) macro
  (synsem_t,
  loc:(cat:(head: case: Case,
   subj: [],
          comps: [],
          marking: unmarked_t),
      cont:index: Index)).

npnc(Case, Index) macro
  (non_clitic_t,
  loc:(cat:(head: case: Case,
   subj: [],
          comps: [],
          marking: unmarked_t),
      cont:index: Index)).

non_verb macro
   (noun_t; adj_t; prep_t; adv_t).

non_verb_synsem macro
  loc:cat:head: @non_verb.

thsg(Gend) macro
  person: third_t,
  number: sing_t,
  gender: Gend.

adjp macro
  loc:cat:(head:(adj_t,
                pred: yes_t),
          subj:[],
          comps:[]).

s(Vform,Marking) macro
  non_clitic_t,
  loc:cat:(head:vform:Vform,
          subj: [],
          comps: [],
          marking: Marking).

prep(Pform,Mod) macro
  loc:cat:(head:(pform:Pform,
                mod:Mod),
          subj:[],
```

```
          comps:[],
          marking:unmarked_t).

vinf macro
  non_clitic_t,         % no clitic verbs considered
  loc:cat:(head:vform:inf_t,
          subj: [@noun(_)],
          comps: [],
          marking:unmarked_t).

inf_comp(Index, List) macro
  non_clitic_t,
  loc:cat:(head:vform:inf_t,
          subj: [@nps(_, Index)],
          comps: List,
          marking:unmarked_t).

inf_no_rm_comp(Index, List) macro
  non_clitic_t,
  loc:cat:(head:vform:inf_t,
          subj: [@nps(_, Index)],
          comps: (List, ([]; [@non_rm];
          [@non_rm, @non_rm]; [@non_rm, @non_rm, @non_rm])),
          marking:unmarked_t).

no_rm_list macro
   []; [@non_rm]; [@non_rm, @non_rm]; [@non_rm, @non_rm, @non_rm].

rm macro
  clitic_t,
  loc:(cat:(head: case: acc_t,
   subj: [],
          comps: [],
          marking: unmarked_t),
       cont: ana_t).

non_rm(Case) macro
   (@npnc(Case, _);
   (@noun(clitic_t, Case),
    loc:cont: ppro_t)).

non_rm macro
  (non_clitic_t;
   (@noun(clitic_t, _),
    loc:cont: ppro_t)).
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% a simple lexicon
%
%%%%%%%%%%%%%%%%%%%%%%%%%%

janek ---> word_t, synsem: @noun(nom_t).

marysi ---> word_t, synsem: @npnc((dat_t; gen_t), @thsg(fem_t)).

mu ---> word_t, synsem: (clitic_t, @nps(dat_t, @thsg(masc_t))).

go ---> word_t, synsem: (clitic_t, @nps((acc_t; gen_t), @thsg(masc_t))).

sie ---> word_t, synsem: @rm.

marysia --->  word_t, synsem: @noun(non_clitic_t, nom_t).

marysie --->  word_t, synsem: @noun(non_clitic_t, acc_t).

stole ---> word_t, synsem: @noun(non_clitic_t, loc_t).

kwiaty ---> word_t, synsem: (non_clitic_t, @noun((acc_t;nom_t))).


idzie ---> word_t,
           synsem:loc:(cat:(head:vform:fin_t,
   subj: [@noun(nom_t)],
                           comps: [],
                           marking: unmarked_t)).

isc ---> word_t,
           synsem:loc:(cat:(head:vform:inf_t,
   subj: [@noun(_)],
                           comps: [],
                           marking: unmarked_t)).

dal ---> word_t,
         synsem:loc:(cat:(head:vform:fin_t,
                 subj: [@noun(synsem_t, nom_t)],
         comps: ([@noun(synsem_t, acc_t), @noun(synsem_t, dat_t)]),
                         marking: unmarked_t)).

dac ---> word_t,
         synsem:loc:(cat:(head:vform:inf_t,
                 subj: [@noun(synsem_t, _)],
         comps: ([@noun(synsem_t, acc_t), @noun(synsem_t, dat_t)]),
```

```
                          marking: unmarked_t)).

czeka ---> word_t,
           synsem: (non_clitic_t,
                  loc:cat:(head:vform:fin_t,
                           subj:[@noun(nom_t)],
                           comps:[@prep(na_t,none_t)],
                           marking: unmarked_t)).

czekac ---> word_t,
            synsem:(non_clitic_t,
                    loc:cat:(head:vform:inf_t,
                            subj:[@noun(_)],
                            comps:[],
                            marking:unmarked_t)).

powiedzial ---> word_t,
                synsem:loc:cat:(head:vform:fin_t,
                                subj:[@noun(nom_t)],
                                comps:[@s(fin_t,ze_t)],
                                marking:unmarked_t).

lubi ---> word_t,
          synsem:loc:cat:(head:vform:fin_t,
                          subj:[@noun(non_clitic_t, nom_t)],
                          comps:[@noun(synsem_t, acc_t)],
                          marking:unmarked_t).

lubic ---> word_t,
           synsem:loc:cat:(head:vform:inf_t,
                          subj:[@noun(_)],
                          comps:[@noun(synsem_t, acc_t)],
                          marking:unmarked_t).

boi ---> word_t,
          synsem:loc:cat:(head:vform:fin_t,
                          subj: [@noun(non_clitic_t, nom_t)],
                          comps: [@rm, @non_rm(gen_t)],
                          marking: unmarked_t).

