Syncretism in German: a unified approach to underspecification, indeterminacy, and likeness of case

Berthold Crysmann

Saarland University & German Research Centre for Artificial Intelligence (DFKI)

Proceedings of the HPSG05 Conference

Department of Informatics, University of Lisbon

Stefan Müller (Editor)

2005

CSLI Publications

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

Abstract

In this paper I address the phenomenon of syncretism in German and show how Flickinger (2000)'s approach to related issues in English can be adapted to provide a compact, disjunction-free representation of German nominal paradigms by means of combined case/number/gender type hierarchies. In particular, I will discuss the issue of case identity constraints in German coordinate structures, which has so far prevented successful application of Flickinger's proposal to German, and show how likeness constraints targetting individual inflectional dimensions of a combined type hierarchy can be expressed by means of typed lists that abstract out the relevant dimension.

I further show that current type-based approaches to feature neutrality are unable to combine the treatment of this phenomenon with the virtues of underspecification. I will then propose a revised organisation of the inflectional type hierarchies suggested by Daniels (2001), drawing on a systematic distinction between inherent and external (case) requirements.

1 Introduction

Nouns, adjectives and determiners in German inflect for case, number and gender. However, as is typical for inflectional languages, these morphosyntactic feature dimensions are not expressed by discrete, individually identifiable affixes. Rather, affixes realise complex feature combinations. Although four case, three gender and two number specifications can clearly be distinguished, the morphological paradigms of the language are characterised by heavy syncretism.[1]

(1)

	Singular	Plural
Nom	der alte Computer	die alten Computer
Gen	des alten Computers	der alten Computer
Dat	dem alten Computer	den alten Computern
Acc	den alten Computer	die alten Computer

As illustrated by the paradigm in (1), German inflected nouns and adjectives are highly ambiguous at the word level. At the phrase level, however, ambiguity is somewhat reduced owing to the fact that, first, German NPs are subject to agreement in case, number, and gender, and second, determiners, adjectives and nouns are subject to different patterns of ambiguity.

[†]I would like to thank Stefan Müller and Michael Jellinghaus for fruitful discussion of several aspects of this work. I am also indebted to the audiences at the HPSG 2005 and FG-MoL 2005 conferences for comments on and discussion of the ideas presented here, in particular Carl Pollard and Shuîchi Yatabe. A great many thanks also to the anonymous reviewers for their invaluable comments.

The work presented in this article was partially supported by research grants from the German Federal Ministry of Education, Science, Research and Technology (BMBF) to the DFKI project Quetal (FKZ 01 IW C02).

[1]See the Surrey Morphology Group syncretism database for a cross-linguistic overview (http://www.surrey.ac.uk/LIS/SMG/).

Often, syncretism cannot be resolved to disjunctive specification or underspecification within a single feature, but it cuts across the three inflectional dimensions: in our example above, the base noun *Computer*, can express either nominative, dative, and accusative singular, or nominative, genitive, and accusative plural. Likewise, the adjectival form *alten* can fill any cell in the (weak masculine) paradigm, except nominative singular. Although in principle, it is possible to provide a compact description of the set of readings in terms of nested disjunctions, one is actually forced to make an arbitrary decision as to which of the dimensions one wants to encode as the outer or inner disjunction (cf. (2) and (3))

(2) $$\begin{bmatrix} \textsc{case} & \mathit{nom} \vee \mathit{dat} \vee \mathit{acc} \\ \textsc{num} & \mathit{sg} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{nom} \vee \mathit{dat} \vee \mathit{acc} \\ \textsc{num} & \mathit{pl} \end{bmatrix}$$

(3) $$\begin{bmatrix} \textsc{case} & \mathit{nom} \vee \mathit{acc} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{dat} \\ \textsc{num} & \mathit{pl} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{gen} \\ \textsc{num} & \mathit{sg} \end{bmatrix}$$

A possible way to circumvent this problem is to revert to disjunctive normal form, as in (4): as a result, however, one will lose the generalisation that all six paradigm cells are actually expressed by one and the same form.

(4) $$\begin{bmatrix} \textsc{case} & \mathit{nom} \\ \textsc{num} & \mathit{sg} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{dat} \\ \textsc{num} & \mathit{sg} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{acc} \\ \textsc{num} & \mathit{sg} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{nom} \\ \textsc{num} & \mathit{pl} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{gen} \\ \textsc{num} & \mathit{pl} \end{bmatrix} \vee \begin{bmatrix} \textsc{case} & \mathit{acc} \\ \textsc{num} & \mathit{pl} \end{bmatrix}$$

This is even more unsatisfactory, if the linguistic expression under consideration is actually the unmarked citation form, as in the case of *Computer*.

Yet, typed feature formalisms, as argued in Flickinger (2000), offer an alternative to the use of disjunction, both within a dimension and across dimensions, namely type underspecification. Flickinger (2000) suggests to combine the inflectional dimensions of number and person in English to arrive at a compact representation of third singular and non-third singular agreement without the use of negation or disjunction. The key is to combine all the inflectional dimensions involved in syncretism into a single over-arching hierarchy.

