

Luca Dini
Scuola Normale Superiore & CSLI

Più and Dictionary Syntax: A Case Study

Abstract

The aim of this paper* is to show that the insertion of appropriate part of speech indicators will ease the work of lexicographers in defining the grammatical context of use for function words.¹ By appropriate lexical categories I mean syntactic labels drawn not only from the traditional descriptive grammars, but also from current (formal) syntactic theories. To prove this hypothesis I have analysed how Italian dictionaries² deal with entries used in comparative constructions. More specifically, the attention has been focused on the *più–di/che* ('more–than') pair, trying to determine how far dictionary definitions miss generalizations and ignore syntactic constraints placed on the use of these words.

0. Introduction

This paper is mainly concerned with the study of the interactions between formal linguistic theories and practical lexicography. I will show that certain lexical items which are usually assigned different syntactic labels in current Italian dictionaries could be more suitably classified by abandoning some of the traditional grammatical distinctions and adopting more structured lexical classifications, such as the one exploited, for instance, in Head-driven Phrase Structure Grammar (in the following HPSG). By analyzing certain properties of Italian comparative constructions,³ I will prove that this move would considerably enhance the capability of dictionaries to capture both distributional and relational properties of the words they classify, thus resulting in a much more reliable tool.

As for terminology, I will make use of the following terms:

- antecedent to mean all the words which can introduce a comparison (e.g. *più* 'more', *meno* 'less', *così* 'as');
- complementizer to mean the word which introduces the phrase in relation to which the comparison is made (e.g. *che* 'than' in *più mele che banane* 'more apples than bananas');
- than–phrase or comparative clause or second comparant to mean the constituent formed by the complementizer plus the phrase in relation to which the comparison is made (e.g. *di Bill* 'than Bill' in *Paula è più intelligente di Bill* 'Paula is more clever than Bill')

The term *quantifier* will be defined *in itinere*, as well as *specifier* and *determiner*. However, to avoid misunderstanding, it is worth noticing that the

words *specifier* and *determiner* will always be used to mean constituents (or lexical categories) able to occur in a pre-head position, irrespective of their traditional classification.

1. A syntactic overview

The main hypothesis underlying our syntactic treatment of comparatives is that the antecedent *più* 'more' belongs to a unique category, this category being able to occur in a number of positions:

noun modifier:	<i>Mangio più ossa di Fido</i> (I) eat more bones than Fido
adjectival modifier:	<i>Anna è più bella di Rita</i> Anna is more beautiful than Rita
adverbial modifier:	<i>Mangio più velocemente di te</i> (I) eat more quickly than you
sentential modifier:	<i>Paperino sta soffrendo più di Topolino</i> Donald Duck is suffering more than Mickey Mouse

Departing from considerations of this sort, Bresnan (1973) argues that *more* always projects to a maximal category QP⁴ which is allowed to occur in all the slots listed above. Here I will not focus the attention on the derivation of the lexeme *more* (or *più*), since the transformational devices used by Bresnan are of no interest in the present work. Indeed my attempt is to give a context-free analysis of comparative structures in order to derive a syntactic classification usable in dictionary definitions.

The idea of having a single category for the antecedent of comparison has been developed by other authors, who, in context-free formalisms, have tried to develop a syntactic theory where a particular slot was foreseen for quantifiers and specifiers. In Generalized Phrase Structure Grammar (GPSG), for instance, it is argued (Gazdar 1981) that *more* is rewritten as a specifier, be it a nominal specifier or an adjectival one. This line has been pursued by most of the "mildly context sensitive formalisms" (Partee & al. 1990), including HPSG, where comparative antecedents are assigned the type *SPEC-SIGN* (Pollard & Sag 1994). Of course the idea of having a unique class of words able to change the status of a phrase has been borrowed from X-bar theory.⁵

1.1 The category of the antecedent

In the following two paragraphs I will try to show that *più* should be theoretically assigned to a unique syntactic category, even if particular features (namely +/- adverbial) distinguish two different uses of the same word. These uses are called *quantificational* and *adverbial*. From a lexicographic point of view I will argue that the usual splitting of the

quantificational use of comparative antecedents into two or even more grammatical categories tends to hide a number of syntactic generalizations.