bac ---> word_t,
          synsem:loc:cat:(head:vform:inf_t,
                          subj: [@noun(non_clitic_t, _)],
                          comps: [@rm, @non_rm(gen_t)],
                          marking: unmarked_t).

stoja --->  word_t,
```

```
             synsem:loc:cat:(head:vform:fin_t,
            subj: [@noun(nom_t)],
                             comps: ([];[@prep(na_loc_t,none_t)]),
                             marking: unmarked_t).

kazal ---> word_t,
            synsem:loc:cat:(head:vform:fin_t,
                             subj:[@noun(nom_t)],
                             comps:([@nps(dat_t, X),@inf_comp(X, List) | List]),
                             marking:unmarked_t).

kazac ---> word_t,
            synsem:loc:cat:(head:vform:inf_t,
                             subj:[@noun(_)],
                             comps:[@nps(dat_t, X),@inf_comp(X, List) | List],
                             marking:unmarked_t).

przeszkodzil ---> word_t,  % no CC
            synsem:loc:cat:(head:vform:fin_t,
                             subj:[@noun(nom_t)],
                             comps:([@nps(dat_t, X),
                               @inf_comp(X, (List, non_clitic_list)) | List]),
                             marking:unmarked_t).

przeszkodzic ---> word_t,  % no CC
            synsem:loc:cat:(head:vform:inf_t,
                             subj:[@noun(_)],
                             comps:([@nps(dat_t, X),@inf_comp(X, [])]),
                             marking:unmarked_t).

obiecal ---> word_t,       % no CC
            synsem:loc:cat:(head:vform:fin_t,
                             subj:[@nps(nom_t, X)],
                             comps:[@noun(dat_t),@inf_comp(X, [])],
                             marking:unmarked_t).

obiecac ---> word_t,       % no CC
            synsem:loc:cat:(head:vform:inf_t,
                             subj:[@nps(_, X)],
                             comps:[@noun(dat_t),@inf_comp(X, [])],
                             marking:unmarked_t).

moze --->  word_t,
            synsem:loc:cat:(head:vform:fin_t,
                             subj:[@nps(nom_t, X)],
                             comps:[@inf_comp(X, List) | List],
                             marking:unmarked_t).
```

```
wyglada ---> word_t,
             synsem:loc:cat:(head:vform:fin_t,
                             subj:[@noun(nom_t)],
                             comps:[@prep(na_t,none_t)],
                             marking:unmarked_t).

na ---> word_t,
        synsem:loc:cat:(head:(pform:na_loc_t,
                             mod:none_t),
                        subj: [],
                        comps: [@noun(loc_t)],
                        marking: unmarked_t).

na ---> word_t,
        synsem:loc:cat:(head:(pform:na_t,
                              mod:none_t),
                        subj: [],
                        comps: ([@noun(acc_t)];[@adjp])).

duzym ---> word_t,
           synsem:loc:cat:(head:(pred:no_t,
                                 mod: @noun(loc_t)),
                           subj:[],
                           comps:[],
                           marking:unmarked_t).

zmeczonego ---> word_t, synsem: @adjp.

piekne ---> word_t,
           synsem:loc:(cat:(head: (adj_t,mod: @noun((nom_t;acc_t))),
                           subj: [],
                           comps: [],
                           marking: unmarked_t)).

bardzo ---> word_t,
            synsem:loc:(cat:(head:(adv_t,
                                   mod:loc:cat:head:adj_t),
                            subj: [],
                            comps: [],
                            marking: unmarked_t)).