Currently, one of the major obstacles for applying this strategy to the case of German is the kind of likeness constraints operative, e.g., in coordinating constructions, where agreement between conjuncts targets only a single inflectional dimension, namely case, to the exclusion of gender and number. I will show, in the first part of this paper, how list types can be fruitfully put to use to abstract out individual featural dimensions from combined case/number/gender type hierarchies, permitting the expression of likeness constraints, or type identity in coordinate structures.

Ambiguous nominal forms in German are also subject to indeterminacy or feature neutrality. Again, indeterminacy is not restricted to individual inflectional

dimensions, but rather follows the patterns of syncretism. Although the notions of ambiguity and indeterminacy are intimately related, there is currently no analysis at hand that is capable of combining the machinery necessary to cover feature indeterminacy with the benefits of underspecification.

In the second part of this paper, I will propose an entirely type-based approach to syncretism that will successfully reconcile Daniels (2001)'s approach to feature indeterminacy with morphosyntactic underspecification across features. As a result, the current proposal presents an entirely disjunction-free approach to syncretism, addressing indeterminacy, underspecification and likeness constraints.

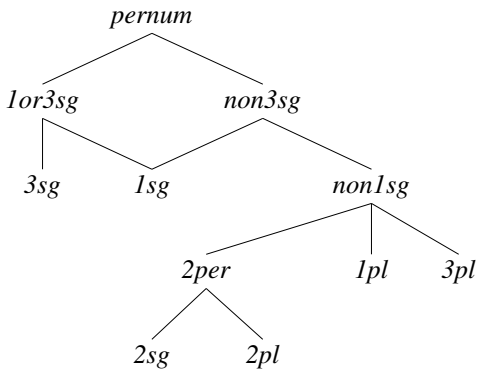
2 Underspecification

In the context of grammar implementation, Flickinger (2000) compares disjunctive and type-based disjunction-free approaches to English verb agreement, in particular non-third singular agreement. Here, the problem is entirely parallel to German case/number/gender syncretism: like with German adjectives as *alten* in table (1), bare simple present forms in English can express every person/number specification except one: third singular. In order to provide a compact description of this unmarked form, one needs nested disjunctions as in (5), if person and number dimensions are to be represented by distinct features.

(5)
$$\left[\begin{array}{ll} \textit{non-3rd-sg-verb} & \\ \text{AGR} & \left[\begin{array}{ll}\text{NUM} & \textit{sg}\\ \text{PER} & \textit{1} \vee \textit{2}\end{array}\right] \vee \left[\text{NUM} \quad \textit{pl}\right] \end{array}\right]$$

Flickinger then showed that by exploiting types it is possible to keep with a very concise representation of non-third singular verbs while at the same time eliminate all disjunctions: the key idea here is to give up the idea of having person and number dimensions of a paradigm represented as individual features, but instead represent these dimensions as part of a single type hierarchy (see (6)), the nodes of which corresponds to linguistically interesting groups of cells in an inflectional paradigm.

(6)



Combined with the fact that complementation within finite sets can always be captured by introducing appropriate supertypes into the hierarchy which subsume the relevant partition of (leaf) types (*non3sg* = *pernum* $\wedge \neg$ *3sg*), the natural class of non-third singular verbs can be described most compactly, without any need for descriptive devices such as disjunction or negation.

(7)
$$\begin{bmatrix} \textit{non-3rd-sg-verb} \\ \text{AGR} \quad \left[\text{PERNUM } \textit{non3sg}\right] \end{bmatrix}$$

Most interestingly, the compactness of linguistic description achieved by the elimination of disjunctive features also pays off very well in terms of processing efficiency: comparing the performance of the disjunctive and the type-based approach, Flickinger (2000) shows that the latter outperforms the former by a factor of 3–4, with an otherwise unchanged grammar (the LinGO English Resource Grammar; ERG Copestake and Flickinger, 2000) running on the same processing platform (PAGE; Uszkoreit et al., 1994).

In the context of more strongly inflecting languages, such as German, where syncretism is the norm rather than the exception, underspecification of inflectional features across different dimensions is probably even more pressing: recall that a typical noun such as *Computer* can express any case/number combination, except genitive singular and dative plural, i.e. 6 in total. Using combined case/number/gender hierarchies, the syncretism between nominative/dative/accusative singular and nominative/genitive/accusative plural can be represented compactly as one entry. The very same holds for German determiners and adjectives: if only disjunctions within a single dimension are eliminated by means of type abstraction, we can still find a residual local ambiguity within each NP, of typically two readings per determiner, adjective, or noun. With disjunctive normal form, local ambiguity would be much higher indeed. Using a combined hierarchy of case/number/gender specifications, local ambiguity can be brought down to 1. Furthermore, such a move will avoid the motivational pitfalls of arbitrary decisions as to the relative nesting of disjunctions.