1.1.1. Comparative antecedents are quantifiers

I will pursue the idea of a single category for the antecedent of comparison by trying to prove its nature of quantifier.⁶ This is done by comparing the behaviour of *molto* ('many/much': a quantifier) with *più* and showing that a single paradigm is required to account for their distribution.

- Quantifiers generally occur in a pre-head position with nouns adjectives and adverbs:⁷
 - (1) *molte/più mele* *molto/più intelligente* *molto/più velocemente*
 many/more apples very/more clever very/more quickly
- Quantifiers occur in a post-head position when the head is a verb:
 - (2) *Giovanni mangia molto/di più*
 John eats very much/more
 - (3) *Giovanni molto/di più mangia.*
 John very much/more eats
- Quantifiers generally alternate with other kinds of specifiers, such as determiners or demonstratives:
 - (4) *molte/più/le mele*
 many/more/the apples
 - (5) * *molte/più le mele*
 many/more the apples
- Normally a single quantifier is allowed for each phrase:
 - (6) * *più molto bello*
 more very beautiful
- Quantifiers can be left-modified by other quantifiers (or degree phrases or negation):
 - (7) *decisamente molto poco più intelligente di Mario*
 decisely very few more clever than Mario
 - (8) *decisamente troppo poco pane*
 decisely too-much few bread

1.1.2 A borderline case.

There are sentences where the comparative antecedent *more/più* does not co-occur with normal quantifiers (henceforth: adverbial comparison):

- (9) *Ha parlato più a John che a Bill*
 He spoke more to John than to Bill
- (10) * *Ha parlato [molto a Giovanni]*⁸
 He spoke [much to John]

The following observations are possible about this kind of construction :

- Normal quantifiers always occur as specifiers of an XP category, while adverbial quantifiers are typical XP adjuncts:
 - XP → QP_[-adv], X'
 - XP → QP_[+adv], XP
- In adverbial comparisons, the *than*-phrase has to be present.⁹
 - (11) *Quest'anno le industrie giapponesi hanno prodotto più automobili*
(normal comparison)
This year Japanese firms have produced more cars.
 - (12) * *Quest'anno le industrie giapponesi hanno prodotto più le automobili* (adverbial)
This year Japanese firms have produced more the cars
- The *than*-phrase always introduces a constituent of the same functional category as the one modified by the antecedent:
 - (13) * *John ha letto più i libri (NP_[obj]) che Bill (NP_[subj])*
John read more the books(NP_[obj]) than Bill(NP_[subj])

In the literature there are two hypotheses about the category of the antecedent involved in this kind of comparison:

- (a) It is some special category belonging to a class which is very close to the one of coordinative conjunction (Napoli-Nespor 1986, Hankamer 1973, Ryan 1983, a.o.).
- (b) It is still a quantifier, even if it is always attached to the S or VP node (Bresnan 1973, Dini 1992).

Here I will not consider hypothesis (a) mainly for two reasons: (i) it leaves a number of facts completely unexplained (e.g. the left modification of the antecedent or the possibility for the *than*-phrase to undergo movements); (ii) even though proved theoretically well-founded, it could not be adopted in the actual coding of a dictionary entry as it would contradict every "traditional" grammatical intuition of the user.

1.2 A frame for quantifiers

In the preceding paragraph I have proved that the introduction of a well defined category *quantifier* allows one to capture a number of syntactic generalizations which would otherwise be split up for every entry belonging to that class. In the present paragraph I will show that *più/more*, like many other quantifiers, has a *frame*, and I will argue that the introduction of such a concept will ease the work of lexicographers in accounting for facts which would remain unexplained otherwise.

The concept of syntactic frame is encoded in various ways in current dictionaries, either as a specification of category (e.g. *transitive*, *intransitive*,

pronominal ...) or as separate syntactic information (e.g. COBUILD: **V+o**, **V+o+c** ...) or as a feature elicitable from the examples ("ascribe (...) ~sth to sb/sth" OALD). This rudimentary notion of frame, while satisfying only to a partial degree the specification for frames (or subcategorization lists) used in most current syntactic theories, seems to be extremely well suited to increase the users' comprehension of some grammatical properties of words. Unfortunately lexicographers tend to ascribe frames *only* to verbs, neglecting the fact that other grammatical categories are able to subcategorize for fixed complements. Certain kinds of quantifiers are among them.