ze ---> word_t,
        synsem:loc:cat:(head: (marker_t, spec: @s(fin_t,unmarked_t)),
                        marking: ze_t,
                        subj: [],
                        comps: []).
```

Bibliography

Abeillé, A. and Godard, D. (1994). The complementation of tense auxiliaries in French. In R. Aranovich, W. Byrne, S. Preuss, and M. Senturia, editors, *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, Stanford. CSLI Publications/SLA.

Abeillé, A. and Godard, D. (1997). The syntax of French negative adverbs. In D. Forget, P. Hirschbühler, F. Martineau, and M.-L. Rivero, editors, *Negation and Polarity*, volume 155 of *Current Issues in Linguistic Theory*, pages 1–27. Benjamins, Amsterdam.

Abeillé, A., Godard, D., and Sag, I. A. (1998). Two kinds of composition in French complex predicates. In Hinrichs *et al.* (1998), pages 1–41.

Ajdukiewicz, K. (1935). Die syntaktische konnexität. *Studia Philosophica*, **1**, 1–27. Translates as 'Syntactic Connexion' in S. McCall, ed., 1967 *Polish logic* Oxford: Clavendom Press, pp. 207–231.

Alsina, A. (1996). *The Role of Argument Structure in Grammar.* Number 62 in CSLI Lecture Notes. CSLI Publications, Stanford.

Anderson, S. R. (1992). *A-Morphous Morphology.* Number 62 in Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.

Backofen, R., Becker, T., Calder, J., Capstick, J., Dini, L., Dörre, J., Erbach, G., Estival, D., Manandhar, S., Mineur, A.-M., van Noord, G., Oepen, S., and Uszkoreit, H. (1996). The EAGLES formalisms working group final report. Technical report, EAGLES.

Bąk, P. (1984). *Gramatyka języka polskiego.* Wiedza Powszechna, Warszawa, fourth edition.

Bąk, P. (1995). *Gramatyka języka polskiego.* Wiedza Powszechna, Warszawa, ninth edition.

Balari, S. and Dini, L., editors (1998). *Romance in HPSG.* Number 75 in CSLI Lecture Notes. CSLI Publications, Stanford.

Bartnicka, B. and Satkiewicz, H. (1990). *Gramatyka języka polskiego dla cudzoziemców.* Wiedza Powszechna, Warszawa.

Bird, S. and Klein, E. (1994). Phonological analysis in typed feature systems. *Computational Linguistics*, **20**, 455–491.

Błaszczak, J. (2000). Towards a binding analysis of negative polarity items in Polish. In *Proceedings of the Second European Conference on Formal Description of Slavic Languages, Potsdam, Germany, November 20–22, 1997.* To appear.

Bogusławski, A. (1977). Polskie *się* — słowo nie do końca poznane. *International Review of Slavic Linguistics*, **2**, 99–124.

Bolc, L., Czuba, K., Kupść, A., Marciniak, M., Mykowiecka, A., and Przepiórkowski, A. (1996). A survey of systems for implementing HPSG grammars. Technical Report 814, Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.

Bolc, L., Czuba, K., Kupść, A., Marciniak, M., Mykowiecka, A., and Przepiórkowski, A. (2000). Formalna gramatyka języka polskiego: Teoria i implementacja. In progress, preliminary title.

Bonet, E. (1995). Feature structure of Romance clitics. *Natural Language and Linguistic Theory*, **13**, 607–647.

Booij, G. E. and Rubach, J. (1987). Postcyclic versus postlexical rules in lexical phonology. *Linguistic Inquiry*, **18**, 1–44.

Borsley, R. D. (1999). Weak auxiliaries, complex verbs and inflected complementizers in Polish. In Borsley and Przepiórkowski (1999), pages 29–59.

Borsley, R. D. and Przepiórkowski, A., editors (1999). *Slavic in Head-Driven Phrase Structure Grammar*. CSLI Publications, Stanford.

Borsley, R. D. and Rivero, M. L. (1994). Clitic auxiliaries and incorporation in Polish. *Natural Language and Linguistic Theory*, **12**, 373–422.

Bouma, G., Malouf, R., and Sag, I. (1998). A unified theory of complement, adjunct, and subject extraction. 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 83–97.

Bouma, G., Malouf, R., and Sag, I. (2000). Satisfying constraints on extraction and adjunction. To appear in *Natural Language and Linguistic Theory*.

Bresnan, J., editor (1982). *The Mental Representation of Grammatical Relations*. MIT Press Series on Cognitive Theory and Mental Representation. The MIT Press, Cambridge, Mass.

Bresnan, J. and Mchombo, S. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory*, **13**, 181–254.

Brown, S. (1996). *The Syntax of Negation in Russian*. Ph.D. thesis, Indiana University.

Brown, S. (1999). *The Syntax of Negation in Russian: A Minimalist Approach*. CSLI Publications, Stanford.

Brown, S. and Franks, S. (1995). Asymmetries in the scope of Russian negation. *Journal of Slavic Linguistics*, **3**(2), 239–87.

Brückner, A. (1957). *Słownik etymologiczny języka polskiego*. Wiedza Powszechna, Warszawa.

Bugajski, M. (1983). *Morfem* nie *we współczesnym języku polskim i zasadach pisowni*, volume 647 of *Acta Universitatis Wratislaviensis*. Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław.

Burzio, L. (1986). *Italian syntax*. Reidel, Dordrecht.

Buttler, D., Kurkowska, H., and Satkiewicz, H. (1971). *Kultura języka polskiego: zagadnienia poprawności gramatycznej*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa.

Calder, J. and Humphreys, K. (1995). Pleuk overview. Unpublished manuscript.