3 Likeness constraints in coordination

It has been argued by Müller (p.c.) that one of the main obstacles for exploiting combined case-number-gender hierarchies to provide an entirely disjunction-free representation of German syncretism surfaces in certain coordinate structures. It is a well-known fact about German that likeness of category in coordinate structures includes likeness of case specification, but excludes, as a rule, requirements concerning the likeness of gender or number specifications in the conjuncts, a pattern which is quite neatly predicted by HPSG's segregation of HEAD features and INDEX features. However, in free word order languages like German, case arguably serves not only a categorial function, but also a semantic one, thereby supporting

the originally morphological motivation towards organising all agreement features into a single hierarchy (see also Kathol (1999) for a similar proposal). Moreover, the mere existence of indeterminacy across case and index features makes combined hierarchies almost inevitable.

Müller discusses syncretive pronominals in German, such as *der*, which is ambiguous, inter alia, between nominative singular masculine, as shown in (8), and dative singular feminine, as illustrated in (9).

(8) Der schläft.
the.N.S.M sleeps
'That one sleeps.'

(9) Ich helfe der.
I help the.D.S.F
'I help that one.'

This ambiguity could be represented by a type *n-s-m+d-s-f*.[2] Subcategorisation for nominative singular (type *n-s-g*) or dative (type *d-n-g*) will disambiguate these forms accordingly.[3]

In coordinate structures, however, we observe that likeness of case equally eliminates one of the possible gender specifications for *der*, as witnessed by the disambiguation (10). Thus, we must be able to distribute the case requirement over the two conjuncts in such a way that it can exert its disambiguatory potential, without actually unifying the entire case/number/gender specifications of the two conjuncts.

(10) Ich helfe der und dem Mann.
I help the.D.S.F and the.D.S.M man
'I help this one and the man.'

In Daniels (2001), this problem was partly anticipated: he suggests to address the issue of likeness of case by means of a relational constraint same-case/2, which restricts the two arguments to satify identical type requirements. This type equality is essentially imposed by disjunctive enumeration of the four possible subcategorised case values. In typed feature formalisms without relational constraints, his solution may be mimicked by means of unfolding the relevant phrase structure schemata into case-specified variants. In both cases, a greater part of the efficiency gains achieved by underspecification may get eaten up by this disjunctive approach to case similarity.

[2] As a convention, I am using the following nomenclature of combined c(ase)-n(umber)-g(ender) types: the three inflectional dimensions are specified in the above order, separated by a hyphen. In the first slot, *c* represents the most general case "value", *n,g,d,a* the most specific. "Disjunctive values" are represented as combinations of case specifications. The very same holds number and gender specifications.

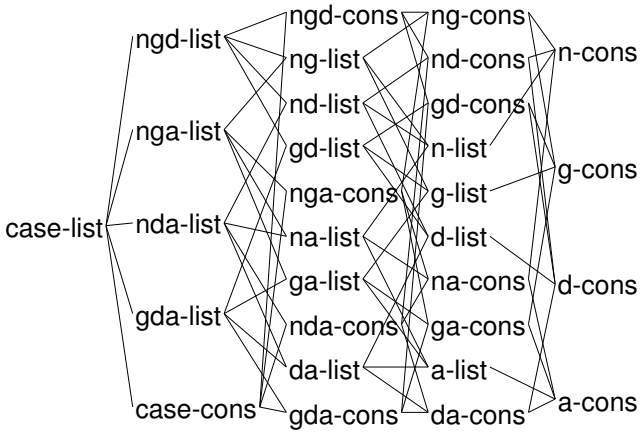
[3] For ease of exposition, I am abstracting away from the internal/external distinction, which is immaterial here, since we are only dealing with underspecification, not indeterminacy.

An alternative, though not fully satisfactory solutiuon would involve retaining a HEAD feature CASE along-side the combined AGR feature. While this move will be at least effective in ruling out unacceptable surface strings, it will fail to impose the disambiguation potential of the subcategorising head onto the individual conjuncts.

What is really needed here is a data structure that may serve to both express the appropriate case-requirements in terms of a combined hierarchy, and permit arbitrarily many specific instantiations of the case constraint. Fortunately, typed feature formalisms do provide for such a data structure, namely typed lists.

To start with, we will set up a hierarchy of case list types, as depicted in figure (11)[4], where each list type immediately subsumes at least one subtype representing a non-empty list of the same case type.

(11)



Types in the combined case-number-gender hierarchy will now restrict their CASE value to an appropriate list type, as given in (12).[5]

(12) *nda-n-g* $\rightarrow \left[\text{CASE} \quad \textit{nda-list}\right]$

Non-empty case lists bear a type constraint restricting the FIRST value to the corresponding agreement type in the combined case/number/gender hierarchy. Actually, thanks to type inference in the hierarchy of case lists, we only need to do this for the 4 immediate subtypes of *case-cons*, namely *ngd-cons*, *nga-cons*, *nda-cons*, and *gda-cons*. In order to propagate the case specification onto all elements of the open list, the tail is constrained to the corresponding list type (see (13)).

