MORPHOLOGICAL NON-DISTINCTIVENESS AND COORDINATION

Annie Zaenen (Cornell University)
Lauri Karttunen (SRI, CSLI)

In traditional linguistic literature it is quite often debated whether forms like the ones given under (1) are to be considered one word or should be considered to represent several homonymns, i.e. whether in the lexicon, or wherever one wants to treat these forms, one should have one entry, partially specified like in (2a) or a set of entries partially specified like in (2b):

(1) take, are, can, etc.

(2) (a) take V { TEN = pres | TEN = -}  
           - [PER = 3  
               NUM = sg]  

(b) take V TEN = pres  
    PER = 1  
    NUM = sg  

take V TEN = pres  
    PER = 2  
    NUM = sg

(b) take V TEN = pres  
    PER = 1  
    NUM = pl

(b) take V TEN = pres  
    PER = 2  
    NUM = pl  

The representation in (2a) specifies take as a form that is either a present tense form but not sg 3 or a tenseless form; (2b) spells out all the alternative feature combinations in full detail.

Bloomfield (1933) expressed the view that the second type of representation was the right one, whereas Bloch (1947) and Karttunen (1984) opt for the former. (An overview of some of the approaches is given in Huddleston 1975).

As far as we know, however, no syntactic arguments have been given to bear upon the choice of two types of representations. In this paper we will give an argument that argues for the first type of representation as the right one in certain cases and we will discuss some of the consequences of that representation for the treatment of coordination.

Some coordination facts.

It is well-known that in most cases coordination is only possible when the syntactic requirements that are imposed on each conjunct by the rest of the sentence are the same; e.g. we cannot say

(3) *The dog took and went a walk.
This is so because whereas *take* takes a DO as a complement, *go* takes a prepositional object and hence on the V level we cannot coordinate.

In languages with case-marking, these identity requirements of course include identity of case. Hence, in Icelandic for instance, one cannot say

(4) *Hann stal og bordadi kókuna (acc)/kókunni (dat).*

He stole and ate the cookie.

This is so because the verb 'eat' takes an accusative object whereas 'steal' takes a dative object. Because the accusative form of 'cookie' is kókuna and the dative form of 'cookie' is kókunni, neither form fulfills the case requirements of both verbs. Another example, from French, is given in (5):

(5) *Je l'ai mis à la porte et donné des coups de pied.*

I him have put at the door and given some kicks.

This in spite of the fact that sentences like (6) are, of course, completely grammatical:

(6) *Je l'ai frappé et mis à la porte.*

I him have hit and put at the door.

So the coordination of the VP material following the auxiliary is normally allowed. What makes (5) ungrammatical is that the accusative form of the 3 pers sg clitic before a vowel is l whereas the dative form is lui.

Similar examples can be given for Finnish, German, Serbo-Croatian and other languages, although it is not totally clear to us that all languages, or speakers, have strict case agreement requirements of the sort illustrated above. Native speakers of Russian, for instance seem to be less adamant about the ungrammaticality of that type of sentence. We will stick here to languages (or speakers?) for which sentences like (4) and (5) are ungrammatical.

For those languages we now want to look at instances where the different cases are morphologically not distinct. We give a French example first. A few possible lexical entries for preverbal clitics in French are given in (7). (7a) gives a representation where all the forms are thought of as independent lexical items whereas (7b) gives a representation in which it is assumed that, when there is only one form, there is only one lexical item. In (7a) there is a dative and an accusative me, in (7b) we just have a single me whose case is either accusative or dative.
(7) (a) me CL CASE = dat
    PER = 1st
    NUM = sg

    me CL CASE = acc
    PER = 1st
    NUM = sg

    le CL CASE = acc
    PER = 3rd
    NUM = sg

    lui CL CASE = dat
    NUM = sg
    PER = 3rd

(b) me CL CASE = {dat | acc}
    PER = 1st
    NUM = sg

le [as in (7a)]
lui [as in (7a)]

A simple interpretation of the difference in representation under (a) and (b) is that (a) predicts that a sentence like (8) should be ungrammatical because the two conjuncts cannot possibly fulfill an identical case requirement, whereas the representation under (b) would be required if the sentences are grammatical.