Cardinaletti, A. and Starke, M. (1994). The typology of structural deficiency on the three grammatical classes. *University of Venice Working Papers in Linguistics*, **4**(2), 41–109.

Carpenter, B. (1992). *The Logic of Typed Feature Structures*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, New York.

Carpenter, B. and Penn, G. (1995). *The Attribute Logic Engine (Version 2.0.1). User's Guide*. Carnegie Mellon University, Pittsburgh.

Carpenter, B. and Penn, G. (1999). *The Attribute Logic Engine (Version 3.2 Beta). User's Guide*. Carnegie Mellon University, Pittsburgh.

Chomsky, N. (1957). *Syntactic Structures*. Mouton, The Hague.

Chomsky, N. (1981). *Lectures on Government and Binding*. Foris, Dordrecht.

Colmerauer, A. (1978). Metamorphosis grammars. In L. Bolc, editor, *Natural Language Communications with Computers*, number 63 in Lecture Notes in Computer Science. Springer-Verlag, Berlin.

Crystal, D. (1997). *A Dictionary of Linguistics and Phonetics*. Blackwell, Oxford, fourth edition.

Czuba, K. (1995a). Multiple inheritance applications to construction of a lexeme lexicon for Polish. Technical Report 780, Institute of Computer Science, Polish Academy of Sciences.

Czuba, K. (1995b). *Zastosowanie dziedziczenia do analizy wybranych aspektów języka polskiego*. Master's thesis, Uniwersytet Warszawski, Warszawa.

Czuba, K. (1997). Towards minimal linguistic description: A hierarchical gender structure for Polish. In Junghanns and Zybatow (1997), pages 347–356.

Czuba, K. and Przepiórkowski, A. (1995). Agreement and case assignment in Polish: An attempt at a unified account. Technical Report 783, Institute of Computer Science, Polish Academy of Sciences.

Davis, A. (1997). *Lexical Semantics and Linking and the Hierarchical Lexicon*. Ph.D. thesis, Stanford University.

Derwojedowa, M. (1998). *Porządek linearny składników zdania elementarnego w języku polskim*. Ph.D. thesis, Uniwersytet Warszawski, Warszawa.

Dłuska, M. (1974). *Prozodia języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa.

Dobrovie-Sorin, C. (1998). Impersonal *se* Constructions in Romance and the Passivization of Unergatives. *Linguistic Inquiry*, **29**, 399–473.

Dornisch, E. (1997). Auxiliaries and functional projections in Polish. In W. Browne, E. Dornisch, N. Kondrshova, and D. Zec, editors, *Annual Workshop on Formal Approaches to Slavic Linguistics, The Cornell Meeting 1995*, volume 39 of *Michigan Slavic Materials*, pages 183–209, Ann Arbor. Michigan Slavic Publications.

Dukiewicz, L. and Sawicka, I. (1995). *Gramatyka współczesnego języka polskiego. Fonetyka i fonologia.* Instytut Języka Polskiego PAN, Kraków.

Dyła, S. (1983). Evidence for S-deletion in Polish. *Folia Linguistica*, **XVII**, 327–337.

Dziwirek, K. (1994). *Polish Subjects*. Garland, New York & London.

Dziwirek, K. (1998). Reduced construction in universal grammar: Evidence from the Polish object control construction. *Natural Language and Linguistic Theory*, **16**, 53–99.

Emele, M. C. (1993). TFS: The typed feature structure representation formalism. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Emele, M. C. (1994). TFS: The typed feature structure representation formalism. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Evans, R. and Gazdar, G. (1989). The DATR Papers: May 1989. Technical report, University of Sussex, Brighton.

Evans, R. and Gazdar, G. (1990). The DATR Papers: February 1990. Technical Report CSRP 139, University of Sussex, Brighton.

Fodor, J. D. (1992). Islands, learnability and the lexicon. In H. Goodluck and M. Rochemont, editors, *Island Constraints*, pages 109–180. Kluwer Academic Publishers.

Fouvry, F. and Meurers, D. (2000). Towards a platform for linearization grammars. In *Proceedings of the ESSLLI'2000 Workshop on Linguistic Theory and Grammar Implementation, August 6–18, 2000, Birmingham, UK*, Birmingham.

Fowler, G. (1993). A syntactic account of derivational *-sja* in Russian. In R. A. Maguire and A. Timberlake, editors, *American Contributions to the Eleventh International Congress of Slavists*, pages 270–284. Slavica Publishers, Columbus, Ohio.

Franks, S. (1998). Clitics in Slavic. Position paper at the Workshop "Comparative Slavic Morphosyntax", Spencer, Indiana, 5–7 June, 1998.

Franks, S. and Bański, P. (1999). Approaches to "schizophrenic" Polish person agreement. 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 123–143, Ann Arbor. Michigan Slavic Publications.

Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Blackwell and Harvard University Press, Cambridge and Cambridge, Mass.

Geach, P. T. (1970). A program for syntax. *Synthese*, **20**, 483–497.

Gołąb, Z., Heinz, A., and Polański, K. (1968). *Słownik terminologii językoznawczej.* Pa'nstwowe Wydawnictwo Naukowe.

Götz, T., Meurers, D., and Gardemann, D. (1997). *The ConTroll Manual*. Universität Tübingen.

Greszczuk, B. (1993). *Składniowe wykładniki negacji i ich funkcje w historii języka polskiego*. Wydawnictwo Wyższej Szkoły Pedagogicznej w Rzeszowie, Rzeszów.

Grimshaw, J. (1982). On the Lexical Representation of Romance Reflexive Clitics. In Bresnan (1982), chapter 2, pages 87–148.