(13) *nda-cons* $\rightarrow \left\langle \textit{nda-n-g} \mid \textit{nda-list} \right\rangle$

Now that we have a data structure that enables us to encode likeness of case for arbitrary instances of case/number/gender types, all we need to do is refine

[4]The type hierarchy has been exported from the LKB: supertypes are on the left, subtypes are on the right.

[5]Recall that, according to our naming convention,the type *nda-n-g* represents all case specification except genitive. Number and gender are fully underspecified.

our existing coordination schemata to distribute the case restriction imposed on the coordinate structure onto the individual conjuncts. In the implemented German grammar we are using, coordinate structures are licensed by binary phrase structure schemata. Thus, all we have to do is to constrain the AGR feature of the left conjunct daughter to be token-identical to the first element on the mother's AGR|CASE list, and percolate the rest of this list onto the (recursive) righthand conjunct daughter's AGR|CASE value:

(14) *coord-phr* →
$$\begin{bmatrix} \text{SS}\,|\,\text{L}\,|\,\text{AGR}\,|\,\text{CASE}\left\langle \boxed{1}\,|\,\boxed{2}\right\rangle \\ \text{COORD-DTRS}\left\langle \left[\text{SS}\,|\,\text{L}\,|\,\text{AGR}\;\boxed{1}\right], \\ \qquad\qquad\quad \left[\text{SS}\,|\,\text{L}\,|\,\text{AGR}\,|\,\text{CASE}\;\boxed{2}\right]\right\rangle \end{bmatrix}$$

Coordinating conjunctions, which combine with a conjunct by way of a head-complement rule, will equate their own AGR|CASE|FIRST value with the AGR value of their complement, percolating the case constraint onto the last conjunct.

(15)
$$\begin{bmatrix} \text{SS}\,|\,\text{L}\left[\text{AGR}\,|\,\text{CASE}\left\langle \boxed{1}\,|\,\textit{list}\right\rangle\right] \\ \text{VAL}\,|\,\text{COMPS}\left\langle \left[\text{L}\,|\,\text{AGR}\;\boxed{1}\right]\right\rangle \end{bmatrix}$$

Besides coordination, the current approach to likeness constraints across syncretive forms can also be applied to case/gender agreement in German constructions involving the phrase *ein- nach d- anderen* 'one after the other', a set of phenomena discussed by Höhle (1983) and Müller (1999, 2001):

(16) Wir$_i$ helfen ihnen$_j$ [einem nach dem anderen]$_{*i/j}$
we.NOM help them.dat one.DAT.M after the.M other
'We help them one after the other.'

(17) Wir$_i$ helfen ihnen$_j$ [einer nach der anderen]$_{*i/j}$
we.NOM help them.dat one.DAT.F after the.F other
'We help them one after the other.'

(18) Wir$_i$ helfen ihnen$_j$ [einer nach dem anderen]$_{i/*j}$
we.NOM help them.dat one.NOM.M after the.M other
'We help them one after the other.'

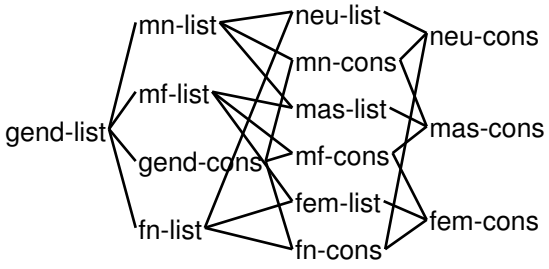
(19) Wir$_i$ helfen ihnen$_j$ [eine nach der anderen]$_{i/*j}$
we.NOM help them.DAT one.NOM.F after the.F other
'We help them one after the other.'

As illustrated by the data in (16–19) above, agreement between antecedent and the phrase *ein- nach d- anderen* 'one after the other' proceeds along two inflectional dimensions: case and gender.Within the phrase *ein- nach d- ander-*, we find

gender agreement between the two pronominal *ein-* and the NP *d- anderen*. Case of the latter is invariantly dative, since it is governed by the preposition *nach*. The important aspect of this construction now is that the gender agreement between the pronominals partially disambiguates the case specification: e.g., the pronominal *einer* displays syncretism between nominative masculine and dative feminine (singular). As witnessed by the contrasts in (17) and (18), disambiguation of case syncretism by means of grammatical gender reduces the semantic attachment potential of the entire phrase, precluding attachment to the subject in (17), and to the object in (18).

