
Underspecification and Neutrality: a unified approach to syncretism

BERTHOLD CRYSMANN [†]

Abstract

In this paper I discuss the phenomenon of syncretism in German and show that current type-based approaches are unable to combine the treatment of feature indeterminacy 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. Finally, I will show how likeness constraints operating over a subset of the inflectional dimensions can be expressed by means of typed lists that abstract out the relevant dimension from the combined case/number/gender hierarchies suitable for syncretism.

Keywords SYNCRETISM, UNDERSPECIFICATION, INDETERMINACY, HPSG, GERMAN

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. Often, syncretism can-

[†]I would like to thank Stefan Müller and Michael Jellinghaus for fruitful discussion of several aspects of this work. Many thanks also to the three 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).

not be resolved to disjunctive specification or underspecification within a single feature, but it cuts across the three inflectional dimensions. However, since disjunctions are in general much harder to process than type inference, type-based underspecification of case/number/gender specifications appears to be the key towards an efficient and concise treatment of syncretism.¹

Ambiguous nominal forms in German are also subject to indeterminacy. 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 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. Furthermore, I will show 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 in coordinate structures. As a result, the current proposal presents an entirely disjunction-free approach to syncretism, addressing indeterminacy, underspecification and likeness constraints.

1.1 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.

- (1) Er findet und hilft Frauen.
he finds.A and helps.D women.A/D
'He finds and helps women.'
- (2) *Er findet und hilft Kindern.
he finds.A and helps.D children.D
- (3) *Er findet und hilft Kinder.
he finds.A and helps.D children.A

Unification-based frameworks such as LFG or HPSG have taken up the challenge, refining the representation of feature constraints in such

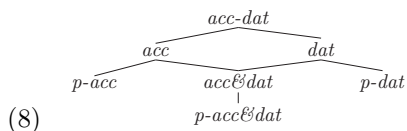
¹See the Surrey Morphology Group syncretism database for a cross-linguistic overview (<http://www.surrey.ac.uk/LIS/SMG/>).

grammar engineering. In the LinGO ERG (Flickinger, 2000), person and number are represented as values of a single feature PNG, permitting the expression of, e.g., non-3rd-singular agreement without the use of negation or disjunction.

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 even more pressing: 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. Intuitively, it would make perfect sense to try and exploit the combined type hierarchies required for the treatment of neutrality in order to arrive at a more concise and efficient representation of syncretism.

1.3 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 Karttunen (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.



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

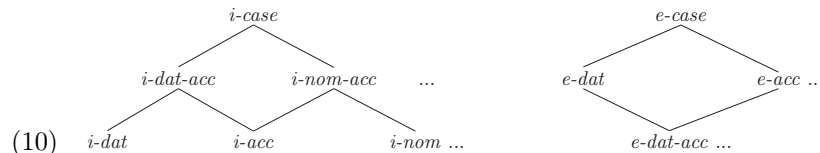
will still be a supertype of the neutral type $acc\mathcal{E}dat$, erroneously licensing the unambiguously non-dative *die Frauen* ‘the women’ in the neutral accusative/dative context of *findet und hilft* ‘finds and helps’.

- (9) * 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.

1.4 A Solution

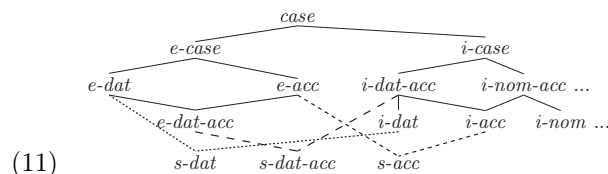
The reason for the apparent incompatibility of underspecification and feature neutrality lies with the attempt to address both aspects within a single type hierarchy. Instead, I shall argue to draw a principled distinction between inherent inflectional feature values, where unification specialises from 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*).³



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 (1) and (2): case ambiguous *Frauen*

³In 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).

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),⁴ and hence, valence lists, “overspecifying” the case requirement as *e-dat-acc*.



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.

1.5 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

⁴For an overview of the treatment of coordination in HPSG, see Crismann (to appear).

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 (12), and dative singular feminine, as illustrated in (13).

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

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

This ambiguity could be represented by a type $n-s-m+d-s-f$.⁵ Subcategorisation for nominative singular (type $n-s-g$) or dative (type $d-n-g$) will disambiguate these forms accordingly.⁶

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 (14). 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.

(14) 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 satisfy identical type requirements. This type equality is essentially imposed by disjunctive enumeration of the four possible subcategorised case values. In typed

⁵As 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.

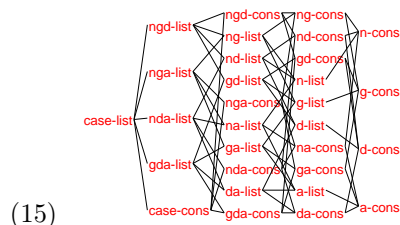
⁶For ease of exposition, I am abstracting away from the internal/external distinction, which is immaterial here, since we are only dealing with underspecification, not indeterminacy.

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.

An alternative, though not fully satisfactory solution 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 (15)⁷, where each list type immediately subsumes at least one subtype representing a non-empty list of the same case type.