(8) Il m'a frappé et donné des coups de pied.

As we have seen above frapper asks for an accusative whereas donner des coups requires a dative clitic object. The lexicon does not contain a form that can fulfill these two requirements at once. Under the representation in (b), however, the lexicon does contain such forms, and, given a compatible representation of coordination, the sentence could be grammatical. It turns out that sentences like (8) are grammatical for most speakers of French, although some prescriptive grammarians condemn them (see Grevisse 1980 and Kayne 1975 for discussion).

Similar judgments obtain for the following German case, that illustrates the same point (Eisenberg 1973).

(9) der Antrag des oder der Dozenten
    the petition of the docent or docents.

This is grammatical because Dozenten can be either plural or singular: if we try a similar coordination replacing 'docent' with 'professor' the result is ungrammatical because the gen sg is Professoren whereas the plural is Professoren:

(10) *der Antrag des oder der Professors
    *der Antrag des oder der Professoren.

Similar examples can marginally be constructed in English:

(11) this and these sheep.
We also find them in Icelandic, as illustrated in (12):

(12) Hann stal og bordáði kóku.
He stole and ate a cookie.

(12) differs from (4) only in that the object is indefinite and
that there is no difference in case-marking for kóku 'cookie' in
the dative and accusative. The difference that is found in (4)
comes from the fact that the postposed and cliticized definite
article exhibits a difference in casemarking in the dative and
accusative singular, feminine forms. Speakers of Icelandic,
however, differ in their judgment about (12). Whereas the
sentence is grammatical for some it is judged ungrammatical by
others (the small sample of informants we have consulted does not
suggest that the difference is regional or linked to age group or
social class).

One interpretation of the Icelandic results is that we need
the representation under (13b) as well as that under (13a):

(13) (a) kóku: N CASE = {dat | acc}    (b) kóku: N CASE = dat
    kóku: N CASE = acc.

Speakers who find (13) ungrammatical could have the
representation under (13b) whereas speakers that find the sentence
grammatical would operate under the representation under (13a).

In the next two subsections we want to investigate further
the consequences of the type of morphological neutralization for
some areas of syntactic theory.

Feature percolation and coordination.

As we said in the introduction, when one has two co-ordinated
parts of a sentence, the constituent(s) they share have to obey
whatever constraints are imposed by each of the conjuncts. (For
the general approach to coordination that we are assuming here see
Sag et al. forthcoming.) The facts given above constrain the
mechanisms by which this can be done and they suggest that the
conjoining node cannot play a role in enforcing these
requirements.

We will illustrate this by looking at the Icelandic example.
First we extend the GFG account of subcategorization in an
obvious way, proposing that verbs that subcategorize for dative
are introduced by rule (14a), whereas verbs that subcategorize for
accusative are introduced by rule (14b).

(14) (a) < 3, [ V NP]
        VP 3 DAT

(b) < 4, [ V NP]
        VP 4 ACC
When two verbs are conjoined, their subcategorization features presumably have to be shared by the coordinated structure to exclude sentences like (15) in a maximally simple way.

(15) 'Hann bakadi og bordali kókunni.
He baked(+acc) and ate(+acc) the-cookie(dat).

Assuming that the subcategorization feature appears on the coordinating V, this sentence is ungrammatical, simply because there is no PS rule that would allow for it.

Given these natural assumptions, however, we run into trouble with sentences like (12). Here the value of the first V for the subcategorization feature clashes with that of the second V, so the intersection is empty, predicting that the sentence is ungrammatical. If we, on the other hand, were to allow the coordinate structure to have no subcategorization feature at all when there is a conflict, then we would expect (4) to be grammatical.

The right generalization seems to be that the constraints on verb coordination should simply require that the subcategorization requirements of all of the conjuncts be fulfilled. It is clear that we cannot rule out all of the bad cases or admit all of the good ones by rules that assign subcategorization values to the coordinate structure as a whole. In the case of non-NP conjunction, it seems best to think of the conjunction as not having any such features at all. (For a general proposal along these lines, see Bresnan et al. forthcoming.)