The situation we encounter here is actually highly parallel to the one we found earlier with likeness of case in coordinate structures: again, agreement only targets a subset of the inflectional dimensions (case and gender) to the exclusion of others (person and number). What is therefore needed, is , again, a mechnism to abstract out the relevant dimensions from our syncretism types. While we can directly reuse our list-valued CASE feature to implement case agreement, we have to provide an analoguous abstraction of the gender dimension, a step, which is very much straightforward:

(20)



(21) $\textit{c-n-mn} \rightarrow \left[\text{GEND} \quad \textit{mn-list}\right]$

(22) $\textit{mn-cons} \rightarrow \left\langle \textit{c-n-mn} \mid \textit{mn-list} \right\rangle$

Again, we need a hierarchy of list types, and connect it — via type constraints — to appropriate types in the combined *c-n-g* hierarchy.

Having established the required abstraction of gender alongside case, we are now in a position to capture the interaction of case and gender agreement. All it needs, is to require that, in the phrase *ein- nach d- anderen*, the PP *nach d- anderen*, which exhibits gender agreement with the pronoun *ein-*, will equate the first element of its GEND list with the AGR value of *ein-*, either constructionally, or via a selection feature, e.g. MOD.

As a result, the entire AGR value of *ein-* will be disambiguated to a *c-n-g* specification compatible with the PP's gender. The AGR value of the entire construction, which represents an aggregate of singular entities, will be the unification of a constructional plural specification (*c-p-g*) with the first elements on both CASE and

GEND of *ein-*. This AGR value will then be unified with that of the antecedent.[6]

(23)
$$\left[\begin{array}{l} \text{SS}\,|\,\text{L}\,|\,\text{AGR}\ \textit{c-n-p} \wedge \boxed{1} \wedge \boxed{2} \\ \text{DTRS}\left\langle \left[\begin{array}{l} \text{PH}\left\langle \text{einer} \right\rangle \\ \text{SS}\,|\,\text{L}\,|\,\text{AGR}\ \textit{n-s-mn+d-s-f} \wedge \boxed{3}\left[\begin{array}{l}\text{CASE}\,|\,\text{FIRST}\ \boxed{1}\\ \text{GEND}\,|\,\text{FIRST}\ \boxed{2}\end{array}\right]\end{array}\right], \left[\begin{array}{l} \text{PH}\left\langle \text{nach der anderen} \right\rangle \\ \text{SS}\,|\,\text{L}\,|\,\text{AGR}\ \textit{d-s-f} \wedge \left[\begin{array}{l}\text{CASE}\ \textit{d-list}\\ \text{GEND}\left[\begin{array}{l}\textit{fem-cons}\\ \text{FIRST}\ \boxed{3}\textit{c-n-f}\end{array}\right]\end{array}\right]\end{array}\right] \right\rangle \end{array}\right]$$

To conclude, we have seen that the approach to likeness of case in coordinate structures can be extended, in a principled way, to other phenomena displaying partial agreement, i.e. agreement involving only a subset of inflectional dimensions. Furthermore, as illustrated by our analysis of the overlapping of gender and case agreement, the combination of dimensions in partial agreement can essentially be reduced to abstracting out each dimension individually and having them interact by means of unification.

Under a more general perspective, the technique employed here to abstract out certain dimensions from a more complex hierarchy by means of typed lists can be regarded as a sort of closed-world variant of type identity. As such, it certainly has an application potential which goes far beyond the concrete problems solved here.

4 Feature neutrality

It has been argued by Ingria (1990) that the phenomenon of feature neutrality in coordination constitutes a severe challenge for unification-based approaches to feature resolution and concludes that unification should rather be supplanted by feature compatibility checks.

(24) Er findet und hilft Frauen.
he finds.A and helps.D women.A/D
'He finds and helps women.'

(25) * Er findet und hilft Kindern.
he finds.A and helps.D children.D

(26) * Er findet und hilft Kinder.
he finds.A and helps.D children.A

[6] In order to make the lexical specification of case/number/gender information more transparent, I have left the unification of values in (23) unresolved.

Unification-based frameworks such as LFG or HPSG have taken up the challenge, refining the representation of feature constraints in such a way that neutrality can be modelled without any substantial changes to the underlying formalism. For HPSG, Daniels (2001) proposed to address these problems by means of enriching the type hierarchy to include neutral types, an idea originally due to Levine et al. (2001).[7]

Daniels (2001) has also discussed cases where the potential for feature indeterminacy does not only involve the values of a single feature: as illustrated in (27), a masculine noun like *Dozenten* can express any cell of the case/number paradigm except nominative singular. Accordingly, one and the same form can be subject to feature indeterminacy regarding number, gender, or even case.

(27) der Antrag des oder der Dozenten
the petition Def.G.Sg or Def.G.Pl lecturer.G/D/A+N.Pl
'the petition of the lecturer(s)'

(28) der oder die Abgeordnete
Def.N.M.Sg or Def.N.F.Sg representative.N.Sg.M/F
'the male or female representative'

(29) Er findet und hilft Dozenten.
he finds.A and helps.D lecturers.A/D
'He finds and helps lecturers.'

