

A New HPSG Approach to Polish Auxiliary Constructions

Anna Kupść<sup>1,2</sup> and Jesse Tseng<sup>1</sup>

<sup>1</sup> CNRS, Loria UMR 7503

<sup>2</sup> Polish Academy of Sciences, Institute of Computer Science

Proceedings of the HPSG05 Conference

Department of Informatics, University of Lisbon

Stefan Müller (Editor)

2005

CSLI Publications

<http://csli-publications.stanford.edu/>

## Abstract

This paper presents an analysis of constructions involving the *l*-form of the verb in Polish, including primarily the past tense, the conditional mood, and the future tense. Previous approaches have attempted to treat these uniformly as auxiliary verb constructions. We argue against a unified treatment, however, in light of synchronic and diachronic evidence that indicates that only the future tense and the conditional still involve auxiliaries in modern Polish. We show that the past tense is now a simple tense, although the *l*-forms appear in combination with agreement affixes that can appear in different places in the sentence. We provide an account of the common linearization properties of the past tense markings and the conditional auxiliary. We present a detailed HPSG analysis of the past tense construction that relies on the introduction of two interacting agreement features. We then discuss the consequences of our proposals for the analysis of the conditional and future auxiliary constructions, and finally, we offer a treatment of constructions involving inflected complementizers in Polish.

## 1 Introduction

The “*l*-participle” form of the verb in Polish (for short: *l*-form, so called because it ends in *l* or *ł*, usually followed by a vowel) is inflected for number and gender and agrees with the subject. As an example, the different *l*-forms for the verb *czytać* ‘read’ are as follows:

- (1) singular: *czytał* (masculine), *czytała* (feminine), *czytało* (neuter);  
plural: *czytali* (masculine human), *czytały* (other).

The *l*-form can appear in the past tense, in the conditional mood, and in the future tense. In the past tense, the *l*-form requires additional endings in 1st and 2nd persons: 1sg *-(e)m*, 2sg *-(e)ś*, 1pl *-śmy*, and 2pl *-ście*, cf. (2a–b).

- (2) a. Ja *czytałem* książkę. My *czytaliśmy* książkę.  
I read.m.sg-1sg book we read.m.pl-1pl book  
b. Ty *czytałeś* książkę. Wy *czytaliście* książkę.  
you read.m.sg-2sg book you read.m.pl-2pl book  
c. On *czytał* książkę. Oni *czytali* książkę.  
he read.m.sg book they read.m.pl book  
‘I/you/he/we/you/they read a book.’

In conditional constructions, the *l*-form appears in combination with the element *by*. In this case, it is *by* that takes the personal endings in 1st and 2nd persons:

---

<sup>†</sup>We would like to thank three anonymous reviewers for the HPSG Conference, the audiences of the IPIAN Linguistic Engineering Group seminar (Warsaw, April 2005) and the HPSG Conference (Lisbon, August 2005), where versions of this paper were presented.

- (3) a. Ja *bym* czytał książkę. My *byśmy* czytali książkę.  
 I CND-1sg read.m.sg book we CND-1pl read.m.pl book
- b. Ty *byś* czytał książkę. Wy *byście* czytali książkę.  
 you CND-2sg read.m.sg book you CND-2pl read.m.pl book
- c. On *by* czytał książkę. Oni *by* czytali książkę.  
 he CND read.m.sg book they CND read.m.pl book
- ‘I/you/he/we/you/they would read a book.’

Finally, in the future tense, the *l*-form combines with future forms of the auxiliary *być* ‘be’, (4).<sup>1</sup> In this use, however, we do not find the 1st and 2nd person endings that characterize the past tense and the conditional.

- (4) a. Ja *będę* czytał książkę. My *będziemy* czytali książkę.  
 I FUT.1sg read.m.sg book we FUT.1pl read.m.pl book
- b. Ty *będziesz* czytał książkę. Wy *będziecie* czytali książkę.  
 you FUT.2sg read.m.sg book you FUT.2pl read.m.pl book
- c. On *będzie* czytał książkę. Oni *będą* czytali książkę.  
 he FUT.3sg read.m.sg book they FUT.3pl read.m.pl book
- ‘I/you/he/we/you/they will read a book.’

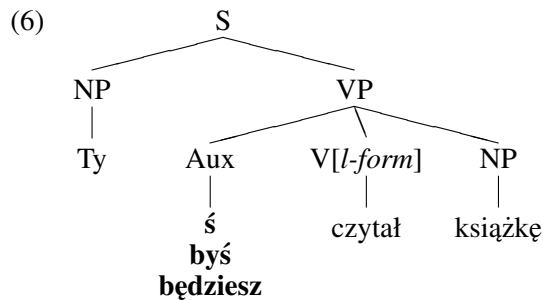
In the past tense, the endings can be attached directly to the *l*-form (*agglutination*) as in (2), or they can appear at a distance, somewhere to the left (*tmesis*, (5)). In the latter configuration, the past tense resembles the conditional and the future, which also involve a “bare” *l*-form.

- (5) a. *Jam* czytał książkę. *Myśmy* czytali książkę.  
 I-1sg read.m.sg book we-1pl read.m.pl book
- b. *Tys* czytał książkę. *Wyście* czytali książkę.  
 you-2sg read.m.sg book you-2pl read.m.pl book

Some previous accounts of Polish verbal constructions, e.g., Borsley and Rivero (1994), Borsley (1999), Kupść (2000), have attempted to provide a unified analysis of the three uses of *l*-form verbs in (3)–(5) as auxiliary constructions, i.e., they are treated as a syntactic combination of the *l*-form with an auxiliary verb, shown schematically in (6).<sup>2</sup>

<sup>1</sup>The future tense can also be formed with the infinitive, instead of the more recent *l*-form construction; the same auxiliary is used in both cases.

<sup>2</sup>In fact, not all the authors assume the same syntactic structure for all the uses or a flat structure as in (6), but these differences are irrelevant for the current discussion.



Such an account, however, overlooks the fact that in the past tense there is no auxiliary for the *l*-form to combine with in the 3rd person. Recall that there is no ending in the 3rd person, singular or plural (2c), and thus, the syntactic structure in (6) is inapplicable in these cases (unless, of course, an empty category is assumed).<sup>3</sup>

In fact, there is further evidence against a uniform treatment of the three constructions involving the *l*-form. In the next section, we will focus on the divergent properties of the past tense and conditional constructions.

## 2 Empirical Observations

There are a number of crucial differences between the conditional particle *by* and the past tense markings that suggest strongly that they do not have the same grammatical status.