Grimshaw, J. (1990). *Argument Structure*. Linguistic Inquiry Monographs. The MIT Press, Cambridge, Mass.

Grzegorczykowa, R., Laskowski, R., and Wróbel, H., editors (1999). *Gramatyka współczesnego języka polskiego: Morfologia*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa, third edition.

Halpern, A. (1995). *On the Placement and Morphology of Clitics*. CSLI Publications, Stanford.

Hinrichs, E. and Nakazawa, T. (1989). Flipped out: Aux in German. In *Papers from the 25th Regional Meeting of the Chicago Linguistic Society*, pages 193–202, Chicago. Chicago Linguistic Society.

Hinrichs, E., Kathol, A., and Nakazawa, T., editors (1998). *Complex Predicates in Nonderivational Syntax*, volume 30 of *Syntax and Semantics*. Academic Press, New York.

Höhle, T. N. (1999). An architecture for phonology. In Borsley and Przepiórkowski (1999), pages 61–90.

Holvoet, A. (1991). *Transitivity and Clause Structure in Polish. A Study in Case Marking*, volume 95 of *Prace Slawistyczne*. Slawistyczny O'srodek Wydawniczy PAN, Warszawa.

Jackendoff, R. (1992). Madame Tussaud meets the Binding Theory. *Natural Language and Linguistic Theory*, **10**, 1–31.

Jaworski, M. (1974). *Podręczna gramatyka języka polskiego*. Wydawnictwa Szkolne i Pedagogiczne, Warszawa, third edition.

Jaworski, M. (1986). *Podręczna gramatyka języka polskiego*. Wydawnictwa Szkolne i Pedagogiczne, Warszawa, fourth edition.

Jodłowski, S. (1976). *Podstawy polskiej składni*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa.

Jodłowski, S. and Taszycki, W. (1968). *Słownik ortograficzny*. Ossolineum, Wrocław.

Johnson, M. (1986). A GPSG account of VP structure in German. *Linguistics*, **24**, 871–882.

Junghanns, U. and Zybatow, G., editors (1997). *Formale Slavistik*, volume 7 of *Leipziger Schriften zur Kultur-, Literatur-, Sprach- und Uebersetzungswissenschaft*, Frankfurt am Main. Vervuert Verlag.

Kallas, K. (1993). *Składnia współczesnych polskich konstrukcji współrzędnych*. Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

Kański, Z. (1986). *Arbitrary Reference and Reflexivity: A Generative Study of the Polish Pronoun* się *and its English Equivalents*. Ph.D. thesis, University of Silesia, Katowice.

Kański, Z. (1992). Impersonal constructions as a strategy for second-order predication. In M. Kefer and J. van der Auwera, editors, *Meaning and Grammar: Cross-Linguistics Perspectives*, volume 10 of *Empirical Approaches to Language Typology*, pages 95–121. Mouton de Gruyter, Berlin.

Kardela, H. (1986). *A Grammar of English and Polish Reflexives*. Uniwersytet Marii Curie-Skłodowskiej, Wydział humanistyczny, Lublin.

Kasper, R. and Rounds, W. (1986). A logical semantics for feature structures. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, Morrisontown, N.J. Association for Computational Linguistics.

Kathol, A. (1995). *Linearization-Based German Syntax*. Ph.D. thesis, Ohio State University.

Kayne, R. (1975). *French syntax: The Transformational Cycle*. MIT Press, Cambridge, MA.

Kim, J.-B. and Sag, I. A. (1995). The parametric variation of English and French negation. In *Proceedings of the Fourteenth West Coast Conference on Formal Linguistics*, pages 303–317.

King, P. J. (1989). *A Logical Formalism for Head-driven Phrase Structure Grammar*. Ph.D. thesis, University of Manchester.

King, P. J. (1999). Towards truth in HPSG. In V. Kordoni, editor, *Tübingen Studies in Head-Driven Phrase Structure Grammar*, Arbeitspapiere des Sonderforschungsbereichs 340 Bericht Nr. 132, Seminar für Sprachwissenschaft, Universität Tübingen, pages 301–352.

King, T. H. (1995). *Configuring Topic and Focus in Russian*. CSLI Publications, Stanford.

Klavans, J. L. (1985). The independence of syntax and phonology in cliticization. *Language*, **61**(1), 101–121.

Klemensiewicz, Z. (1960). *Podstawowe wiadomości z gramatyki języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa, third edition.

Klemensiewicz, Z. (1968). *Podstawowe wiadomości z gramatyki języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa, fifth edition.

Klemensiewicz, Z., Lehr-Spławiński, T., and Urbańczyk, S. (1955). *Gramatyka historyczna języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa.

Koenig, J.-P. and Davis, A. R. (1999). Linking as constraints on word classes in a hierarchical lexicon. In V. Kordoni, editor, *Proceedings of the ESSLLI-99 Workshop on* Lexical Semantics and Linking in Constraint-Based Theories, pages 1–45.

Koenig, J.-P. and Jurafsky, D. (1994). Type underspecification and on-line type construction in the lexicon. In R. Aranovich, W. Byrne, S. Preuss, and M. Senturia, editors, *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, volume 13, pages 270–285, Stanford University. CSLI Publications/SLA.

Kopcińska, D. (1997). *Strukturalny opis składniowy zdań z podmiotem-mianownikiem we współczesnej polszczyźnie*. Dom Wydawniczy Elipsa, Warszawa.

Kordoni, V. (1999). Lexical semantics and linking in HPSG: The case of psych verb constructions. In V. Kordoni, editor, *Tübingen Studies in Head-Driven Phrase Structure Grammar*, number 132 in Arbeitspapiere des Sonderforschungsbereichs 340, pages 494–527. Universität Tübingen.