1.2.1 Features of frames

Elements bearing a frame specification (or, in theories such as HPSG, a subcategorization list) have to fulfill two main requirements: (i) they must be the syntactic head of the phrase containing the complements they select; (ii) they must specify the syntactic features of their complements (e.g. the preposition by which a PP should be introduced).

It is trivial to show that the comparative antecedent fulfills both these requirements and that the *than*-phrase is exactly its complement. Indeed :

- comparative antecedents have the distribution of QP's, independently of the presence of a *than*-phrase (they are heads):
 - (14) *Quest'anno ho letto più libri di Bill*
This year I have read more books than Bill
 - (15) *Quest'anno ho letto più libri*
This year I have read more books
 - (16) * *Quest'anno ho letto libri di Bill*
This year I have read books than Bill
- comparative antecedents select the particle (or complementizer) by which the *than*-phrase is introduced:
 - (17) *Più dischi che libri*
More records than books
 - (18) *Tanti dischi quanti libri*
As many records as books
 - (19) **Più dischi quanti libri*
More records as books
 - (20) **Tanti dischi che libri*
As many records than books

2. The current coding of *più*

The main conclusions of Section 1 were that (i) the antecedent of the comparison has to be ascribed to the category *quantifier* possibly specified as for the [+/- adverbial] feature; (ii) the quantifiers involved in comparison

have a subcategorization frame (i.e. they select the second comparant). As for (i), all the Italian dictionaries consulted tend to neglect the [+ adverbial] use of *più* and, by contrast, show a systematic splitting of *più* [- adverbial] as adjective and adverb, the former occurring in a pre-nominal position, the latter in a pre-adjectival one (cf. 2.1). As for (ii) the concept of frame is either completely neglected (thus causing redundancies and inconsistencies) or introduced in some indirect form (Palazzi Folena)(2.2).

2.1 Problems of categorial assignment

In the following I will show some of the inconsistencies generated by the grammatical assignments of the consulted dictionaries. In addition I will show how the [+adverbial] use of *più* is ignored by most Italian dictionaries.

2.1.1 The adjectival reading

All the consulted Italian dictionaries assign the category of adjective to the reading of *più* as a nominal specifier. This strategy (i.e. the assumption of the identity between quantifiers and adjectives) does not cover semantic properties of nominal quantifiers and, in addition, it allows the following (incorrect) predictions (under this hypothesis all the ungrammatical phrases of the list below would be considered grammatical):

word order:	<i>ragazze belle / * ragazze più</i> girls beautiful/girls more
coordination:	<i>belle e intelligenti ragazze / * belle e più ragazze</i> beautiful and clever girls / beautiful and more girls
interactions with determination:	<i>Ho visto delle belle/*molte/*più ragazze</i> I have seen some beautiful/many/more girls

The list could be further extended, but I think that these reasons are enough to justify at least the introduction of a new category able to replace the label *adjective* in the case of quantifiers.

2.1.2 Adverbial Reading.

The assignment of *più* (and other quantifiers of this sort) to the category adverb (needed to justify its pre-adjectival occurrence), leads to incorrect predictions about the positions where it can occur, which are in fact more restricted than normal adverbials. Consider for example:

- (21) *Ha parlato a Giovanni e, probabilmente/*più/*molto, ad Antonio*
He has spoken to Giovanni and, probably/more/much to Antonio
- (22) *Potrebbe tranquillamente/*più parlare con chiunque*
He could easily/more speak with everybody

The fact that adverbial and quantifier distribution is not isomorphic is noted in some way by the lexicographers, but the attempts to restrict the context of use of such a widely distributed category (adverbs) produces very approximate results. Zingarelli, for instance, restricts the context of occurrence of *più* as a sentential modifier to post-verbal position ('V _'), saying that it introduces a comparative proposition (where this ambiguous statement has to be interpreted as "it must introduce"). Note that this coding has two effects: it constrains the position of *più* as though it were a verbal quantifier (redundant information), and it misses the fact that comparative QP's, if saturated (i.e. with the argumental slot filled), have a wider distribution than normal quantifiers (e.g. **abbiamo parlato a Michele molto* 'We have spoken to Michele much', but *abbiamo parlato a Michele più di Antonio* 'We have spoken to Michele more than Antonio'). The Devoto "solves" every problem allowing premodification by *più* only for adjectives, adverbs and verbs. As a consequence, clearly ungrammatical sentences like **Più Giovanni ha lavorato* 'More Giovanni has worked' might be accepted, while *Ha lavorato più di Giovanni* 'He has worked more than Giovanni' should be rejected. Only in the Palazzi–Folena the occurrence of *più* within sentential constructions is restricted to a post verbal position (not necessarily to an *immediately* post verbal position), which seems rather plausible, even though, as already said, redundant.