### 2.1 Past tense

In the past tense, the personal markings have different properties when they are attached to or detached from the *l*-form—compare (2) and (5).

When the personal markings directly follow the *l*-form, they induce morphophonological changes in their host. With a masculine singular subject, an epenthetic vowel *e* must be inserted before the singular markings *-m* and *-ś*, cf. (2a-b) and (7a). This creates an additional syllable, which results in stress shift, and with certain verbs, leads to a vowel shift *ó* to *o* (7a).<sup>4</sup> In the plural, the addition of the markings *-śmy*, and *-ście* can, for some speakers or in fast speech, shift the stress one syllable to the right (7b).

- (7) a. P**O**mógl → po**MO**gł**e**m  
 help.m.sg help.m.sg-1sg
- b. po**MO**gli → ?pomog**LI**śmy  
 help.m.pl help.m.pl-2pl

<sup>3</sup>The striking absence of a 3rd person ending in modern Polish has a historical explanation, which will be briefly sketched in §3.

<sup>4</sup>Capital letters mark lexical stress. With isolated exceptions, words in Polish have penultimate stress.

These observations suggest that the postverbal markings are suffixes. Another piece of evidence comes from coordination data, discussed also by Bański (2000). The personal ending has to be repeated on all conjuncts if it is realized to the right of the *l*-verb (8). (For some speakers this requirement can be relaxed in the plural).

- (8) Często [czytałem i pisał\*(em)].  
often read-1sg and write(-1sg)  
'I often read and wrote.'

According to the criteria of Miller (1992), the obligatory repetition of an item in coordination speaks in favor of its affix status. Therefore, (8) further supports the suffix status of the personal markings in agglutinated past tense forms.

When the personal markings are realized at a distance from the *l*-form, they are quite particular about the phonological properties of their host. Bański (2000) characterizes the conditions in terms of phonological "friendliness" between the host and the marking. Broadly speaking, the host must end in a vowel or another highly sonorous segment, but the different markings impose specific constraints, which are subject to wide variation among speakers (especially in the plural). The restrictions seem to be weaker than Bański suggests:

- (9) a. The 1sg marking (-*m*) can only follow a word ending in a non-nasal vowel (i.e., not *ɛ* or *ɔ*), or (possibly) the glide *j*;  
b. The 2sg marking (-*ś*) can additionally (but somewhat marginally) follow a nasal vowel or *j*, and possibly the sonorants *l*, *r*, *l* in a simple coda;  
c. The 1-2pl forms (-*śmy* and -*ście*) can follow any vowel, but words ending in a single consonant other than a sibilant (e.g., *sz*, *ż*, *cz*) are also potential (marginal) hosts.

Such combinatory restrictions are common for affixes but not for sequences of syntactic items. It should be noted that the evidence is less compelling for the plural marking.

Epenthetic *e*-insertion before -*m* and -*ś* is only possible with a few lexical items, e.g., *już* 'already', *jak* 'as', *tam* 'there' or *choć* 'although', and the resulting suffixed forms (e.g., *jużem*, *jakeś*) sound distinctly archaic. There is no vowel change or stress shift with hosts other than the *l*-form. Finally, wide scope over coordination is possible (10) (Bański (2000) overlooks this possibility).

- (10) Często*m* [czytał i pisał].  
often-1sg read and write  
'I was often reading and writing.'

The possibility of wide scope over coordination does not distinguish between affix and syntactic clitic status, according to the criteria in Miller (1992).

## 2.2 Conditional constructions

The placement of the conditional element *by* is analogous to that of the personal endings in the past tense: it can be attached directly to the *l*-form, e.g., *czytałbym*, *czytałbyś*, *czytaliby*, otherwise it has to appear somewhere to its left, as illustrated in (3). Other properties of conditional constructions, however, are quite different from the past tense.

First, unlike in the past tense, the element *by* is present in all persons (3): 1sg *bym*, 2sg *byś*, 1pl *byśmy*, 2pl *byście*, 3sg/pl *by*. Second, the forms of conditional *by* are phonologically weak, but they impose no phonological restrictions on the preceding word. They can follow a word ending in any segment (i.e., any of the vowels and consonants that appear word-finally in Polish); this is the same behavior as observed for Polish pronominal clitics. Also, the presence of conditional *by* never has a morphophonological effect on the preceding material (again, as in the case of pronominal clitics, e.g., Dłuska (1974), Rappaport (1988)). Finally, the conditional particle can take wide scope over a coordination of VPs in both preverbal (11a) and postverbal (11b) positions.

- (11) a. Często *bym* [czytał i pisał].  
often CND-1sg read and write
- b. Często [czytałbym i pisał(*bym*)].  
often read-CND-1sg and write(-CND-1sg)  
'I would often read and write.'

According to Miller (1992), the optional repetition of the postverbal conditional particle in (11b) excludes an affix analysis. On the other hand, the wide scope over coordination in preverbal positions, (11a), does not distinguish between affix and syntactic clitic status.

## 2.3 Common properties

Despite the differences, there are also some similarities between the past tense and conditional forms. First of all, the *l*-form in the past tense and the conditional element *by* take identical personal endings: *by-m*, *by-ś*, *by-śmy*, *by-ście*.

The past tense markings and the forms of *by* are subject to the same placement restrictions: they can immediately follow the *l*-form, e.g., (2) and (11b), or they appear somewhere to its left, (5) and (3), but without escaping from the clausal projection of the *l*-form, (12). Also, all positions further to the right of the *l*-form are excluded (13). There are similar constraints on the position of Polish pronominal clitics with respect to the verb (Kupść, 2000).

- (12) a. Mówi, [że ty przeczytałeś / przeczytałbyś tę książkę].  
says that you read-2sg / read-CND-2sg this book  
'(S)he says that you read / would have read this book.'
- b. \*Mówiś / byś, [że ty przeczytał tę książkę].  
says-2sg / CND-2sg that you read this book

- (13) \* Ty przeczytał książkęś / byś.  
 you read book-2sg / CND-2sg  
 ‘You read / would have read the book.’

The past tense endings and the forms of *by* always require a prosodically appropriate host. An immediate consequence of this is that they can never appear sentence-initially:<sup>5</sup>

- (14) \* Ś / Byś ty przeczytał książkę.  
 2sg / CND-2sg you read book  
 ‘You read / would have read the book.’