(30) a. mit jedem Mann oder Kind
with every.D.Sg.M/N man.M or child.N
'with every man or child'

b. * jeder Mann oder Kind
every.N.Sg.M man.N.Sg.M or child.N.Sg.N
'with every man and child'

c. * jedes Mann oder Kind
every.N.Sg.N man.N.Sg.M or child.N.Sg.N
'with every man and child'

A determiner like *der* is neutral between nominative singular masculine and genitive/dative plural. However, indeterminacy with respect to number is not independent of case, as illustrated by (31), where the unavailability of a nominative singular reading for *Dozenten* is responsible for the illformedness of the sentence.

(31) * der Dozenten ist hier
the.N.Sg.M+G/D.Sg.F+G.Pl lecturer.G/D/A+N.Pl is here

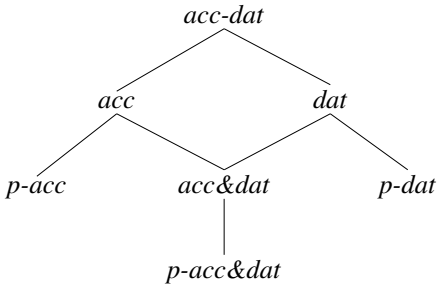
[7] Within LFG, a technically different, though conceptually similar approach has been developed by Dalrymple and Kaplan (2000). See Levy and Pollard (2001) for a comparison.

To incorporate the issue of neutrality across features, Daniels suggests to combine values of different inflectional features into an overarching type hierarchy, the nodes of which are essentially derived by building the Cartesian product of the types within each inflectional dimension.

4.1 The Problem

Although both feature indeterminacy and ambiguity do call for type hierarchies combining different inflectional dimensions, these two approaches have not yet received a unified treatment to date: it has been recognised as early as Zaenen and Karttunnen (1984) that in unification-based formalisms feature neutrality cannot be reduced to underspecification. The apparent incompatibility of neutrality and underspecification is even more surprising, as these two notions are intimately related: i.e., the ambiguity of a form between two values is a necessary prerequisite for this form to be embeddable in a neutral context.

(32)



Taking as starting point the case hierarchy proposed by Daniels (2001), one might be tempted to assign a case-ambiguous form like 'Frauen' a supertype of both *acc* and *dat*, e.g. *acc-dat*, which can be resolved to *p-acc* ('die Frauen') or *p-dat* ('den Frauen'), depending on context. However, to include feature-neutrality, it must also be possible to resolve it to the neutral type *acc&dat*. Suppose now that a form like *die* 'the' is itself ambiguous, i.e. between nominative and accusative, representable by a type *nom-acc*, again a supertype of *acc*. Unification of the case values of *die* 'the' and *Frauen* 'women' will yield *acc*, which will still be a supertype of the neutral type *acc&dat*, erroneously licensing the unambiguously non-dative *die Frauen* 'the women' in the neutral accusative/dative context of *findet und hilft* 'finds and helps'.

(33) * Er findet und hilft [die Frauen]
he finds.A and helps.D [the women].A

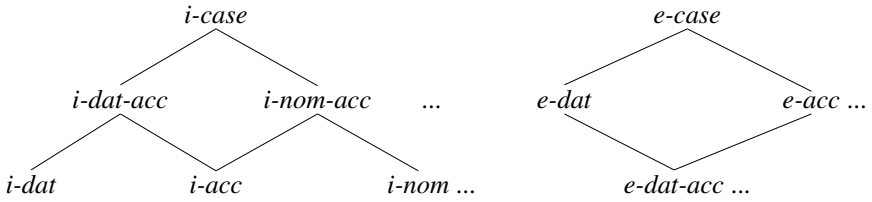
Thus, under Daniels's account, lexical items are explicitly assigned leaf type values, so-called "pure types". While successful at resolving the issue of indeterminacy, this approach in fact drastically increases the amount of lexical ambiguity, having to postulate distinct entries for type-resolved pure accusative, pure dative, pure nominative, pure genitive, as well as all pair-wise case-neutral variants of a

single form like *Frauen* 'women'. Ideally, all these different readings should be representable by a single lexical entry, if only underspecification could be made to work together with indeterminacy.