2.1.3 Missing a use of *più*

As I said in Section 1.1.2 there are cases when comparative antecedents are used in front of XP rather than X' (cf. *leggo più i libri che le riviste* 'I read more the books than the reviews' vs. *leggo più libri che riviste* 'I read more books than reviews'). In the same paragraph I also showed that in such constructions there is some reason to consider the category of *più* as an adverb behaving like a quantifier (or a quantifier with very idiosyncratic features). The striking fact is that these constructions (which are very ordinary in Italian) are completely neglected in Italian dictionaries. At the best (Palazzi–Folena), they tend to identify one of the possible meanings assumed by adverbial antecedents (namely the metacomparative reading of Pinkham (1982) or the absolute degree reading of Napoli–Nespor (1986)) with adverbial quantifier constructions, i.e. they simply assume a coincidence of a semantic reading and a syntactic phenomenon. Sentences like the following prove that this generalization does not hold:

- (23) *Ho parlato più ad Antonio che a Michele* (+adverbial, –met)
I have spoken more to Antonio than to Michele
- (24) *Amo gli animali più che i loro padroni* (+adverbial, +met)
I love the animals more than their bosses

2.2 Problems caused by the absence of a concept of frame

The absence of the concept of frame for adverbs and quantifiers usually causes a lot of redundancies and inconsistencies among the entries. Devoto, for instance, codes the same meaning of *più* twice, firstly introducing the definition "...Comparativo nel sistema di *molto*..." ("Comparative within the system of *molto* ['many/much']") and then asserting that "...Con gli agg. e avv. seguito da un termine di paragone, funge da segnale di comparativo di maggioranza..." ("with adj. and adv., if followed by a second comparant, it serves as a marker of comparison to a higher degree"). The fact which is patently missed here is that the antecedent of the comparison is the *main* element (head) in this kind of construction, the second comparant being nothing more than an optional complement. The "dramatic" effect of this repetition is that the reader could be led to think that two different kinds of *più* exist in Italian, the second one being actualized only in the presence of an overt second comparant. Zingarelli and Palazzi-Folena, more coherently, rely on the definition of the grammatical notion of "comparativo di maggioranza" ("comparison to a higher degree"), presupposing that the reader is acquainted with the fact that this grammatical construction introduces a second term of comparison. Unfortunately, since comparative constructions are defined in traditional grammars only with respect to pre-adjectival and pre-adverbial uses of *più*, they miss completely the fact that a second comparant may occur also when the antecedent is in a pre-nominal position.

Note that none of the three dictionaries makes any reference to the particle (*di/che*) by which the second comparant can be introduced (i.e. any reference to the syntactic subcategorization of the quantifier), which may cause a lot of trouble to readers whose mother tongue (e.g. English) allows only one comparative particle:

(25) *Lavoro più di te*

I work more than you

(26) *John e' più astuto che intelligente*

John is more witty than clever

3. Conclusions

At the light of the data exposed in Section 1, it seems that there are strong reasons to introduce a new categorial label (namely the one of *quantifier*). This new label, besides solving (at least in part) the problems raised in Section 2, would reveal extremely useful to handle a number of Italian words (such as *meno* 'less', *poco* 'few', *molto* 'many/much', *tanto* 'many/much', *troppo* 'too-many/too-much',...) which exhibit, in many respects, the same behaviour of *più*:

- (27) *Molte/troppe/tante/più mucche*
 Many/too-many/many/more cows
- (28) *Molto/troppo/tanto/più stupido*
 Much/too/much/more stupid
- (29) *Lavora molto/troppo/tanto/di_più*
 He works much/too-much/much/more

Note that some of the words listed above can select a complement, which confirm the fact that quantifiers should be allowed to have a frame (cf. 1.2):

- (30) *Tante mucche da ottenere tonnellate di latte*
 So-many cows to obtain tons of milk
- (31) *Tanto stupido da disgustare chiunque*
 So-much stupid to disgust everybody
- (32) *Lavora tanto da perdere l'autocoscienza*
 He works so-much to lose himself

Of course there are properties which are peculiar only to certain lexical items, but these properties are exactly the ones which should be mentioned in the dictionary definition.