## 2.4 Summary

These observations suggest that the past tense endings, both in pre- and postverbal positions, are much more closely bound to the preceding word than the conditional particle. In fact, their behavior is more typical of morphological suffixes than of independent syntactic items. Therefore, we will treat the past tense markings as inflectional elements. On the other hand, the forms of conditional *by* are syntactic words, but they are clitics, subject to special word order constraints.

The data presented above highlight distinct properties of conditional and past tense constructions and indicate that, despite certain similarities, the two constructions should be analyzed independently. In §4, we will present a proposal along these lines.

Before continuing to the next section, we should mention one final construction in Polish involving the *l*-form, illustrated by the embedded clauses in the following examples:

- (15) a. On powiedział mi, żebym / abym / bym ja przyszedł.  
 he say me.dat COMP-1sg I come  
 ‘He told me to come.’
- b. Ostrzegam was, żebyście / abyście / byście tego nie robili.  
 warn.1sg you, COMP-2pl that NEG do  
 ‘I warn you not to do that.’
- c. Nie chcę, żebyśmy / abyśmy / byśmy się spóźnili.  
 NEG want.1sg COMP-1pl REFL be-late  
 ‘I don’t want us to be late.’

“Subjunctive” clauses of this type are usually described as special uses of the conditional mood (Swan, 2002), but in fact they have quite different properties, as Borsley (1999) points out. They do involve an element superficially identical to

<sup>5</sup>There is a complementizer homophonous with *byś* in (14) that does appear initially (it is not a clitic); see the discussion of (15).

conditional *by*, with the same inflectional endings. But in this construction, its position is completely fixed: it must introduce the clause, possibly preceded by *że-* or *a-*, from which it cannot be separated. We agree with Borsley's conclusion that these cases cannot involve the conditional auxiliary, and that the forms *żeby*, *aby*, and *by* should be treated as complementizers that introduce clauses with an *l*-form verb, and they show the peculiar property of inflecting to agree with the subject of this verb. We do not adopt Borsley's analysis of these forms, however; we return to this question at the end of the paper.

### 3 Historical Development

In order to better understand the complex behavior of the Polish past tense forms, we sketch their historical development, based on Andersen (1987).

The modern Polish past tense endings evolved from Old Polish forms of the auxiliary verb BE given in (16).

(16)	Old Polish		Modern Polish	
	strong	weak	weak	strong
1sg	jeśm	-(e)śm / (e)m	-(e)m	jestem
2sg	jeś	-(e)ś	-(e)ś	jesteś
3sg	jest / je	∅	∅	jest
1pl	jesm(y)	-(e)smy	-(e)śmy	jesteśmy
2pl	jeście	-(e)ście	-(e)ście	jesteście
3pl	są	∅	∅	są

In Old Polish, there were two forms of BE: strong (orthotonic) and weak (phonologically reduced, atonic). The modern Polish past tense endings evolved from the Old Polish weak forms of BE. Note that already in Old Polish (13th century), there was no weak form in the 3rd person. The modern Polish strong form of BE serves only as a present tense form and cannot participate in the formation of the past tense.

The origins of the modern Polish past tense date back to pre-Polish. At that time, only one (strong) form of BE was available and the counterpart of the contemporary past tense was a construction formed by the *l*-participle and the auxiliary BE (e.g., 3sg: [mlūvilū jestū] '(he) has said').

In Old Polish, the two forms of BE (16) could participate in the formation of the past tense. The latter were more common and indicated the unmarked use, whereas the strong 3rd person forms *jest* (sg.) and *są* (pl.) could be added for emphasis in all persons (with agreement in number with the subject):

- (17) a. 1sg: [...-(e)m ... mówił (jest)]  
 b. 1pl: [...-(e)smy ... mówili (są)]  
 c. 3sg: [mówił (jest)], 3pl: [mówili (są)]

Old Polish had no weak form in the 3rd person and so only emphatic constructions were still auxiliary constructions. The placement of the weak BE was quite

rigid and it could appear only after the first stressed word of the clause (“second position”).

In early modern Polish, the optional emphatic forms *jest/są* fall out of use, while weak forms of BE are no longer restricted to second position, but instead they can appear after any stressed word in the clause (to the left of the verb). In the modern language, the original forms of the auxiliary BE have been reinterpreted as person/number agreement markings and the *l*-participle has become a finite (non-present) verb. Postverbal position is also possible, i.e., the *l*-form combines directly with the personal marking (agglutination).

Andersen (1987) quotes statistics from Rittel (1975) indicating that in current Polish there is a strong preference for the agglutinated forms (e.g., *czytałem*), while the endings in other positions appear much less frequently. Hence, the past tense markings in Polish are still undergoing a centuries long transition from second-position clitics to verbal affixes.

## 4 Proposed Analysis

### 4.1 Auxiliaries

#### 4.1.1 The conditional auxiliary *by*

Given the empirical properties identified in the previous sections, we believe that Borsley (1999) and Kupść (2000) are correct in treating inflected forms of conditional *by* appearing to the left of the *l*-form verb as auxiliary verbs, and in fact we extend the same analysis to *by* when it appears immediately to the right of the *l*-form. In all cases, the forms of *by* satisfy the following partial lexical description:

$$(18) \left[ \begin{array}{l} \text{word} \\ \left[ \begin{array}{l} \text{clitic} \\ \left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \text{verb} \\ \text{VFORM } \textit{cond} \\ \text{AUX } + \\ \text{NEG } - \end{array} \right] \\ \text{ARG-ST } \left\langle \begin{array}{l} \boxed{1} \text{ NP, VP} \left[ \begin{array}{l} \text{HEAD | VFORM } \textit{l-form} \\ \text{SUBJ } \langle \boxed{1} \rangle \\ \text{COMPS } \langle \rangle \end{array} \right] \end{array} \right\rangle \end{array} \right] \end{array} \right] \end{array} \right]$$

This description essentially reflects the standard HPSG analysis of auxiliaries as subject-raising verbs, selecting a verbal complement with a specific verbal form. As observed in Kupść (2000), there is no direct evidence for the flat structure of conditional auxiliary constructions postulated in Borsley (1999), and so we assume simple VP-complementation here. The feature  $[-\text{NEG}]$  is specified in order to

ensure that *by* is never negated; in conditional structures in Polish, negation can only be expressed (in the form of the element *nie*) on the *l*-form:

- (19) a. \*Ty nie *byś* wczoraj widział tego filmu  
 you NEG CND-2sg yesterday see that film  
 b. Ty *byś* wczoraj nie widział tego filmu.  
 you CND-2sg yesterday NEG see that film  
 ‘You wouldn’t have seen that film yesterday.’