 Even if we do not percolate features from conjoined nodes up to the mother node, the phenomenon at hand is still problematic for all frameworks that use unification as a general operation for computing feature values (Shieber and Pereira 1984). Consider again the German example given above. We will assume as just argued that the num feature does not show up on the coordinating node, but that agreement between a DET and a N is done by merger or unification of the feature values of the DET and those of the N, as diagrammed in (16):

(16)

```
       NP
        /  \
       DET  N
      /    |
     DET   [NUM = {sg | pl}]
    /  \
  des  oder
   /    |
  der  Dozenten
```
Even under the representation of Dozenten proposed above, simple unification of this N with both DET will not work because the operation is transitive. When the first determiner's [NUM = sg] is merged with the noun's [NUM = {sg | pl}], both operands become the same, i.e. [NUM = sg]. Consequently, the unification of the noun's feature with the second determiner feature [NUM = pl] now fails. The simple reason is that, while Dozenten can be either singular or plural, it cannot be both at once. No matter what the right feature values are for the N and the NP, we cannot get it by unification alone.

The basic problem is that a simple idea about how merging of features works entails that the operation should fail when two conflicting values for the number feature are assigned to one and the same noun. This is obviously the wrong result. In the case at hand it is even difficult to see what the right result ought to be. It is unclear whether des oder der Dozenten is singular or plural or whether it really has any number at all.

Which features can be neutralized?

Regardless of the right way to handle the problem pointed out in the previous section, we can ask another question about the cases of morphological neutralization illustrated in (8), (9), (12). Can any old feature be underspecified? Can we get conjunction of any morphologically non-distinct material? We of course know that this is not the case. For example, we cannot take advantage of the two homonymous readings of flying to form conjunctions like (17).

(17) *Flying planes can be easy and invisible.

We would like to predict that this type of coordination is impossible, even when it can be plausibly argued that the category of the two conjuncts is the same. Where is the cut-off point between the possible and impossible cases? As the following example from Finnish shows, acceptable cases of non-distinctness may involve more than just one feature dimension. In that respect it is more exotic than the French and German cases discussed above.

(18) He lukivat hänen uusimman _ ja me
    They read his newest (sg gen) and we

    hänen parhaat _ kirjansa.
    his best (pl nom) book/books.

The facts about Finnish morphology that make this possible are summarized in (19):
(19) 1. Possessive suffixes obliterate some case distinctions.

Sg Nom: kirja 'book' kirja-nsa 'his book'
Sg Gen: kirja-n 'book' kirja-nsa 'his book'
Pl Nom: kirja-t 'book' kirja-nsa 'his book'

2. The case of the object alternates between partitive, genitive, and nominative.

Gen: He lukivat kirja-n They read the book.
Nom: Me luettlin kirja We read the book.

3. Adjectives agree in number and case with their head.

Sg Nom: uusin kirja newest book
uusin kirjansa his newest book

Sg Gen: uusimman kirjan newest book
uusimman kirjansa his newest book

Pl Nom: uusimmat kirjat newest books
uusimmat kirjansa his newest books

The possessive suffix -nsa 'his/her/their' wipes out any preceding case ending which consists of a single consonant. The same is true of other possessive suffixes. This results in a three-way loss of distinctions. A form like kirjansa can be either sg nom, sg gen, or pl nom. It is the non-distinctness of the last two feature combinations that makes (18) possible.

The following case from German appears at first very similar to the Finnish example because it, too, involves neutralization along two dimensions. For certain type of nouns, such as Dozent, singular accusative and plural dative are the same: den Dozenten. One might expect to be able to take advantage of this by having a noun phrase of this form serve as the object for two conjoined verbs which differ with respect to the case they assign to their object. One such a pair is sehen (+acc) 'see' and helfen (+dat) 'help.' However, examples of this sort, such as (20), turn out to be totally unacceptable to German speakers.\3

(20) *I habe den Dozenten gesehen und geholfen.
   I have seen the docent (sg acc)
   and helped the docents (pl dat).

   Another case where morphological non-distinctness fails to
   sanction conjunction is shown in (21).

(21) *Zij schijnt en kijkt onbevangen.
   She seems and sees candid(ly).
(22) (a) Zij schijnt onbevangen.
She seems candid.
(b) Zij kijkt onbevangen.
She is looking candidly (around).