4.2 A Solution

The reason for the apparent incompatibility of underspecification and feature neutrality lies with the attempt to address both aspects within a single type hierarchy. Instead, I shall argue to draw a principled distinction between inherent inflectional feature values, where unification specialises from underspecified or ambiguous types to unambiguous types, and external or subcategorised feature values where unification proceeds from non-neutral, though generally unambiguous to neutral types. As a result we will have two partially independent hierarchies, one for ambiguity (*i-case*) and an inverse one for neutrality (*e-case*).[8]

(34)

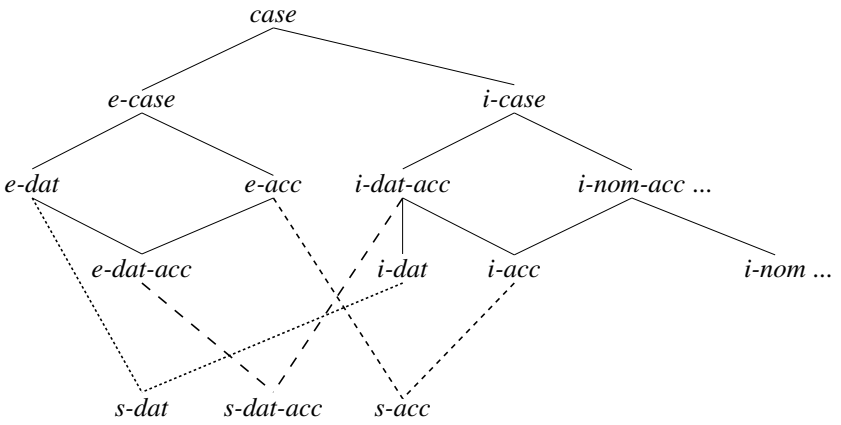


Inherent case specifications of dependents will be types in the *i-case* subhierarchy (for inherent case), whereas case requirements imposed by a subcategorising head will be values in the *e-case* subhierarchy (for external case). Unification of internal case specifications will result in disambiguation of underspecified case values, whereas unification of external case requirements will result in feature indeterminacy. To illustrate this, take the examples in (24) and (25): case ambiguous *Frauen* will be specified *i-dat-acc*, whereas unambiguous *Kindern* will carry the more specific value *i-dat*. Likewise, the verbs *finden* and *helfen* will subcategorise for an *e-acc* and *e-dat* complement, respectively. Coordination of the two lexical verbs will lead to unification of CAT values (Pollard and Sag, 1994),[9] and hence, valence lists, "overspecifying" the case requirement as *e-dat-acc*.

[8] In essence, the inverse layouts of the two subhierarchies correspond quite closely to the different behaviour of functor and argument categories with respect to strengthening/weakening in the approach of Bayer and Johnson (1995).

[9] For an overview of the treatment of coordination in HPSG, see Crysmann (in press).

(35)



In order to permit satisfaction of any subcategorised case by some inherent case, all we need to do is define the greatest lower bound for any pair of internal and external case specification.

Thus, underspecified internal cases will unify with a corresponding neutral case, whereas specific internal cases will only unify with their corresponding non-neutral cases. As depicted above, more specific types in one hierarchy will be compatible with less specific types in the other, and vice versa. Returning to our example above, underspecified *i-dat-acc*, as in *Frauen* unifies with overspecified *e-dat-acc*, as required by the coordination *findet und hilft*, whereas unambiguous *Kindern* does not, since no greatest lower bound is defined for *i-dat* and *e-dat-acc*. Thus, disambiguation of *i-case* values will always reduce the potential for neutrality, as required. On a more conceptual level, these cross-classifications between the two hierarchies embody the logical link between underspecification and neutrality.

4.3 Discussion

The reader familiar[10] with recent work on non-constituent coordination within HPSG (Yatabe, 2003; Crysmann, 2003, to appear) may have noticed that these accounts already provide an alternative solution to the problem addressed by Daniels (2001): instead of coordinating two verbs with conflicting subcategorisation requirements, one might equally well assume coordination of VP or S, where identical peripheral material is simply suppressed:

(36) Er findet (Frauen) und hilft Frauen.
he finds and helps women
'He finds and helps women.'

Although, purely theoretically, this is indeed a valid objection, once we look at available implementations of the HPSG formalism, we must conclude that sharing of domain objects is unsupported. As a consequence, in the light of implemented

[10]This issue has actually been brought to my attention by Carl Pollard.

HPSG grammars, Daniels (2001)'s approach to neutrality is still without competition.

On the other hand, the realisation of closed-world type identity by way of typed list constraints may equally well prove as an alternative approach to non-constituent coordination. One of the main concerns in current linearisation-based approches to the phenomenon (Crysmann, to appear; Yatabe, 2003) is to ensure that instantiations of valence lists of a head within one conjunct do not, inadvertantly, get identified with the valence lists of the shared, unexpressed head in the other conjunct. While Yatabe (2003) adresses the issue by explicitly composing pairs of valence list instantiations from both conjuncts, Crysmann (to appear) chooses to restrict sharing of domain objects to head information, basic phonology and the key semantic relation, thereby ensuring a sufficient degree of relatedness, without requiring identity of dependents, or even events.

Yet, once we subscribe to the idea that valence patterns within a language draw from a finite set, and that these patterns can be compactly represented as types, we have the necessary prerequite in place for an account of head-sharing which is independent of domain object sharing, or even non-continuity: in essence, a phenomenon like conjunction reduction can then be modelled by creating a type-identical copy of the overt head, and saturate its valence lists with the non-head constituents of the second conjunct.