#### 4.1.2 [ $\pm$ CLITIC-HOST]

As specified in (18), the forms of *by* are syntactic clitics, and therefore subject to particular linearization constraints. First of all, *by* must appear either in the sentence field<sup>6</sup> to the left of the *l*-form verb, or immediately following the *l*-form. Its exact position is determined primarily by prosodic structure (see for example Mikoś and Moravcsik (1986) and Bański (2000)). We believe that a DOMAIN-based analysis (Reape, 1994) is the best way to handle the linearization possibilities, although we cannot offer a full account in this paper. We simply introduce a shorthand boolean feature CL(ITIC)-HOST to identify words that satisfy (marked [+CL-HOST]) or do not satisfy ([−CL-HOST]) the prosodic and other conditions for hosting a clitic immediately to the right. Typical clitic hosts include subject pronouns, *wh*-words, and complementizers, but in principle the range of possibilities is very large. Monosyllabic prepositions are typical words which are [−CL-HOST].

Non-prosodic conditions on CL-HOST are most apparent in the post-verbal sentence field. All verbs (including *l*-forms) can be [+CL-HOST], so clitics such as conditional *by* and pronominal clitics<sup>7</sup> can appear immediately to their right. But after the rightmost verb in a clause (and the clitics that it hosts, if any), all other words are [−CL-HOST]. Consequently, no clitics (or the past tense endings) can appear in this field—recall also example (13):

- (20) a. My zobaczyli(*by*)śmy dawno słońca.  
 we see(-CND)-1pl long-ago elephant  
 ‘We saw / would have seen an elephant a long time ago.’  
 b. \*My zobaczyli dawnośmy / byśmy słońca.  
 c. \*My widzieli dawno słońcaśmy / byśmy.

The suffixed adverb and noun in the last two examples are phonologically well-formed, and they would be grammatical in the preverbal sentence field. But here they are both [−CL-HOST], and this is a constraint determined simply by linear

<sup>6</sup>We use the term “field” in a purely descriptive way, without suggesting that any version of the topological fields approach, as used for the analysis of German word order, would be applicable to Polish.

<sup>7</sup>As argued in Kupś (2000), Polish pronominal clitics are syntactic items.

order, one that cannot be overridden by prosodic or syntactic considerations in this case.<sup>8</sup>

To account for clitic clusters (including those immediately to the right of the *l*-form verb), we assume that clitics can themselves be [+CL-HOST] and host clitics to their right. As noted in Witkoś (1997), the relative order of pronominal and conditional clitics is very constrained as pronominal clitics tend to follow rather than precede the conditional auxiliary, (21a) vs. (21b).

- (21) a. Ty *byś* go widział. / Ty widział*byś* go.  
           you CND-2sg him.cl see        you see-CND-2sg him.cl
- b. ?\* Ty go *byś* widział. / \*Ty widział go *byś*.  
           you him.cl CND-2sg see        you see        him.cl CND-2sg

Borsley (1999) argues for two different analyses of *by*, depending on its position to the left or to the right of the *l*-form. For the combination of an *l*-form verb followed by *by*, he proposes a special rule of morphological compound formation. We see no evidence (stress shift or vowel quality alternations, for example) to motivate a distinct treatment of *by* in this case. In our account, *by* is always a clitic, and has to be hosted by a [+CL-HOST] element, whether this happens to be the *l*-form verb itself or some other word to the left.

#### 4.1.3 The future auxiliary

The forms of the future auxiliary (4) differ from conditional *by* in that they are full syntactic words, not clitics. It therefore does not depend on a [+CL-HOST] element, and can appear in a wider range of positions, in particular to the right of the *l*-form and in sentence-initial position:<sup>9</sup>

- (22) a. (Ty) *będziesz* widział ten film.  
           (you) FUT.2sg see        this film  
           ‘You will see this film.’
- b. Ty widział *będziesz* ten film. / Ty widział ten film *będziesz*.  
           you see        FUT.2sg this film    you see        this film FUT.2sg

Furthermore, the VP complementation proposed for *by* in (18) may be inadequate for the future auxiliary. “Clitic climbing” phenomena in Polish, which are discussed in Kupść (2000), may be better analyzed by assuming complement raising and a flat structure, as proposed for French auxiliary constructions by Abeillé and Godard (2002). (We will not go into the details in this paper.)

<sup>8</sup>But see the discussion of (32) below.

<sup>9</sup>Swan (2002) claims that the inverted order illustrated in (22b) is only possible if the auxiliary combines with an infinitival complement (see fn. 1), but in fact an *l*-form is also possible, as this example shows.

## 4.2 Past tense agreement markings

Like the conditional auxiliary, the past tense elements *-m*, *-ś*, *-śmy*, and *-ście* can appear either in the preverbal sentence field, or immediately to the right of the *l*-form verb. Although these two options result in superficially distinct constructions, we present a single analysis that covers both cases.

### 4.2.1 Floating suffixes

We begin with the analysis of past tense elements that “float” in the preverbal sentence field. We are guided by the following empirical observations, discussed in detail in §2.

- The past tense elements are not independent syntactic items in modern Polish, but suffixes.
- These suffixes can appear in a variety of positions and attach to a variety of hosts, and the possibilities cannot be effectively characterized in terms of syntactic category, syntactic function, or phrase structure.
- The suffixes are obligatory with 1st and 2nd person subjects, but completely absent in the 3rd person.

Taken together, these facts pose serious problems for any analysis of the past tense elements as auxiliary verbs. In fact, we find the last point to be a convincing argument on its own, but technically it is not an insurmountable obstacle. One could appeal to a phonologically empty auxiliary, or propose an auxiliary-less account just for the 3rd person, as Borsley (1999) does.

The first two points provide arguments against an auxiliary treatment of cases where the past tense elements are actually present (in the 1st and 2nd persons). One might suggest, for example, that the Polish phenomenon is similar to auxiliary contraction in English (e.g., *I'll*, *we've*, *you'd*). But the English facts are much simpler, in that the contracted auxiliary always appears in the same position as the full auxiliary, and it always contracts with the subject. In Polish, the varied placement possibilities for the past tense elements and the wide range of possible hosts make the analyses proposed for English (see Bender and Sag (2001) and references therein) inapplicable. Another crucial difference is that in Polish, the past tense suffixes have no corresponding full form; this seems to exclude an analysis where the forms are produced by late phonological reduction (as proposed for syllabic contracted auxiliaries in English), because such processes are not usually obligatory.