In Dutch there is no morphological difference between adverbs and adjectives but verbs like schijnen 'seem' take AP's as possible complements whereas verbs like kijken 'look' in the sense of 'see' take ADVP. The morphologically non-distinct form cannot be used to fill a different position in both conjuncts, although the language doesn't make a morphological distinction between adverbs and adjectives across the board.

It seems to us that the unacceptability of examples like (20) and (21) is essentially semantic, not syntactic. We think it is due to the following simple constraint, the Anti-Pun Ordinance (APO):

ANTI-PUN ORDINANCE: A phrase cannot be used in two different senses at the same time.

In other words: no puns. The effect of the APO can be seen most easily by considering examples like (23a) and (23b).

(23) (a) She is married to John and she is married to Bill.
(b) She is married to John and to Bill.

Because it contains two occurrences of she, (23a) leaves open the possibility that there are two individuals involved. This option is is not available in (23b); for it to be true, some woman has to be married to both John and Bill. To put it in a more general way: in interpreting a conjunction of the form

\[ X \land (A \land B) \land Y \]

the content of \( X \ldots Y \) has to remain constant. It is not possible, for example, to have she in (23b) refer to one individual in relation to Bill and to another one in relation to John.

The ungrammaticality of the German example in (20) is a simply another manifestation of the same constraint. The NP den Dozenten can be used to refer to a single individual or to a group of individuals but it cannot be used in both ways at the same time, as (20) requires. The same is obviously true of the Dutch example; they involve two different semantic interpretations of onbevangen. Consequently, (21) is not possible. This is also the most straight-forward explanation for the ungrammaticality of (17), although one could also attribute it to some violation of syntactic structure. The NP flying planes can designate a set of individuals or a set of situation types (?) but not both.
While ruling out examples like (24), the Anti-Pun Ordinance does allow examples like (25).

(24) *Yesterday John, and tomorrow his friends, hit the nail on the head.

(25) He has and will hit the nail on the head.

The anomaly of (24) is due to the same reason that makes (23b) unambiguous. The phrase hit the nail on the head could refer either to a past or non-past time but it can't do both at the same time. In (25), on the other hand, tense reference is contributed by the two conjoined auxiliaries and hit the nail on the head is only interpreted in one way. The Finnish example in (18) is parallel to (25). The shared part, kirjansa, is not a noun phrase but a common noun. Although it is part of a singular noun phrase in one conjunct and a plural noun phrase in the other, there is no violation of the APO. We assume here that reference is assigned to noun phrases, not to common nouns. This is the one crucial difference between (18) and (20).

While the APO is clearly necessary, it doesn't seem to be sufficient, however, to explain all the data we have come across. Consider the following case in Finnish. Given the fact that (20) and (21) are both grammatical, one would expect (23) to be just as grammatical in Finnish as it is in English. But it is not, and it is not obvious that there is any way to explain this by the APO.

(26) Vaimoni voi siivota.
    My wife (sg nom) can clean

(27) Vaimoni täyttyy siivota.
    My wife (sg gen) must clean

(28) *Vaimoni voi ja täyttyy siivota.
    My wife (sg nom.sg gen) can and must clean

The reason for the ungrammaticality of (28) is not entirely clear, although it obviously has to do with the fact that, in Finnish, necessity modals like täyttyy 'must' require that the NP that plays the role of the subject for the infinitive phrase have genitive case, whereas voi 'can' goes with nominative NPs. We are not sure whether the difference is syntactic or semantic or both.

The traditional view is that the two classes of verbs have distinct syntactic properties. In standard grammars of Finnish, necessity modals are said to be "impersonal." According to that view 'my wife' in (27) is not a subject but some kind of adverbial. From that point of view, it would be better to gloss it as 'it is necessary for my wife to clean.' If this were so, it would distinguish the case in hand from the French (8) and Icelandic (12) examples discussed earlier which seem to involve no difference in grammatical function. If we were to say that 'my
wife' is the subject in both (26) and (27) and that the necessity modals just assign a quirky case to their subject, then it would be a mystery why (28) isn't just as good as (8) and (12). Note also that in (26) 'my wife' controls the number of the verb but the verb in (27) would be singular even if the NP were plural.

