Lexicalizing number and gender in Colonnata

Knut Tarald Taraldsen
Center for Advanced Study in Theoretical Linguistics
University of Tromsø
knut.taraldsen@uit.no

1. Introduction

Current "late insertion" accounts of the relationship between syntactic structure and vocabulary items elements (morphemes) tend to share the two assumptions in (1) and (2):

(1) Vocabulary insertion targets only terminal nodes
(2) *The Subset Principle*
A vocabulary item A associated with the feature set F can replace a terminal X with the feature set F' if and only if F is a subset of F'.

However, a growing body of conceptual and empirical considerations suggests that (1) and (2) should be replaced with (3) and (4), as argued extensively by M. Starke (CASTL research seminars); cf. Abels & Muriungi (2008), Caha (2007, in prep.), Muriungi (2008), Taraldsen (2009):

(3) Vocabulary insertion targets subtrees
(4) *The Superset Principle*
A vocabulary item A associated with the feature set F can replace a subtree X with the feature set F' if and only if F is a superset of F'.

The feature set of a subtree will be the set of features associated with its terminal nodes, and analyses adhering to (3)-(4) typically also assume that each head corresponds to a single feature.
On the conceptual side, we note, in particular, that if each syntactic head actually corresponds to exactly one syntactic feature, increasingly fine-grained syntactic analyses will favor (3) over (1) simply because the heads in a tree will generally outnumber the morphemes used to lexicalize it.
The purpose of this squib is to provide an additional empirical argument for moving from (1)-(2) to (3)-(4).

2. Feminine plural marking in Lunigiana

Manzini & Savoia (2005: III, 618 ff) report that various Italian dialects in the Lunigiana area never use the standard Italian *–e* as a marker of f.pl. Instead, one finds *–ya* on determiners, nouns and adjectives which would have *–a* in the f.sg, e.g. *∂ona* "woman" vs *∂onja* "women" in the Colonnata variety. Like Manzini & Savoia, we think the null hypothesis should be that *–ja* is decomposable as *–i -a* where *–a* is the usual feminine gender marker also found in the singular forms, while *–i* - is the plural affix also found in the plural of masculine nouns, both in Lunigianese, e.g. *kwanti omi* "how many men" (Filattiera), and in Standard Italian. So, both on the standard approach and on ours, lexical insertion will target two syntactic nodes independently:

(5) [N [pl [f]]]
| |
i *a*

We take it that the configuration in (5) is created by successive movement of the N to the specifier of the pl through the specifier of the f, as shown in (8):

(6) [N [pl [~~N~~ f [~~N~~]]]

That is, we assume that the heads pl and f appear in their underlying order. One reason for assuming this, is that it seems natural to assume that the category Gender, being directly involved in the selection of the N, should be below Number (pl).
We also assume that traces can be ignored, when the lexicalization procedure parses a syntactic structure. Then, [pl [f]] is a subtree in (6), and it could be targeted by vocabulary insertion of *–e*, on our assumptions, assuming that *–e* has the lexical entry in (7)a:

(7)a e ←→ {pl,f}
b [N [pl [f]]]
|___|
|
e

This would produce the Standard Italian f.pl forms. Lunigianese, however, doesn't have the Standard Italian *–e*. So, instead, f and pl are lexicalized separately, as would, of course, be entirely possible also on standard accounts.

3. Silent pl in Colonnata

In the Colonnata dialect of Lunigianese, the plural *–i-* on the feminine noun seems to be in complementary distribution with the plural marking on determiners and quantifiers, i.e. when the noun cooccurs with an article or quantifier marked with *–i-* (reflected only in the palatalization of the *l-* of the article in (8)a), only the gender marker *–a* appears on the noun:

(8)a λa ∂ona
the-pl-f woman-f
"the women"
b tantja ∂ona
so.many-pl.f woman-f
"so many women"

Otherwise, the noun has the full f.pl inflection *–ia-*, e.g. *tre d'donja* "three women".
We take this to be an ellipsis phenomenon in the specific sense that whenever the element lexicalizing pl is spelled out on a determiner or a quantifier, it is not also pronounced on the noun. But from this perspective, it is surprising that *–i* is in fact always spelled out on a m.pl form (with the exceptions discussed in section 6):

(9) kwanti omi (Filattiera)
how.many-pl man-pl
"how many men"

(10)a a λ o camata (Colonnata)
scl ocl-f.pl I-have called-*a*
"I have called them (feminine)."
b a j o camati
scl ocl-m.pl I-have called-*i*
"I have called them (masculine)."