We therefore reject the auxiliary approach. We consider the past tense elements to be personal agreement markings; they therefore do not have syntactic head or functor status. We treat the past tense in Polish as a simple tense, with the *l*-form as the head of the structure. A uniform analysis applies in all three persons, but the *l*-form requires the presence of an agreement marking in the 1st and 2nd persons.

The past tense agreement markings are unlike ordinary suffixes, which attach to a particular kind of host. They cannot be analyzed as phrasal suffixes, either; phrasal affixes do combine with a variety of lexical hosts, but they can be characterized as combining with a specific type of phrasal host (NP, VP, etc.), and appearing in a specific position with respect to this phrase (at the left or right edge, typically). This is not the case for the past tense suffixes. In some sense they could be thought of as sentential or clausal affixes, but again, they do not occupy a fixed position in the sentence/clause. Since no existing technical machinery seems to cover this kind of behavior, we have to introduce special mechanisms for the realization of the past tense suffixes (at the morphological level) and for the propagation of information about their presence (in the syntax).

The realization of the floating agreement marking is subject to a strict surface order constraint: it must appear exactly once, somewhere to the left of the *l*-form verb. And unlike in ordinary cases of agreement, no particular word or constituent is targeted to receive the marking. The host can be of practically any category (but never a verb), it can be a complement, adjunct, filler, or complementizer, or embedded inside another phrase with one of these grammatical functions (and within this phrase, the host of the suffix can be the head, the specifier, an adjunct, etc.):

- (23) a. *Dlaczegoś tu przyszedł?* (suffix host: *wh*-adverb filler)  
 why-2sg here come  
 ‘Why did you come here?’
- b. [*Bardzo często*] *widział ten film.* (head of AdvP adjunct)  
 very often-1sg see that film  
 ‘I have seen that film very often.’
- c. ... *ale* *widział ten film wczoraj.* (conjunction)  
 but-1sg see that film yesterday  
 ‘... but I saw that film yesterday.’
- d. *Już dawno [dobrego filmu] nie widział.* (adjunct in NP object)  
 long time good-1sg film NEG see  
 ‘I haven’t seen a good film in a long time’

To begin with, then, we need a mechanism to introduce the agreement markings morphologically. The following constraint partitions words into suffixed and unsuffixed classes:

$$(24) \text{ word} \Rightarrow \left[ \begin{array}{l} \text{PHON } \mathbf{F}_{agr}(\underline{1},\underline{2},\underline{3}) \\ \text{MORPH } \left[ \text{FORM } \underline{1} \right] \\ \text{SS } \left[ \begin{array}{l} \text{LOC | CAT | HEAD } \underline{2} \\ \text{CL-HOST } + \\ \text{AGR-MARK } \mathbf{F}_{I2}(\underline{3}) \end{array} \right] \end{array} \right] \vee \left[ \begin{array}{l} \text{PHON } \underline{4} \\ \text{MORPH } \left[ \text{FORM } \underline{4} \right] \\ \text{SS } \left[ \text{AGR-MARK } \langle \rangle \right] \end{array} \right]$$

$$(25) \quad \mathbf{F}_{12}(\langle \mathbb{I} \text{ [PER 1st} \vee \text{ 2nd]} \rangle) = \langle \mathbb{I} \rangle$$

$$\mathbf{F}_{12}(\langle \text{[PER 3rd]} \rangle) = \langle \rangle$$

The first disjunct in (24) corresponds to suffixed words. Note first of all that these words are required to be [+CL-HOST]; this constrains the possible surface positions of agreement markings, just as for the clitic *by* discussed in the preceding section. The phonological realization of the suffixed word is determined by the function  $\mathbf{F}_{agr}$ , which takes into account the host word's morphological form, its HEAD value, and the index of the personal suffix to be realized. The function has to have access to the HEAD value because the phonological properties and effects of suffixation depend on the identity of the host (whether it is an *l*-form or not), as discussed in §2. In particular, the definition of  $\mathbf{F}_{agr}$  incorporates the phonological restrictions on the host identified in (9). For incompatible combinations (e.g., a word ending in a consonant like [t] cannot take any suffix), the function is undefined and no valid description can be constructed.

We introduce a list-valued attribute AGR-MARK to record the presence and identity of the agreement suffix. The function  $\mathbf{F}_{12}$  serves as a filter to make sure that only 1st and 2nd person suffixes are recorded.<sup>10</sup> The second disjunct of the constraint applies to unsuffixed words, which have an empty AGR-MARK list.<sup>11</sup>

Suffixed words with a non-empty AGR-MARK value participate normally in syntactic combinations, with all possible grammatical functions (head, specifier, adjunct, and so on). The presence of the agreement affix has no effect on the syntactic properties of the host. As mentioned already, a suffix does influence the linearization potential of its host, because the specification [+CL-HOST] requires the suffixed word to end up in a surface position that is compatible with this feature. The exact location of the suffixed word within a phrase cannot be specified: it can be the first word, the last word, or somewhere in the middle. But in all cases, information recording the presence of the affix must be projected. This means that the value of AGR-MARK must be amalgamated and propagated from all daughters in every phrasal combination. This formal mechanism is presented at the end of the next section in (27).

#### 4.2.2 AGR-TRIG and *l*-forms

The agreement marking is required by the *l*-form verb. We encode this by introducing another feature AGR(ELEMENT)-TRIG(GER), which (like AGR-MARK) takes a list of *index* objects as its value. Elements on AGR-TRIG must be discharged by the realization of the corresponding agreement suffix. The value of AGR-TRIG on *l*-forms is determined by the constraint in (26), which also relies on the function  $\mathbf{F}_{12}$  defined in (25):

<sup>10</sup>The function  $\mathbf{F}_{agr}$  can be defined for 3rd person indices (simply returning the original, unsuffixed form of the word), but this is not technically necessary for this constraint.

<sup>11</sup>This is obviously a simplified formulation that ignores other morphophonological processes in Polish that might cause the PHON value to be different from the MORPH | FORM value.

$$(26) \left[ \begin{array}{l} \text{word} \\ \text{HEAD} \mid \text{VFORM} \quad l\text{-form} \end{array} \right] \Rightarrow \left[ \begin{array}{l} \text{ARG-ST} \quad \langle \text{NP}_{\boxed{1}}, \dots \rangle \\ \text{AGR-TRIG} \quad \mathbf{F}_{I2}(\boxed{1}) \end{array} \right]$$

The result of constraint (26) is that *l*-forms with a 1st or 2nd person subject put their subject's index on their AGR-TRIG list (and thus trigger the presence of the corresponding agreement suffix), while 3rd person *l*-forms have an empty AGR-TRIG value. For the moment let us assume that all words in Polish other than *l*-forms have an empty AGR-TRIG list (although we will see some possible exceptions to this in §4.3).