Types in the combined case-number-gender hierarchy will now restrict their CASE value to an appropriate list type, as given in (16).⁸

$$(16) \quad nda-n-g \rightarrow [CASE \quad nda-list]$$

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 (17)).

⁷The type hierarchy has been exported from the LKB: supertypes are on the left, subtypes are on the right.

⁸Recall that, according to our naming convention, the type *nda-n-g* represents all case specification except genitive. Number and gender are fully underspecified.

$$(17) \quad nda-cons \rightarrow \langle nda-n-g \mid nda-list \rangle$$

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 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:

$$(18) \quad coord-phr \rightarrow \left[\begin{array}{l} SS \mid L \mid AGR \mid CASE \langle \boxed{1} \mid \boxed{2} \rangle \\ \text{COORD-DTRS} \left\langle \begin{array}{l} [SS \mid L \mid AGR \boxed{1}], \\ [SS \mid L \mid AGR \mid CASE \boxed{2}] \end{array} \right\rangle \end{array} \right]$$

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.

$$(19) \quad \left[\begin{array}{l} SS \mid L \left[AGR \mid CASE \langle \boxed{1} \mid list \rangle \right] \\ VAL \mid COMPS \left\langle [L \mid AGR \boxed{1}] \right\rangle \end{array} \right]$$

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):

$$(20) \quad \text{Wir}_i \quad \text{helfen ihnen}_j \quad [\text{einem} \quad \text{nach dem} \quad \text{anderen}]_{*i/j}$$

we.NOM help them.dat one.DAT.M after the.M other
'We help them one after the other.'

$$(21) \quad \text{Wir}_i \quad \text{helfen ihnen}_j \quad [\text{einer} \quad \text{nach der} \quad \text{anderen}]_{*i/j}$$

we.NOM help them.dat one.DAT.F after the.F other
'We help them one after the other.'

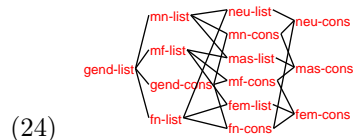
$$(22) \quad \text{Wir}_i \quad \text{helfen ihnen}_j \quad [\text{einer} \quad \text{nach dem} \quad \text{anderen}]_{i/*j}$$

we.NOM help them.dat one.NOM.M after the.M other
'We help them one after the other.'

- (23) 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 (20–23) 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- anderen-*, 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 (21) and (22), disambiguation of case syncretism by means of grammatical gender reduces the semantic attachment potential of the entire phrase, precluding attachment to the subject in (21), and to the object in (22).

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 mechanism 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 analogous abstraction of the gender dimension, a step, which is very much straightforward:



$$(25) \quad c-n-mn \rightarrow [\text{GEND} \quad mn\text{-list}]$$

$$(26) \quad mn\text{-cons} \rightarrow \langle c-n-mn \mid mn\text{-list} \rangle$$

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.⁹

$$(27) \left[\begin{array}{l} \text{SS | L | AGR } c-n-p \wedge \boxed{1} \wedge \boxed{2} \\ \left\langle \begin{array}{l} \text{PH } \langle \text{einer} \rangle \\ \text{SS | L | AGR } n-s-mn+d-s-f \wedge \boxed{3} \left[\begin{array}{l} \text{CASE | FIRST } \boxed{1} \\ \text{GEND | FIRST } \boxed{2} \end{array} \right] \\ \text{PH } \langle \text{nach der anderen} \rangle \\ \text{SS | L | AGR } d-s-f \wedge \left[\begin{array}{l} \text{CASE } d\text{-list} \\ \text{GEND } \left[\begin{array}{l} fem\text{-cons} \\ \text{FIRST } \boxed{3}c\text{-n-f} \end{array} \right] \end{array} \right] \end{array} \right\rangle \end{array} \right]$$

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.

1.6 Conclusion

In this paper 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. We have further shown how likeness constraints abstracting out a particular inflectional dimension from a combined inflectional type hierarchy can still be expressed concisely by means of typed lists.

⁹In order to make the lexical specification of case/number/gender information more transparent, I have left the unification of values in (27) unresolved.

References

- Bayer, S. and M. Johnson. 1995. Features and agreement. In *Proceedings of the 33rd Annual Meeting of the ACL*, pages 70–76.
- Crysmann, Berthold. to appear. Coordination. In K. Brown, ed., *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 2nd edn.
- Dalrymple, Mary and Ron Kaplan. 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 R. Levine and G. Green, eds., *Studies in Contemporary Phrase Structure Grammar*, pages 209–260. Cambridge and New York: Cambridge University Press.
- Levine, Robert, Thomas Hukari, and Michael Calcagno. 2001. Parasitic gaps in English: Some overlooked cases and their theoretical implications. In P. Culicover and P. Postal, eds., *Parasitic Gaps*, pages 181–222. Cambridge, MA: MIT Press.
- Levy, Roger and Carl Pollard. 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 of free relative clauses in German. *Grammars* 2:53–105.
- Pollard, Carl and Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Stanford: CSLI and University of Chicago Press.
- Zaenen, Annie and Lauri Karttunen. 1984. Morphological non-distinctiveness and coordination. In *Proceedings of the First Eastern States Conference on Linguistics (ESCOL)*, pages 309–320.