In particular, this is surprising if *–i-* is just specified as {pl}, as it would have to be on an analysis of f.pl-*ia* adhering to (2). On an account assuming (4), the Superset Principle, rather than (2), however, we can take advantage of the fact that even with feminine nouns the gender marker is always spelled out. The following sections will show how this works.

3. Lexicalizing m in Lunigiana

We know from (8) that although the exponent of pl is not pronounced in the presence of an inflected determiner, the exponent of Gender (*-a*) is. But most masculine nouns do not have an exponent of gender in the singular, i.e. no *–o* comparable to the Standard Italian m.sg. *–o*:

(11) om fradel
man brother

As for those few that do, we will provisionally assume that the final vocalic element is epenthetic, but return to the issue below.

Rather than posit Ø ←→ {m} , we assume that the masculine gender m is lexicalized by the root along with N (a set of features, i.e. a phrase), e.g. /om/ = [N [m]] (assuming also privative pl, although that is not critical):

(12) [N [m]]
|___|
|
om

But in the masculine plural forms (produced by N-movement without pied-piping), N and m do not form a subtree:

(13) [**N** [pl [**m**]]]

Therefore, no root can lexicalize N and m together in the plural forms.

However, if there is no zero exponent for m, the existence of plurals like *omi* "men" in non-elliptic contexts will force the language learner to set up the entry in (14) for *-i*, assuming, as seems plausible, that every node must be lexicalized; cf. Fábregas (2007):

(14)a *i* ←→ { pl, m}
b [N [pl [m]]]
|___|
|
i

Notice that the Superset Principle makes this fully consistent with the hypothesis that *i* lexicalizes just pl in f.pl forms.

4. Why ellipsis preserves the *i* in the m.pl

Suppose now the structure in (13) is embedded under a determiner with its own pl marking. On the basis of (8), we now expect the pl head in (13) not to be pronounced. Yet, as (9) shows, the noun will have a final *–i*. The conclusion reached in section 3 provides two ways of explaining this, corresponding to two different ways of modeling the ellipsis effect.
If the silent pl head is never lexicalized, the Superset Principle allows the two distinct lexicalization patterns depicted in (15):

(15)a [N [pl [m]]]
| |
om *i*
b [N [pl [m]]]
| |
om *om*

To weed out option (15)b, we must then assume that the set of features abbreviated by N in *om* ←→ {N, m} has at least two members, and invoke the principle in (16) (see Taraldsen (2009) for other applications of this):

(16) *The Best Fit Principle*
If two vocabulary item A and B are both candidates for lexicalizing the subtree X, and the number of features of B not matched by X exceeds the number of features of A not matched by X, A is preferred.

The Best Fit Principle picks up on one aspect of the Elsewhere Principle familar from other frameworks. As formulated in Distributed Morphology, for example, the Elsewhere Principle gives precedence to the candidate that matches the most of the features of the lexicalization target. In a theory assuming the Subset Principle, this is equivalent to saying that the candidate with most restricted distribution wins. In a theory based on the Superset Principle, saying that the candidate with the most restricted distribution wins, translates to (16).
Alternatively, we view the ellipsis effect seen in (8) as a matter of not pronouncing a pl after it has been lexicalized. Thinking of lexicalization as associating a vocabulary item simultaneously to all the terminals of the subtree it lexicalizes, ellipsis could then be viewed as delinking. The surface form of the nouns in (8) would then be produced via the two-step process in (17):

(17)a [N [pl [f]]]
| | |
∂on i a (lexicalization)
b [N [pl [f]]]
| -|- |
∂on i a (ellipsis)

In completely analogous fashion, the derivation of the form *omi* ”men” in (9), would proceed as in (18):

(18)a [N [pl [m]]]
| |__|
| |
om i (lexicalization)
b [N [pl [m]]]
| -|-__|
| |
om i (lexicalization)

After the vocabulary item *i* has been delinked from pl, it is still attached to m, and so, will be pronounced, unlike the *i* in (17).
In this scenario, the lexicalization of m by *om* is excluded independently of the Best Fit Principle and concomitant assumptions about the cardinality of N. The framework we are adopting, also includes the principle in (19), which, as given here, maximizes lexicalization targets; cf. Caha (2007, in prep.), Muriungi (2008), Taraldsen (2009) for slightly different formulations:

(19) *The Maximize Target Princple*
If a subtree X is part of a larger tree Y that can be lexicalized by some vocabulary item A, X cannot be lexicalized separately.