The AGR-TRIG value propagates along the head projection of the *l*-form. The combined constraint that determines the values of AGR-TRIG and AGR-MARK in phrasal combinations is defined as follows:

$$(27) \left[ \begin{array}{l} \text{phrase} \\ \text{HD-DTR} \mid \text{SS} \quad \left[ \begin{array}{l} \text{AGR-MARK} \quad \boxed{1} \\ \text{AGR-TRIG} \quad \boxed{0} \end{array} \right] \\ \text{NON-HD-DTRS} \quad \langle \left[ \text{SS} \mid \text{AGR-MARK} \quad \boxed{2} \right], \dots, \left[ \text{SS} \mid \text{AGR-MARK} \quad \boxed{n} \right] \rangle \end{array} \right] \\ \Rightarrow \left[ \begin{array}{l} \text{AGR-MARK} \quad \boxed{1} \oplus \boxed{2} \oplus \dots \oplus \boxed{n} \\ \text{AGR-TRIG} \quad \boxed{0} \end{array} \right]$$

#### 4.2.3 Interaction of AGR-TRIG and AGR-MARK

Given constraint (27), the values of the two agreement features will propagate all the way to the maximal clausal projection of the *l*-form. At this point, matching AGR-TRIG and AGR-MARK specifications (which may have originated in very different parts of the clause) are brought together and discharged. This is achieved by means of the following non-branching ID schema (a *hd-only-ph* in the system of Ginzburg and Sag (2001), whose notation we adopt):<sup>12</sup>

$$(28) \left[ \begin{array}{l} \text{phrase} \\ \text{HEAD} \mid \text{TENSE} \quad \text{past} \\ \text{AGR-MARK} \quad \langle \rangle \\ \text{AGR-TRIG} \quad \langle \rangle \end{array} \right] \rightarrow \mathbf{H} \left[ \begin{array}{l} \text{HEAD} \mid \text{VFORM} \quad l\text{-form} \\ \text{AGR-MARK} \quad \mathbf{F}_{I2}(\boxed{1}) \\ \text{AGR-TRIG} \quad \mathbf{F}_{I2}(\boxed{1}) \end{array} \right]$$

The mutual discharging of AGR-MARK and AGR-TRIG results in the introduction of past tense (represented here simply as a head feature). The *l*-form itself must be

<sup>12</sup>This representation of the rule is simplified in two ways. First, an additional (most likely semantic) specification is needed to prevent iteration of the rule, because it can be triggered by “matching” empty lists for the 3rd person. Second, the rule can potentially apply at different points in the *l*-form projection, introducing a degree of spurious structural ambiguity. In general, non-branching rules should apply as “late” or as “high” as possible, but this cannot be simply encoded in terms of saturation of VAL and SLASH, given the possibility of coordinating non-maximal verbal projections (with distinct tenses).

lexically underspecified for tense, given the variety of its uses in Polish; it could perhaps be specified as  $\neg present$ . This non-branching schema stops the propagation of AGR-MARK and AGR-TRIG. Note that the resulting phrase is not subject to the constraint in (27), which is formulated only for branching phrases.

Let us work through some examples to illustrate our proposals. Polish allows subject drop, so a complete past tense clause can consist of a single word (a suffixed *l*-form), as in the following analysis:

$$(29) \begin{array}{c} \left[ \begin{array}{l|l} \text{HEAD} & \text{TENSE } \textit{past} \\ \text{AGR-MARK} & \langle \rangle \\ \text{AGR-TRIG} & \langle \rangle \end{array} \right] \\ | \\ \left[ \begin{array}{l|l} \text{AGR-MARK} & \langle \underline{\text{I}} \rangle \\ \text{AGR-TRIG} & \langle \underline{\text{I}} \rangle \end{array} \right] \\ \textit{przyszede\u0144em} \\ \textit{come-1sg} \\ \textit{I came} \end{array}$$

In this case, the first disjunct of (24) applies, so the *l*-form has a non-empty AGR-MARK list. Recall that the phonological function  $\mathbf{F}_{agr}$  has access to the HEAD value, and so the specific properties of *l*-form suffixation can be handled correctly, taking into account the phonological effects illustrated in (7). At the same time, constraint (26) requires the *l*-form also to have a non-empty AGR-TRIG value. The *l*-form thus satisfies the conditions for schema (28).

Fig. 1 is the analysis of the floating suffix example in (23b). It shows how the amalgamation and propagation mechanisms defined in (27) unite the corresponding AGR-MARK and AGR-TRIG specifications, which are introduced quite far apart from one another structurally, and trigger the application of the past tense schema.

Two final constraints need to be defined to complete the analysis. The agreement marking cannot appear to the right of the verb that selects it (13), (20). To block such structures, we formulate the following linear precedence rule:

$$(30) \left[ \text{SS} | \text{AGR-MARK} \langle \underline{\text{I}} \rangle \right] < \overset{\text{HD-DTR}}{\left[ \text{SS} | \text{AGR-TRIG} \langle \underline{\text{I}} \rangle \right]}$$

And finally, at the clausal level, there can be no unlicensed agreement markings (AGR-MARK elements) and no unsatisfied agreement requirements (AGR-TRIG elements):<sup>13</sup>

$$(31) \textit{clause} \Rightarrow \left[ \text{SS} \left[ \begin{array}{l|l} \text{AGR-MARK} & \langle \rangle \\ \text{AGR-TRIG} & \langle \rangle \end{array} \right] \right]$$

<sup>13</sup>The type *clause* is meant to subsume independent root clauses and embedded CPs.