Beyond the specific problems of *più/more*, two conclusions can be drawn from the present study:

(a) Current formal syntactic theories cannot simply be by-passed, as the amount of linguistic work undertaken in such frameworks could reveal very useful in a lexicographic context. There are essentially two ways of taking these theories into account:

(1) The information thus acquired is used to create lexicographic paradigms of classes of words (cf. Atkins & al. (1988)). These paradigms are used by the lexicographers to *predict* all possible types of behaviour of the headwords they are dealing with, even if the actual coding is maintained within the framework of traditional classifications.

(2) A global reconsideration of the use of certain labels in dictionaries is reevaluated in the light of recent syntactic developments. A more in-depth investigation should be made in order to understand how far the lexical devices used in formal theories (e.g. GB feature bundles, LFG and HPSG feature structures, lexical rules and metarules of most of the current syntactic approaches...) are transferable directly to lexicography.

(b) There is a certain variability about the possible degree of reusability of 'theoretical guidelines' coming from different frameworks. In particular it seems that theories relying on surface-based descriptions of the language are able to provide, all the rest being equivalent, analyses which can be more readily integrated into a lexicographic context. It is not a case that some of these theories (for instance: HPSG) have adopted a *hierarchical, type based*

organization of the lexicon which reflects, to a certain extent, a traditional methodology of lexicographic work. After all, the whole analysis sketched here about comparative antecedents was aimed to prove the existence of a *supertype* (the label 'quantifier') grouping together words traditionally classified under different syntactic categories (*più*-adj, *più*-adv, *più*-det). My intention here is not to say that all the burden of linking formal linguistic theories to lexicography should be undertaken only by lexicographers. On the contrary, there is a role for linguists in synthesizing from their theories guidelines for practical lexicographers: this task, beyond making the work of commercial lexicography easier and increasing the quality of dictionaries, would constitute a concrete beta-test for theories themselves.

Notes

- * This paper would have never come to light without the aid of Carla Marelli, whose hints have helped me in every stage of the preparation. I am also indebted to Sue Atkins whose meticulous revision has greatly improved the quality of this paper. A further revision is due to Chris Chambers. Of course all errors are mine.
- 1 Informally, we call function words all the categories whose import in the semantic interpretation of a phrase can be represented in terms of operators, system defined relations, and quantifiers.
- 2 The analysis was limited to *Vocabolario della lingua italiana Zingarelli* (1993) (Zingarelli), *Dizionario della lingua italiana Devoto-Oli* (1990) (Devoto), *Dizionario della lingua italiana Palazzi-Folena* (1992) (Palazzi-Folena).
- 3 The scope of the present work will be limited to two main aspects of comparative constructions, namely the assignment of the grammatical category to the antecedent of comparison (*più*) and its relation to the comparative clause (*che*/*di* ('than') + XP). No attention will be paid to idiomatic uses of *più*.
- 4 QP' is a notation for Quantificational Phrase bar one, which, in Bresnan's approach means the maximal projection of a quantifier.
- 5 I will not explain the details of X-bar theory. Here it is enough to say that a lexical head, say of category X, projects to X' together with its complements, and then to XP when the determiner is also found:

$$X' \rightarrow X, \text{COMP}$$

$$XP \rightarrow \text{DET}, X'$$
- 6 Within the present context, no semantic analysis is sketched, so the term *quantifier* will be used in an informal way, mainly referring to a cluster of syntactic properties rather than to common properties of interpretation.
- 7 The slash ("/") is used as a marker signalling that the phrase has to be considered *alternatively* with the various elements which surround it.
- 8 The square brackets are used to mark the interpretation in which sentences are grammatical/ungrammatical.
- 9 Omission of the second comparant is possible only in very marked contexts, and in this sense it differs very much from the omission of the second comparant in quantificational constructions.