Krasnowolski, A. (1911). *Główne zasady składni polskiej*. Wydawnictwo M. Arcta, Warszawa, second edition.

Kubiński, W. (1982). Polish *się* constructions and their English counterparts. *Papers and Studies in Contrastive Linguistics*, **XV**, 55–65.

Kubiński, W. (1987). *Reflexivization in English and Polish: An Arc Pair Grammar Analysis*, volume 178 of *Linguistische Arbeiten*. Niemeyer, Tübingen.

Kupść, A. (1999a). Clitic climbing and tense auxiliaries in Polish and Romance. In *Proceedings of 6th Annual TALN Conference*, pages 421–426, Cargése, Corse.

Kupść, A. (1999b). Clitic climbing in Polish tense auxiliary constructions. In A. Todirascu, editor, *Proceedings of the ESSLLI Student Session*, pages 43–52, Utrecht.

Kupść, A. (1999c). Clitic climbing in Polish verb clusters: an HPSG approach. In G.-J. M. Kruijff and R. T. Oehrle, editors, *Proceedings of Formal Grammar 1999*, pages 75–82, Utrecht.

Kupść, A. (1999d). Haplology of the Polish reflexive marker. In Borsley and Przepiórkowski (1999), pages 91–124.

Kupść, A. (1999e). Position of Polish clitics: An HPSG approach. In F. V. Eynde, I. Schuurman, and N. Schelkens, editors, *Computational Linguistics in the Netherlands 1998*, number 29 in Language and Computers: Studies in Practical Linguistics, pages 109–123. Amsterdam – Atlanta, GA.

Kupść, A. (1999f). A syntactic analysis of clitic climbing in Polish. Paper delivered at the HPSG99 conference, Edinburgh, Scotland August 1999.

Kupść, A. (2000a). A lexical account of clitic climbing in Polish. In P. Bański and A. Przepiórkowski, editors, *Proceedings of GLiP*, pages 94–103. Institute of Computer Science, Polish Academy of Sciences, Warsaw.

Kupść, A. (2000b). Lexical analysis of Polish multifunctional reflexive marker. In: T. H. King and I. Sekierina (eds.), Proceedings of the Eighth Annunal Workshop on Formal Approaches to Slavic Linguistics, Philadelphia, PA, May 21-23, 1999.

Kupść, A. (2000). The morphosyntax of Polish verbal negation: Towards an HPSG account. To appear in *Slavic Negation*, Sue Brown and Adam Przepiórkowski (eds.).

Kupść, A. and Przepiórkowski, A. (1997). Morphological aspects of verbal negation in Polish. In *Proceedings of the Second European Conference on Formal Description of Slavic Languages, Potsdam, Germany, November 20–22, 1997*. To appear.

Laka, I. (1990). *Negation in Syntax: On the Nature of Functional Categories and Projections*. Ph.D. thesis, Massachusetts Institute of Technology.

Lehr-Spławiński, T. and Kubiński, R. (1952). *Gramatyka języka polskiego*. Zakład narodowy im. Ossolińskich, sixth edition.

Levin, B. and Rappaport Hovav, M. (1995). *Unaccusativity: At the Syntax–Lexical Semantics Interface*. Massachusetts Institute of Technology, Cambridge, MA and London, UK.

Lidz, J. (2000). Condition R. *Linguistic Inquiry*, **32**(1). Unpublished manuscript. To appear. `http://www.cis.upenn.edu/~jlidz/pub.html`.

Lubowicz, A. (2000). Two views on Polish reflexives. In *Proceedings of WCCFL 18*. To appear.

Manning, C. D. and Sag, I. A. (1998). Argument structure, valence, and binding. *Nordic Journal of Linguistics*, **21**(2), 107–144.

Marciniak, M. (1999). Toward a binding theory for Polish. In Borsley and Przepiórkowski (1999), pages 125–147.

Menantaud, H. (1989). O negacji jako kategorii fleksyjnej. *Język Polski*, **6**, 361–366.

Meurers, W. D. (1994). On implementing an HPSG theory. In E. Hinrichs, D. Meurers, and T. Nakazawa, editors, *Partial VP and Split NP Topicalization in German: An HPSG Analysis*, volume 58 of *Arbeitspapiere des Sonderforschungsbereichs 340*.

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

Mikoś, M. and Moravcsik, E. A. (1986). Moving clitics in Polish and some crosslinguistic generalizations. *Studia Slavistica. Academiae Scientiarum Hungaricae*, **32**(1–4), 327–336.

Miller, P. H. (1992a). *Clitics and Constituents in Phrase Structure Grammar*. Garland, New York.

Miller, P. H. (1992b). Postlexical cliticization vs. affixation: Coordination criteria. In C. Canatsis, G. Chan, and J. Benton, editors, *Papers from the 28th Regional Meeting of the Chicago Linguistic Society*, Chicago. Chicago Linguistic Society.

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

Misz, H. (1966). Szyk *się* we współczesnej polszczyźnie pisanej. *Język Polski*, **46**, 102–110.

Monachesi, P. (1993). Restructuring verbs in Italian HPSG grammar. In K. Beals, G. Cooke, D. Kathman, S. Kita, K. McCullough, and D. Testen, editors, *Proceedings of the 29th Regional Meeting of the Chicago Linguistic Society*, pages 281–295, Chicago.

Monachesi, P. (1995). *A Grammar of Italian Clitics*. Ph.D. thesis, Tilburg University, Tilburg. ITK Dissertation Series 1995-3 and TILDIL Dissertation Series 1995-3.

Monachesi, P. (1998a). Italian restructuring verbs: A lexical analysis. In Hinrichs *et al.* (1998).

Monachesi, P. (1998b). The Morphosyntax of Romanian Cliticization. In P. A. Coppen, H. van Halteren, and L. Teunissen, editors, *Computational Linguistics in the Netherlands 1997. Selected Papers from the Eight CLIN Meeting*, number 25 in Language and Computers: Studies in Practical Linguistics, pages 99–118. Amsterdam – Atlanta, GA.