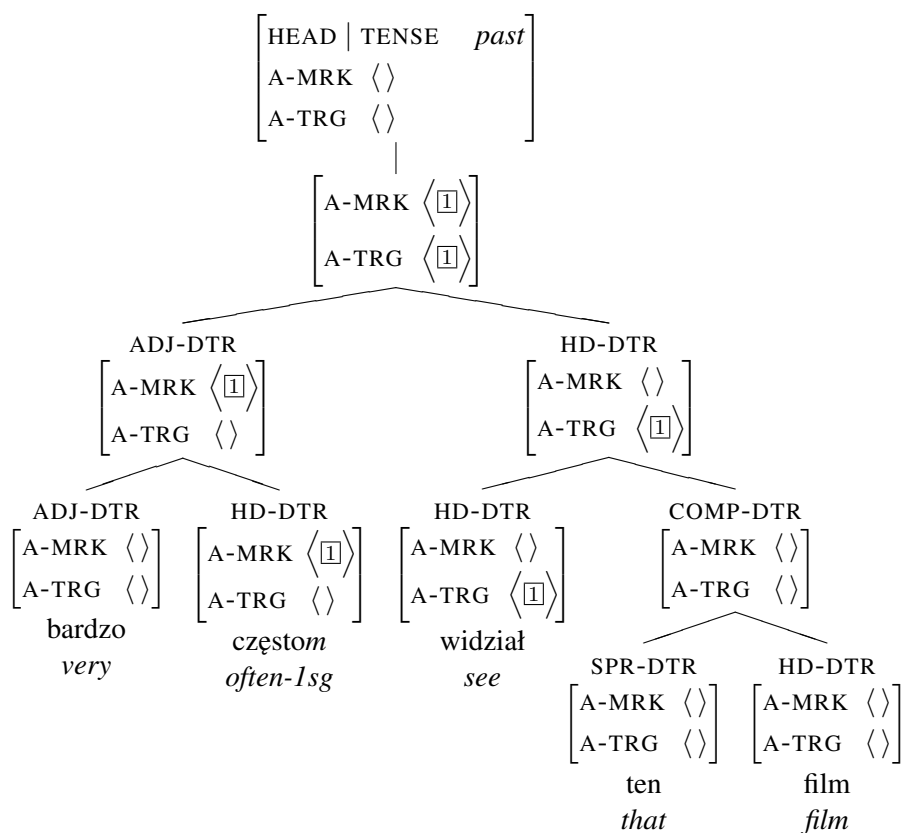


Figure 1: AGR-MARK and AGR-TRIG: Analysis of example (23b)

This constraint accounts for the clause-boundedness of past tense agreement. It excludes the ungrammatical examples in (12), where the required agreement marking is realized outside of the clause headed by the *l*-form. It also blocks the appearance of agreement markings inside a clausal dependent (i.e., sentential complement) of the *l*-form:

- (32) a. *Wszyscy chcieliśmy, [żeby zobaczyć słonia].*  
 all want-*1pl* COMP see.*inf* elephant  
 ‘We all wanted to see an elephant.’
- b. \**Wszyscy chcieli, [żebyśmy zobaczyć słonia].*  
 all want COMP-*1pl* see.*inf* elephant
- c. \**Wszyscy chcieli, [żeby zobaczyć słoniaśmy].*  
 all want COMP see.*inf* elephant-*1pl*

The ungrammatical examples above already fall under the descriptive generalization that agreement suffixes cannot appear in the sentence field to the right of the triggering *l*-form. But the treatment assumed above in §4.1.1—that all words in

this field are simply [−CL-HOST]—cannot apply here, because the sentential complement *can* contain [+CL-HOST] words. But the clitics (and suffixes) that appear in this domain must originate in the embedded clause; the constraint in (31) formalizes this restriction.

### 4.3 Other *l*-form constructions

Let us summarize the analysis just proposed. The *l*-form of the verb introduces an AGR-TRIG specification, and personal agreement markings introduce an AGR-MARK specification. These values propagate to the clause level, but a well-formed clause must have empty values for both features. So an *l*-form must co-occur with the corresponding agreement marking, allowing both features to be discharged, and giving rise to a past tense structure.

But the *l*-form behaves very differently in the other constructions where it appears: the conditional, the future, and with inflected complementizers. As we have seen, these constructions have quite divergent properties, but all three involve the “bare”, unsuffixed *l*-form. None of the dependents of the *l*-form can carry a “floating” agreement marking, either. This is illustrated for the future below:

- (33) a. Ty *będziesz* go widział.  
           you FUT.2sg him see  
           ‘You will see him.’
- b. \*Ty *będziesz* go widziałeś.  
           you FUT.2sg him see-2sg
- c. \*Ty *będziesz* goś widział.  
           you FUT.2sg him-2sg see

At first sight, it seems that the grammatical sentence in (33a) should violate the constraint on clauses in (31): the *l*-form introduces an AGR-TRIG element, but there is no agreement suffix in the clause to discharge it. One possible (but undesirable) solution would be to assume that the future (and the other constructions considered in this section) involve a different *l*-form from the past tense, one that is not subject to the AGR-TRIG constraint in (26).

Actually, we can avoid this move because our analysis already accommodates sentence (33a). According to (27), AGR-TRIG is shared between a phrase and its head daughter, and in this example, the head daughter is the future auxiliary, not the *l*-form. So the *l*-form’s AGR-TRIG value is not propagated to the clause level, and nothing requires it to be discharged.

#### 4.3.1 Auxiliaries

Our treatment of the future and conditional auxiliaries is quite straightforward. They simply require their *l*-form complement to have an empty AGR-MARK list, as in the following description (to be unified with the description of conditional *by* in (18), for example):

$$(34) \left[ \begin{array}{l} \text{word} \\ \text{HEAD} \mid \text{AUX} \quad + \\ \text{ARG-ST} \left\langle \text{NP}, \left[ \begin{array}{l} \text{HEAD} \left[ \begin{array}{l} \text{VFORM} \quad \textit{l-form} \\ \text{TENSE} \quad \neg \textit{past} \end{array} \right] \right] \right\rangle \\ \text{AGR-MARK} \quad \langle \rangle \end{array} \right] \end{array} \right]$$

In combination with the amalgamation of AGR-MARK formulated in (27), the empty list specification ensures that no agreement suffixes appear anywhere in the *l*-form complement. This is only true, however, if the past tense schema in (28) has not applied, with the effect of discharging the agreement lists. This possibility is excluded by the additional specification [TENSE  $\neg \textit{past}$ ], which ensures that the auxiliary sees the “initial” values of AGR-MARK and AGR-TRIG.

Something should be said about the values of AGR-MARK and AGR-TRIG on the auxiliaries themselves. The simplest solution is to assume empty lists, and this is a completely unproblematic analysis for the future auxiliary. For conditional *by*, the situation may be more complex, because the endings it takes are exactly the same as the past tense agreement suffixes, suggesting that the same function  $\mathbf{F}_{agr}$  as in (24) may be involved.<sup>14</sup> In that case, the forms of *by* could have a non-empty AGR-MARK specification, which would then require a non-empty AGR-TRIG specification (inherited from the *l*-form complement). A head-only schema analogous to (28) would also have to be defined for the conditional. We will not pursue this discussion any further here, but we would like to point out that conditional *by* seems to be slowly losing its auxiliary status in the same way as the former past tense auxiliary, and the indeterminacy in its analysis can be explained as a reflection of this transitional status.