References

- Atkins, B. T., Kegl, J. and Levin, B. 1988, "Anatomy of a Verb Entry: From Linguistic Theory to Lexicographic Practice", *International Journal of Lexicography*, 1:84-126. Oxford: Oxford University Press.
- Bresnan, J. 1973, "Syntax of the Comparative Construction in English", *Linguistic Inquiry* 4:275-343.

- Dini, L. 1992, La sintassi della comparazione in italiano: analisi e trattamento automatico, Unpublished Master Dissertation, Università degli Studi di Torino, 1992.
- Gazdar, G. 1981, "A Phrase Structure Syntax for Comparative Clauses" in T. Hoekstra & al. (eds.) *Lexical Grammar*, Dordrecht: Foris Publications. 379–393.
- Hankamer, J. 1973, "Why there are two Than's in English", *CLS* 9:179–191.
- Napoli, D.J. and Nespor, M. 1986, "Comparative Structures in Italian", *Language* 62:622–653.
- Partee, B.H., ter Meulen A. and R. Wall., 1990, *Mathematical Methods in Linguistics*, Dordrecht: Kluwer Academic Publishers.
- Pinkham, J. 1982, *The Formation of Comparative Clauses in French and English*, Harvard: Harvard University Press.
- Pollard, C. and Sag, I. 1994. *Head Driven Phrase Structure Grammar*. Stanford: University of Chicago Press and CSLI Publications.
- Ryan, K.L. 1983, "Than as Coordinator", *CLS* 19:352–361.

Appendix

In this appendix I show how a proper lexical entry for *più* could be arranged if the insertion of a new grammatical label (**quant**) were allowed. The reader can check that the new entry is at the same time more complete and less redundant than other dictionary codings (which are not reported for reasons of space), as a lot of properties can be directly inferred from the grammatical label. Idiosyncratic meanings of *più* have not been listed.

Quantificatore, gram, Parola che precede nomi, aggettivi o avverbi per specificarne la quantità o l'intensità. In costruzioni verbali segue il verbo ed indica la frequenza o l'intensità dell'azione descritta. E' flessa davanti ai nomi e inflessa in tutte le altre costruzioni: *molto bella, molto velocemente, molte ragazze*. Può essere preceduta da negazione, avverbio, o altro quantificatore: *non molto bella, non molti ragazzi, molto poco interessante*. Se seguita dalla preposizione *di* ed un nome plurale può dare origine a costruzioni di tipo partitivo: *molti di noi, troppe delle ragazze*. Talvolta regge una frase o un nome preceduto da preposizione: *troppo per poterlo sopportare, più di Carlo*.

più, quant. [solitamente seguito da un secondo termine di paragone, che può essere costituito da *di*+nome (*più alto di Matteo*), da *che*+qualunque altra parte del discorso (*più alto che largo*), da *di quanto* + frase (*compra più libri di quanti ne riesca a leggere*).] I. E' utilizzato quando si vuole descrivere la quantità espressa da un nome o l'intensità espressa da un aggettivo, verbo o avverbio facendo riferimento ad un'altra quantità (espressa dal secondo termine di paragone, solitamente ellittico) *Leggo più libri di Antonio, E' più alto di Mario, Lavora più di Caterina*. Talvolta il secondo termine di paragone resta sottinteso e può essere recuperato solamente nell'ambito del contesto: *Quest'anno il Giappone ha prodotto più automobili*. In caso di costruzioni verbali, se il secondo termine di paragone è assente *più* è sostituito da *di più*: *Oggi ho lavorato di più*. Quando *più* precede una preposizione o altra categoria contenente un articolo, prende obbligatoriamente il secondo termine di paragone introdotto da *che*: *Ho parlato più ad Angelo che a Michele, Leggo più i libri che i romanzi*. Quando il secondo termine di paragone è introdotto da *che* deve contenere una parola di categoria sintattica uguale a quella preceduta da *più*: *più alto che bello, più libri che riviste*. II. Le frasi comparative contenenti un secondo termine di paragone introdotto da *che* possono avere significato assoluto, ovvero esprimere negazione del primo termine di paragone: *la sua pazzia è più presunta che reale* (= non reale, ma presunta)