Monachesi, P. (1999). *A lexical approach to Italian cliticization.* CSLI Publications, Stanford.

Mykowiecka, A. (1998). *Opis składniowy polskich konstrukcji względnych w formalizmie HPSG.* Ph.D. thesis, Institute of Computer Science, Polish Academy of Sciences, Warszawa.

Nagórko, A. (1996). *Zarys gramatyki polskiej.* Pa'nstwowe Wydawnictwo Naukowe, Warszawa.

Nerbonne, J., Netter, K., and Pollard, C., editors (1994). *German in Head-Driven Phrase Structure Grammar.* Number 46 in CSLI Lecture Notes. CSLI Publications, Stanford.

Nevis, J. A. (1985). *Finnish Particle Clitics and General Clitic Theory.* Ph.D. thesis, Ohio State University. Published as *OSU WPL* 33, 1986.

Penn, G. (1998). Parametric types for typed attribute-value logic. In *Proceedings of th 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics, Montreal, Canada.*

Penn, G. (1999). Linearization and *WH*-extraction in HPSG: Evidence from Serbo-Croatian. In Borsley and Przepiórkowski (1999), pages 149–182.

Polański, K., editor (1993). *Encyklopedia językoznawstwa ogólnego.* Ossolineum, Wrocław.

Pollard, C. (1999). Strong generative capacity in HPSG. In G. Webelhuth, J.-P. Koenig, and A. Kathol, editors, *Lexical and Constructional Aspects of Linguistic Explanation*, pages 281–297. CSLI Publications, Stanford.

Pollard, C. and Sag, I. A. (1994). *Head-driven Phrase Structure Grammar.* Chicago University Press, Chicago.

Pollock, J.-Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, **20**, 365–424.

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). Case assignment in Polish: Towards an HPSG analysis. In Junghanns and Zybatow (1997), pages 307–319.

Przepiórkowski, A. (1999). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach.* Ph.D. thesis, Universität Tübingen, Germany.

Przepiórkowski, A. (2000a). Optional and multiple long distance genitive of negation in Polish. In P. Bański and A. Przepiórkowski, editors, *Proceedings of the First Generative Linguistics in Poland Conference*, pages 135–146. Institute of Computer Science, Polish Academy of Sciences, Warsaw.

Przepiórkowski, A. (2000b). Slavic linguistics formally — HPSG. To appear in G. Fowler (ed.), *The Future of Slavic Linguistics*, Slavica Publishers, Bloomington, Indiana.

Przepiórkowski, A. and Kupść, A. (1997a). Negative concord in Polish. Technical Report 828, Institute of Computer Science, Polish Academy of Sciences, Warsaw.

Przepiórkowski, A. and Kupść, A. (1997b). Unbounded negative concord in Polish: A lexicalist HPSG approach. In J. Landsbergen, J. Odijk, K. van Deemter, and G. V. van Zanten, editors, *Computational Linguistics in the Netherlands 1996: Papers from the Seventh CLIN Meeting*, pages 129–143, Eindhoven. IPO, Center for Research on User-System Interaction, Technische Universiteit Eindhoven.

Przepiórkowski, A. and Kupść, A. (1997c). Verbal negation and complex predicate formation in Polish. In R. C. Blight and M. J. Moosally, editors, *Proceedings of the 1997 Texas Linguistics Society Conference on the Syntax and Semantics of Predication*, volume 38 of *Texas Linguistic Forum*, pages 247–261, Austin, Texas.

Przepiórkowski, A. and Kupść, A. (1999). Eventuality negation and negative concord in Polish and Italian. In Borsley and Przepiórkowski (1999), pages 211–246.

Przepiórkowski, A. and Świdziński, M. (1997). Polish verbal negation revisited: A metamorphosis vs. HPSG account. Technical Report 829, Institute of Computer Science, Polish Academy of Sciences, Warsaw.

Rappaport, G. C. (1988). On the relationship between prosodic and syntactic properties of pronouns in the Slavic languages. In A. M. Schenker, editor, *American Contribution to the Tenth International Congress of Slavists*, pages 301–327. Slavica Publishers, Columbus, Ohio.

Rappaport, G. C. (1998). Clitics as features: A non-semiotic approach. In R. A. Maguire and A. Timberlake, editors, *American Contributions to the Twelfth International Congress of Slavists*, pages 460–478. Slavica Publishers.

Reape, M. (1992). *A Formal Theory of Word Order: A Case Study in West Germanic*. Ph.D. thesis, University of Edinburgh.

Recourcé, G. (1995). *L'association négative en français. Etude linguistique et formelle de la particule* ne. Ph.D. thesis, Université Paris 7, UFRL.

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

Reinhard, S. (2000). *Deverbale Komposita an der Morphologie-Syntax-Semantik-Schnittstelle: ein HPSG-Ansatz*. Ph.D. thesis, Universität Tübingen. In progress.

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

Richter, F. and Sailer, M. (1999a). LF conditions on expressions of Ty2: An HPSG analysis of Negative Concord in Polish. In Borsley and Przepiórkowski (1999), pages 247–282.

Richter, F. and Sailer, M. (1999b). On the left periphery of German finite sentences. To appear in a CSLI volume edited by Tibor Kiss and Detmar Meurers.

Riehemann, S. (1998). Type-based derivational morphology. *Journal of Comparative Germanic Linguistics*, **2**, 49–77.

Rivero, M. L. (1999). Impersonal SIĘ in Polish: a pronominal anaphor. Unpublished manuscript, February 1999.