This excludes lexicalizing the plP as in (20), since *i* alone can lexicalize [pl [m]] :

(20) [N [pl [m]]]
| | |
* *om i om*

For the purposes of this discussion, the principle in (19) is given in a user-friendly version. However, (16) and (19) are both shown to be derivable from a single more general principle in the appendix to Taraldsen (2009), incorporating ideas by M. Starke. It also follows from that account that (19) will take precedence over (16).
Either way, we have the result that the *i* of the m.pl forms is retained under ellipsis for exactly the same reason as the Gender-marking *a* of the f.pl forms: Ellipsis doesn’t involve Gender.

5. *–o* again

The view of ellipsis adopted in the second scenario described in section 4 must be adopted, if the final vowel (apparently a schwa, but written as*–o* below) that appears on some m.sg nouns is not epenthetic, but rather an exponent of masculine Gender:

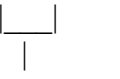
(21) *o* ←→ { m }

We then need to exclude m.pl forms **N-i-o* and, under ellipsis, **N-o*, parallelling the f.pl forms.

Given the entry in (14)a (repeated below) for *i*, the Maximize Target Principle automatically delivers the desired result outside the ellipsis contexts, since *–i* can lexicalize all of [pl [m]]:

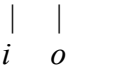
(14)a i ←→ { pl, m}

(22)a [N [pl [m]]]



i (conforms to (19))

b [N [pl [m]]]



**i* *o* (violates (19))

The elliptical m.pl **N-o* will be excluded exactly the same way, if ellipsis is delinking subsequent to full lexicalization.

But if the pl head is not lexicalized at all under ellipsis, as in the first scenario in section 4, the existence of (21) would predict m.pl *N-o* in ellipsis contexts. The Best Fit Principle would favor the lexicalization in (23)b over the one in (23)a:

(23)a [N [pl [m]]]

**i* (violates (16))

b [N [pl [m]]]

o (conforms to (16))

Thus, if we are forced to analyze the *–o* seen on some m.sg nouns and adjectives as a lexicalization of masculine gender, we are also forced to assume that ellipsis works as in the second scenario in section 4, involving delinking as assumed for independent reasons in an analysis of Bantu concords in Taraldsen (2009).

6. Another declension class

Manzini & Savoia's (op. cit.) description indicates that even some m.pl forms may in fact lose the *i* in Lunigianese dialects. But the only examples given are the adjective *zoven* "young" (Standard Italian pl *giovani*) in Filattiera and the noun *kan* "dogs" (Standard Italian *cani*) from Bedizzano. These contrasts with m.pl adjectives like *bravi* "good, clever" and nouns like *omi* "men" in the same contexts.

If the examples are representative, we have a contrast between declension classes. Adjectives like *zoven* and nouns like *kan* fall into the class of masculine nouns and adjectives whose sg forms end in *–e* rather than *–o* in Standard Italian. In Lunigiana, they apparently have no final vowel in the singular.
We suggest that the members of this declension class have a different syntactic derivation of the plural forms. Instead of the straight successive cyclic derivation in (24), they have the "roll-up" derivation in (25), where the raising N pied-pipes the GenderP at the second step of the derivation:

(24) [N [pl [~~N~~ m [~~N]]]~~
(25) [[N [m [~~N~~ ...]]] [pl ~~[[N [m [N~~]]]

Assuming that the Lunigiana roots in this declension class have entries like (26), this allows a straightforward derivation of m. pl forms like *zoven* and *kan* with pl-ellipsis:

(26)a *zoven* ←→ { A, m }
b *kan* ←→ { N, m }
(27)a [[A [m]] pl]
|___| |
| |
zoven *i* (lexicalization)
b [[A [m]] pl]
___	
zoven *i* (ellipsis)
(28)a [[N [m]] pl]
|___| |
| |
kan *i* (lexicalization)
b [[N [m]] pl]
___	
kan *i* (ellipsis)