#### 4.3.2 Inflected complementizers

Given the classical HPSG treatment of complementizers as marker daughters (i.e., non-heads), the data involving inflected subjunctive complementizers in (15) could be handled exactly like the past tense, in terms of AGR-MARK/AGR-TRIG interaction, with the additional constraint that the agreement suffix must appear on the complementizer. But we follow a more recent trend in HPSG (Ginzburg and Sag, 2001; Tseng, 2002) that treats complementizers as syntactic heads. This is also the approach adopted by Borsley (1999) for the Polish elements *żeby/aby/by*. His analysis, however, involves a very unusual argument inheritance mechanism (a completely flat structure in which the complementizer inherits the *l*-form’s subject and “demotes” it to a complement) that we find quite unmotivated.

It would be convenient for the complementizer to have access to the subject in this way, because it has to inflect to agree with it just like the auxiliaries do, but at

<sup>14</sup>The historical evidence also points in this direction, because the conditional forms used to have a completely idiosyncratic set of endings, which have been “regularized” in modern Polish.

the same time there is no evidence to suggest that the complementizer combines with anything other than a saturated sentence. In our analysis, we do not have to resort to argument manipulation, because the information that the complementizer needs is visible in the sentence's AGR-TRIG value:

$$(35) \left[ \begin{array}{l} \text{word} \\ \text{HEAD} \quad \text{comp} \\ \\ \text{COMPS} \left\langle \text{S} \left[ \begin{array}{l} \text{HEAD} \quad \left[ \begin{array}{l} \text{VFORM} \quad \textit{l-form} \\ \text{TENSE} \quad \neg \textit{past} \end{array} \right] \\ \text{AGR-TRIG} \quad \mathbf{F}_{12} (\textit{index}) \\ \text{AGR-MARK} \quad \langle \rangle \end{array} \right] \right\rangle \end{array} \right]$$

Again, as in the auxiliary description in (34), the non-past specification ensures that the agreement features have not been discharged by rule (28). Consequently, if the *l*-form has a 1st or 2nd person subject, its index will still be on the complement's AGR-TRIG list, and the complementizer can take the appropriate person and number inflection. If the subject is 3rd person, AGR-TRIG is empty (thanks to  $\mathbf{F}_{12}$ ) and in this case there is only one form, *żeby/aby/by*, for both singular and plural.

## 5 Conclusion

We have developed analyses for all uses of the *l*-form in Polish (past tense, conditional, future, and inflected complementizer constructions), taking into account their very distinct grammatical properties. In contrast to many previous approaches, we have not tried to offer a uniform picture, although many analytical building blocks are shared across the analyses. Taking a global view of the phenomena we have examined, at one extreme we have the future tense, which is an ordinary auxiliary verb construction, and at the other extreme the past tense, where the elements that were historically auxiliaries are now simply agreement markings. The conditional is in a transitional state between an auxiliary construction and a simple verb construction. We have presented an auxiliary analysis here, but various aspects of the construction are open to reanalysis. Finally, the inflected complementizers are unusual elements, but they are nevertheless handled straightforwardly in our framework.

## References

- Abeillé, Anne and Godard, Danièle. 2002. The syntactic structure of French auxiliaries. *Language* 78, 404–452.
- Andersen, Henning. 1987. From auxiliary to desinence. In Martin Harris and Paolo Ramat (eds.), *Historical Development of Auxiliaries*, pages 21–52, Berlin: Mouton de Gruyter.

- Bański, Piotr. 2000. *Morphological and Prosodic Analysis of Auxiliary Clitics in Polish and English*. Ph. D.thesis, Uniwersytet Warszawski, Warsaw.
- Bender, Emily M. and Sag, Ivan A. 2001. Incorporating Contracted Auxiliaries in English. In Ronnie Cann, Claire Grover and Philip Miller (eds.), *Grammatical Interfaces in HPSG*, pages 17–32, Stanford, CA: CSLI Publications.
- Borsley, Robert D. 1999. Auxiliaries, Verbs and Complementizers in Polish. In Robert D. Borsley and Adam Przepiórkowski (eds.), *Slavic in Head-Driven Phrase Structure Grammar*, pages 29–59, Stanford, CA: CSLI Publications.
- Borsley, Robert D. and Rivero, María Luisa. 1994. Clitic Auxiliaries and Incorporation in Polish. *Natural Language and Linguistic Theory* 12, 373–422.
- Dłuska, Maria. 1974. *Prozodia Języka Polskiego*. Warsaw: PWN.
- Ginzburg, Jonathan and Sag, Ivan A. 2001. *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. Stanford, CA: CSLI Publications.
- Kupść, Anna. 2000. *An HPSG Grammar of Polish Clitics*. Ph. D.thesis, Université Paris 7 and Polish Academy of Sciences.
- Mikoś, M. and Moravcsik, Edith. 1986. Moving Clitics in Polish and Some Cross Linguistic Generalizations. *Studia Slavica* pages 327–336.
- Miller, Philip H. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York: Garland.
- Rappaport, Gilbert C. 1988. On the Relationship between Prosodic and Syntactic Properties of Pronouns in the Slavic Languages. In Alexander M. Schenker (ed.), *American Contribution to the Tenth International Congress of Slavists*, pages 301–327, Columbus, OH: Slavica.
- Reape, Mike. 1994. Domain Union and Word Order Variation in German. In John Nerbonne, Klaus Netter and Carl Pollard (eds.), *German in HPSG*, pages 151–197, Stanford, CA: CSLI Publications.
- Rittel, Teodozja. 1975. *Szyk wyrazów w obrębie form czasu przeszłego i trybu przypuszczającego*. Wrocław: Ossolineum.
- Swan, Oscar E. 2002. *A Grammar of Contemporary Polish*. Bloomington: Slavica.
- Tseng, Jesse. 2002. Remarks on Marking. In Frank van Eynde, Lars Hellan and Dorothee Beermann (eds.), *Proceedings of the 8th International HPSG Conference*, pages 267–283, Stanford, CA: CSLI Publications.
- Witkoś, Jacek. 1997. Polish Inflectional Auxiliaries Revisited, paper presented at the 30th Poznań Linguistic Meeting, Adam Mickiewicz University, Poznań, Poland, May 1–3, 1997.