Rizzi, L. (1982). *Issues in Italian Syntax*. Foris Publications, Dordrecht.

Rozwadowska, B. (1992). *Thematic constraints on selected constructions in Polish and English*. Wydawnictwa Uniwersytetu Wrocławskiego, Wrocław.

Rubach, J. and Booij, G. E. (1985). A grid theory of stress in Polish. *Lingua*, **66**, 281–319.

Safarewiczowa, H. (1954). Rzeczowniki "zwrotne" w języku polskim. *Język Polski*, **XXXIV**(5), 332–348.

Sag, I. A. (1997). English relative clause constructions. *Journal of Linguistics*, **33**(2), 431–483.

Sag, I. A. and Fodor, J. D. (1994). Extraction without traces. In R. Aranovich, W. Byrne, S. Preuss, and M. Senturia, editors, *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, Stanford University. CSLI Publications/SLA.

Sag, I. A. and Godard, D. (1993). Extraction of *De*-phrases from the French NP. In *Proceedings of the North East Linguistic Society 24*, Department of Linguistics, University of Massachusetts. GLSA.

Sag, I. A. and Wasow, T. (1999). *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford.

Sailer, M. (2000). *Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar*. Ph.D. thesis, Universität Tübingen.

Saloni, Z. (1976). *Cechy składniowe polskiego czasownika*. Ossolineum, Wrocław.

Saloni, Z. and Świdziński, M. (1985). *Składnia współczesnego języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, second edition.

Saloni, Z. and Świdziński, M. (1998). *Składnia współczesnego języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, fourth (changed) edition.

Schenker, A. (1993). Does the Polish reflexive exist? In G. Hentschel and R. Laskowski, editors, *Studies in Polish Morphology and Syntax*, volume 99 of *Specimina Philologiae Slavicae*, pages 63–70. Verlag Otto Sagner, München.

Schenker, A. M. (1985). W sprawie *się* raz jeszcze. *Język Polski*, **LXV**(1), 9–24.

Spencer, A. (1991). *Morphological Theory*. Blackwell, Oxford.

Śpiewak, G. and Szymańska, I. (1995). A poll on Pollock — functional categories in Polish. In E. Gussmann, editor, *Licensing in Syntax and Phonology*, volume 1 of *PASE Studies and Monographs*, pages 125–145. Folium, Lublin.

Sproat, R. (1992). *Morphology and Computation*. The MIT Press, Cambridge, Mass.

Stalmaszczyk, P., editor (1998). *Projections and Mapping: Studies in Syntax*, volume 5 of *PASE Studies and Monographs*. Folium, Lublin.

Świdziński, M. (1992). *Gramatyka formalna języka polskiego*, volume 349 of *Rozprawy Uniwersytetu Warszawskiego*. Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Świdziński, M. (1994). Syntactic dictionary of Polish verbs. Unpublished manuscript.

Szober, S. (1953). *Gramatyka języka polskiego*. Nasza Księgarnia, Warszawa, third edition.

Szpakowicz, S. (1978). *Automatyczna analiza składniowa zdań pisanych*. Ph.D. thesis, Uniwersytet Warszawski, Warszawa.

Szpakowicz, S. (1986). *Formalny opis składniowy zdań polskich*. Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Szpakowicz, S. and Świdziński, M. (1981). Zarys klasyfikacji schematów zdaniowych we współczesnej polszczyźnie pisanej. *Polonica*, **VII**, 5–35.

Szymańska, I. (1998). Polish reflexiva tantum: A construction-based interpretation. In Stalmaszczyk (1998), pages 99–118.

Szymańska, I. and Śpiewak, G. (1998). Some remarks on the role of *się* in organising the argument structure of Polish verbal predicates. In Stalmaszczyk (1998), pages 119–136.

Szymczak, M. (1975). *Słownik ortograficzny języka polskiego*. Pa'nstwowe Wydawnictwo Naukowe, Warszawa, first edition.

Trask, R. L. (1993). *A Dictionary of Grammatical Terms in Linguistics*. Routledge, London.

van Noord, G. (1999). *Hdrug Reference Manual*. University of Groningen.

van Noord, G. and Bouma, G. (1994). Adjuncts and the processing of lexical rules. In *Fifteenth International Conference on Computational Linguistics (COLING '94)*, pages 250–256, Kyoto, Japan.

Wanner, D. (1977). On the order of clitics in Italian. *Lingua*, **43**, 101–128.

Wechsler, S. (1995). *The Semantic Basis of Argument Structure*. CSLI Publications, Stanford.

Wilczewska, K. (1966). Czasowniki zwrotne we współczesnej polszczyźnie. *Towarzystwo Naukowe w Toruniu, Prace Wydziału Filologiczno-Filozoficznego*, **XVII**(3).

Willim, E. (1990). On case-marking in Polish. *Papers and Studies in Contrastive Linguistics*, **XXV**, 203–220.

Witkoś, J. (1993). *Some Aspects of Phrasal Movement in English and Polish*. Ph.D. thesis, Uniwersytet im. Adama Mickiewicza, Poznań.

Witkoś, J. (1996a). On NegP and the structure of the Polish clause. *Papers and Studies in Contrastive Linguistics*, **XXXI**, 65–96.

Witkoś, J. (1996b). Pronominal argument placement in Polish. *Wiener Linguistische Gazette*, **57–59**, 147–194.

Witkoś, J. (1998). *The Syntax of Clitics: Steps towards a Minimalist Account.* motivex, Poznań.

Zwicky, A. M. (1977). *On Clitics.* Indiana Univerity Linguitics Club, Bloomington.

Zwicky, A. M. (1985). Clitics and particles. *Language*, **61**(2), 283–305.

Zwicky, A. M. (1987). Suppressing the Zs. *Journal of Linguistics*, **23**, 133–148.

Zwicky, A. M. and Pullum, G. K. (1983). Cliticization vs. inflection: English *n't*. *Language*, **59**(3), 502–513.