5 Conclusion

In this paper we have discussed how Flickinger (2000)'s type-based approach to the representation of inflectional feature specifications can be applied to syncretism in German. In particular, we have shown how likeness constraints abstracting out a particular inflectional dimension from a combined inflectional type hierarchy can be expressed concisely by means of typed lists, representing a closed-world analogue to type-identity. Furthermore, we have argued for an extension to Daniels (2001) original approach to feature indeterminacy in HPSG which makes it possible to combine the empirical virtues of his type-based approach to the phenomenon with the advantages of underspecified representation of syncretism across features, namely generality of specification and efficiency in processing.

References

Bayer, S. and Johnson, M. 1995. Features and Agreement. In *Proceedings of the 33rd Annual Meeting of the ACL*, pages 70–76.

Copestake, Ann and Flickinger, Dan. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*, Athens.

Crysmann, Berthold. 2003. *Constraint-based Coanalysis. Portuguese Cliticisation and Morphology–Syntax Interaction in HPSG*. Saarbrücken Dissertations in Computational Linguistics and Language Technology, No. 15, Saarbrücken: Computational Linguistics, Saarland University and DFKI LT Lab.

Crysmann, Berthold. in press. Coordination. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, Oxford: Elsevier, second edition.

Crysmann, Berthold. to appear. An Asymmetric Theory of Peripheral Sharing in HPSG: Conjunction Reduction and Coordination of Unlikes. In G. Jaeger, P. Monachesi, G. Penn and S. Wintner (eds.), *Proceedings of FGVienna: The 8th Conference on Formal Grammar, Aug 16–17 2003, Vienna*, Stanford: CSLI Publications.

Dalrymple, Mary and Kaplan, Ron. 2000. Feature Indeterminacy and Feature Resolution. *Language* 76(4), 759–798.

Daniels, Michael. 2001. On a Type-Based Analysis of Feature Neutrality and the Coordination of Unlikes. In *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, CSLI Online Proceedings, pages 137–147, Stanford: CSLI Publications.

Flickinger, Daniel P. 2000. On Building a More Efficient Grammar by Exploiting Types. *Natural Language Engineering* 6(1), 15–28.

Höhle, Tilman. 1983. Topologische Felder, ms., University of Cologne.

Ingria, R. J. P. 1990. The limits of Unification. In *Proceedings of the 28th Annual Meeting of the ACL*, pages 194–204.

Kathol, Andreas. 1999. Agreement and the Syntax-Morphology Interface in HPSG. In Robert Levine and Georgia Green (eds.), *Studies in Contemporary Phrase Structure Grammar*, pages 209–260, Cambridge and New York: Cambridge University Press.

Levine, Robert, Hukari, Thomas and Calcagno, Michael. 2001. Parasitic Gaps in English: Some Overlooked Cases and their Theoretical Implications. In Peter Culicover and Paul Postal (eds.), *Parasitic Gaps*, pages 181–222, Cambridge, MA: MIT Press.

Levy, Roger and Pollard, Carl. 2001. Coordination and Neutralization in HPSG. In *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, CSLI Online Proceedings, pages 221–234, Stanford: CSLI Publications.

Müller, Stefan. 1999. An HPSG-Analysis for Free Relative Clauses in German. *Grammars* 2(1), 53–105.

Müller, Stefan. 2001. Case in German – Towards an HPSG Analysis. In Walt Detmar Meurers and Tibor Kiss (eds.), *Constraint-Based Approaches to Germanic Syntax*, Studies in Constraint-Based Lexicalism, No. 7, pages 217–255, Stanford: CSLI Publications.

Pollard, Carl and Sag, Ivan. 1994. *Head-Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.

Uszkoreit, Hans, Backofen, Rolf, Busemann, Stephan, Diagne, Abdel Kader, Hinkelman, Elizabeth, Kasper, Walter, Kiefer, Bernd, Krieger, Hans-Ulrich, Netter, Klaus, Neumann, Günter, Oepen, Stephan and Spackman, Stephen P. 1994. DISCO - An HPSG-Based NLP System and its Application for Appointment Scheduling. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94), August 5-9*, volume 1, pages 436–440, Kyoto, Japan.

Yatabe, Shûichi. 2003. A Linearization-based Theory of Summative Agreement in Peripheral-Node Raising Constructions. In Jong-Bok Kim and Stephen Wechsler (eds.), *Proceedings of the 9th International Conference on Head-driven Phrase Structure Grammar, Kyung Hee University, Seoul, 5–7 August, 2002*, pages 391–411, Stanford: CSLI Publications.

Zaenen, Annie and Karttunnen, Lauri. 1984. Morphological Non-Distinctiveness and Coordination. In *Proceedings of the First Eastern States Conference on Linguistics (ESCOL)*, pages 309–320.