(As before, we assume that the lexicalization procedure disregards traces.)
This account can be extended to the corresponding declension class in Standard Italian, which has sg *giovane*/ pl *giovani* and sg *cane*/ pl *cani*, provided we now take the Standard Italian sg *–e* to be epenthetic, as suggested by the fact that it may fail to appear in certain contexts, unlike the f.pl *–e*. This, in turn, will lead us to say that the Standard Italian m.sg *–o* of the other declension class is not epenthetic, but rather a lexicalization of the gender head m. Notice, however, that this conclusion does not necessarily extend to the final vowel that sometimes appears in m.sg forms in Lunigiana, apparently always a schwa. To determine the status of this element, we will have to see if it appears in both of the two declension classes we have now posited, under conditions that can be characterized in phonological terms. If it does, we will treat it as epenthetic. If it doesn't, we will adopt the analysis suggested in section 5.

The declension class now characterized by the roll-up derivation in (25) also comprises feminine nouns and adjectives in Standard Italian, e.g. *luce* "light" (pl *luci*)and *felice* "happy" (m/f.pl *felici*). A feminine noun like *luce* can be associated with the entry in (29), while accommodating adjectives like *felice* and ensuring that the pl ending is *–i* (rather than *–e*, as in the f.pl forms of the other declension), requires elaborating the representation of gender as proposed below in section 7:

(29) *luc* ←→ { N, f }

For Standard Italian, this will lead to the lexicalization patterns in (30) (the sg form) and (31) (the pl form):

(30)a [N [f]]
|__|
|
luc
b epenthesis: *luce*

(31) [[N [f]] pl]
|___| |
| |
luc *i*

However, transposing this to the Lunigiana varieties leads to the prediction that feminine nouns in the "*luce*-declension" should never have pl forms in *–ia* (as opposed to just *–i*) and should be just the bare root under pl-ellipsis, exactly like masculine nouns and adjectives of the *kan*/ *zoven*-type. As yet, we have not been able to verify whether the prediction is borne out.

7. The representation of feminine gender

In some Italian dialects all plural nouns and adjectives end in *–i* in both genders (Manzini & Savoia 2005:III, 583ff). This cannot be accounted for by extending the analysis assumed above for nouns and adjectives like *giovane* and *cane* in Standard Italian and in the dialects of the Lunigiana. In the varieties we now turn to, the m.sg has the ending *–u* and f.sg *–a* (except for nouns and adjectives of the *cane*-class). That is, m and f are lexicalized by *–u* and *–a* rather than by the root. Correspondingly, it seems that *–i* lexicalizes pl together with either one of the two gender heads m and f.

This can be easily accommodated in a system based on (1)-(2): Assuming fused Gender/Number heads, it suffices to specify *–i* only as pl. But in a system adhering to (3)-(4), it appears that we have to posit two distinct lexical entries *i* ←→ {pl, f} and *i* ←→ {pl, m}.

In some of the dialects discussed by Manzini & Savoia there may in fact well be two distinct *–i*'s, since the m.pl *–i* produces morphophonological alternations in the root not found with the f.pl *–i*, e.g. m.pl *nøvi* "new" vs f.pl *novi* in Borgomanero. But other varieties in the relevant group, e.g. the dialect of S. Lorenzo del Vallo, offer no such indication that m.pl *–i* is

distinct from the f.pl *–i*. So, I will tentatively assume that some dialects do in fact have a single *–i* lexicalizing pl in both geneders, and explore the modifications needed to make this consistent with (3)-(4).

In fact, the only option seems to be to represent one of the two genders as a span of two heads where the second head by itself would represent the other gender. Assume, for example, that the feminine gender is characterized as f [m, while m alone identifies the masculine gender. Conceptually, we can think of m as representing a set out of which f extracts a subset. This move doesn't affect the analyses given previously. Where some morpheme was taken to lexicalize f, we now think of it as lexicalizing f [m. Where N raised past f without pied-piping, it will now raise past f [m without pied-piping, and where N pied-piped f, it will now pied-pipe f and m.