Unfortunately there are other cases of similar nature where it is much harder to make a case for a difference in syntactic function. We consider one more example from Finnish and then show that the same problems arise in French.

(29) Hän ihastui Kaliforniaan
She fell in love with California (illative).

(30) Hän muutti Kaliforniaan.
She moved to California (illative).

(31) *Hän ihastui ja muutti Kaliforniaan.
She fell in love with and moved to California.

The problem with this case does not have to do with morphological non-distinctness. In (29) and (30), Kaliforniaan is in the same locative case but it appears to play a different semantic role. It is much less clear than the previous case that there is any difference in constituent structure or in syntactic function. It is, of course, common to distinguish locative, temporal, and other kinds of adverbial phrases but the criteria for making such distinctions seem to be primarily semantic. It is hard to find any clearly syntactic criteria that would necessitate creating distinct syntactic roles or categories for all the different kinds of adverbials that one may wish to distinguish on the basis of their semantic interpretation. This looks like a violation of the APO but we are not sure how to explain it in detail.

French also shows that "surface" grammatical function alone is also not enough to insure conjoinability:

(32) *Je l'ai fait embrasser à Marie et sortir
I had Mary kiss him and (made) him leave.

This in spite of the fact that (33) and (34) are grammatical and that all theories that we know of would analyze je as the surface object of both (faire)-sortir and (faire)-embrasser:

(33) Je l'ai fait embrasser à Marie.

(34) Je l'ai fait sortir.

On the other hand, we also know of cases where difference in syntactic function does not seem to block coordination. For example, in Finnish a noun phrase can serve simultaneously as the subject of a verb and as the object of an impersonal "passive" verb:
(35) Kirje saapul ja luettlin.
The letter arrived and was read.

It is generally assumed that kirje 'letter' (sg nom) is the subject of saapul 'arrive' and surface object of luettlin 'read.' The conjunction works because both verbs agree on the case of the shared constituent. On the other hand, one cannot say:

(36) *Kirje saapul mutta el luettu.
The letter arrived but was not read.

This is ungrammatical because there is a case conflict. The negative verb phrase ei luettu 'was not read' requires a partitive object, kirjettä, which is distinct from the nominative form.

We leave the statement of the exact conditions for further research.

FOOTNOTES

*We thank Wayne Harbert, Joan Maling, Geoff Pullum, Hans Uszkoreit, Tom Wasow, and Arnold Zwicky for their help.

\(^1\) As the two gaps in (18) indicate, we analyze it as a "Right Node Raising" type of construction. Because Finnish in general is very liberal in allowing adjectives to be used like nouns, one could take the view that 'his newest' in (18) is a complete noun phrase on its own that has no syntactic relationship to the dangling noun at the end. However, if that were the case, one would expect to find a possessive suffix on the adjective: hänen uusimpana 'his newest (one). Without the suffix, hänen uusin __ can only be construed as an incomplete NP whose head with its attached suffix is missing.

\(^2\) This is colloquial Finnish. In Standard Finnish one says Me lukinme kirja-n (sg gen) 'We read the book.'

\(^3\) The existence and relevance of this type of example has been pointed out to us by Wayne Harbert, Geoffrey Pullum, and Arnold Zwicky. See Pullum and Zwicky (forthcoming).

\(^4\) (28) contrasts with Talon voi ja täytyy silvota 'My house can and must be cleaned.' This is grammatical in Finnish because there is no overt subject at all; taloni 'my house' (sg nom) is the object of the embedded verb with respect to both modals.

\(^5\) Finnish does not have real passives. The construction that traditionally is called "passive" is a close relative of German and Swedish sentences with man as an unspecified human subject. Even intransitive verbs have "passive" forms in Finnish. Note also that the choice of 'arrive' as the first verb in (35) is not significant. Any VP that requires a nominative subject would be just as good here.
REFERENCES

Bresnan, J., Kaplan, R., and P. Peterson (forthcoming) "Coordination and the Flow of Information in Phrase Structure."
Pullum, G. and A. Zwicky (forthcoming) "Phonological Resolution of Syntactic Feature Conflict," to be presented at the 59th Annual Meeting of the LSA, Baltimore, Ma.