But the innovation obviously makes a difference for the analysis of *–i* lexicalizing pl in both genders, since we can now posit a single entry *–i* ←→ {pl, f, m}, which, by the Superset Principle, allows *–i* to lexicalize both [pl [f [m]]] (f.pl) and [pl [m]] (m.pl).

It also makes a difference for the analyses of adjectives like *felice*, which must lexicalize agreement features of both genders. Each of these can now have a single entry rather than two, e.g. *felic* ←→ {pl, f, m}. Moreover, we can ensure that feminine nouns and adjectives in this class take *-i* rather than *–e* in the plural even in Standard Italian by positing the entries *e* ←→ {pl, f, m} and *i* ←→ {pl, m}:

(32)a [[A [f [m]] pl]
|___|__| |
| |
felic ***i*** (conforms to (16))

b [[A [f [m]] pl]
|___|__| |
| |
* *felic* ***e*** (violates (16))

Notice finally that the analysis suggested for dialects like S. Lorenzo del Vallo predicts that f.pl nouns and adjectives in such dialects will preserve the *–i* as a lexicalization of f [m even if pl is subject to ellipsis, provided ellipsis is delinking subsequent to lexicalization, as suggested in section 4. The examples given in Manzini & Savoia (2005: III, 584-588) do indeed indicate that the f.pl *–i* is never lost in these dialects. But then again, we cannot really tell whether the ellipsis process is operative in these dialects to begin with, and the example from Celle di Bulgheria in (33), with *nuci* probably in the declension class discussed in section 6 (like Standard Italian *noce*), may suggest it is not:

(33) i n'nuci
the.f.pl nut-pl

8. Conclusion

We have seen that an analysis based on the assumptions in (3) and (4) can make sense of the contrast between f.pl and m.pl nouns with respect to the retention of *–i-* under pl-ellipsis in the Lunigiana varities of Italian. The basic factual assumption is that the *i* lexicalizes just the number head pl in the f.pl forms, but just the gender head m in the m.pl. This assumption is

consistent with the Superset Principle, but not with the Subset Principle. Hence, if correct, our analysis favors adopting (4), the Superset Principle, over (2), the Subset Principle.
As far as we can see, though, the choice between the Superset Principle and the Subset Principle is partially independent of (1) vs (2):

(1) Vocabulary insertion targets only terminal nodes
(3) Vocabulary insertion targets subtrees

In particular, adopting the Superset Principle appears compatible both with (1) and (3) (whereas adopting the Subset Principle within a theory based on (3) would allow dramatic underlexicalization of subtrees).
However, other parts of the analysis are clearly inconsistent with (1). Consider, for example, the fact that *i* must lexicalize both pl and m in the m.pl forms. This is consistent with (1) only if pl and m fuse into a single terminal prior to lexical insertion. But introducing fusion amounts to add a new kind of grammatical process just in order to immunize (1) against empirical observations favoring (3).
Going beyond the issue of (1)-(2) vs (3)-(4), we would also like to point out that if the tentative account of declension classes in section 6 is on the right track, it appears that the derivations underlying inflected forms may employ much the same mechanisms as the derivations described by Cinque (2005) in his account of Greenberg's Universal 20, i.e. they must be part of the syntax, a conclusion reached on the basis of different facts in Muriungi (2008).

References

Abels, K. & P. Muriungi (2008), The focus particle in Kîîtharaka: Syntax and semantics, Lingua 118: 687-731
Caha, P. (2007), Case movement in PPs, in Nordlyd 34.2, www.ub.uit.no/munin/nordlyd
Caha, P. (in preparation), The nanosyntax of case, PhD dissertation, CASTL
Cinque, G. (2005), Deriving Greenberg's universal 20 and its exceptions, Linguistic Inquiry 36: 315-332
Fábregas, A. (2007), An exhaustive lexicalization account of directional complements, in Nordlyd 34.2, www.ub.uit.no/munin/nordlyd
Manzini, M. R. & L. M. Savoia (2005), I dialetti italiani e romanci, Edizioni dell'Orso, Alessandria
Muriungi, P. (2008), Phrasal movement inside Bantu verbs, PhD dissertation, CASTL
Taraldsen, K. T. (2009), The nanosyntax of Nguni noun class prefixes and concords, ms, CASTL (available at lingBuzz/000876)