# Parting ways with the partitive view:
# a corpus-based account of the Italian particle "ne"

**Alice Mariotti**
University of Bologna
`alice.mariotti11@gmail.com`

**Malvina Nissim**
U. of Bologna / U. of Groningen
`m.nissim@rug.nl`

## Abstract

**English.** The Italian clitic "ne" is traditionally described as a partitive particle. Through an annotation exercise leading to the creation of a 500 instance dataset, we show that the partitive feature isn't dominant, and the anaphoric properties of "ne", syntactically and semantically, are what we should focus on for a comprehensive picture of this particle, also in view of its computational treatment.

**Italiano.** *Il clitico "ne" è noto come 'ne partitivo'. Attraverso un esercizio di annotazione che ha condotto alla creazione di un corpus di 500 esempi, osserviamo che il tratto partitivo non è affatto dominante nell'uso del "ne", e che per avere un quadro completo di questo clitico è necessario concentrarsi sulle sue caratteristiche anaforiche, sia a livello sintattico che, specialmente, semantico, anche per lo sviluppo di sistemi di risoluzione automatica.*

## 1 Introduction and Background

The Italian particle "ne" is a clitic pronoun. Traditionally, linguistic accounts of "ne" focus on two of its aspects: its syntactic behaviour and its being a conveyor of partitive relations.

Syntactically, this particle has been studied extensively, especially in connection with unaccusative verbs (Belletti and Rizzi, 1981; Burzio, 1986; Sorace, 2000). In Russi's volume specifically dedicated to clitics, the chapter devoted to "ne" only focuses on the grammaticalisation process which brought the clitic to be incorporated in some verbs, causing it to lose its pronominal properties. It is referred to as "the 'partitive' ne" (Russi, 2008, p. 9).

In (Cordin, 2001), the clitic is described in detail, and shown to serve three main uses. It can be a partitive pronoun, usually followed by a quantifier, as in (1). It can be used purely anaphorically to refer to a previously introduced entity, such as

"medicine" in (2). The third use is as a locative adverb, like in (3).[1]

(1)  Quanti <u>giocatori di quell'U17-U19</u> quest'anno o l'anno scorso hanno giocato minuti importanti in prima squadra? A me **ne** risultano *2 o 3*.
How many <u>players of that U17-U19 [team]</u> this year or last year have played important minutes in the first team? I think *2 or 3* [**of them**]

(2)  Tu sai che <u>la medicina</u> fa bene e pretendi che il palato, pur sentendo**ne** *l'amaro*, continui a gustarla come se fosse dolce.
You know that <u>the medicine</u> is good for you, and you ask your palate to enjoy it as if it was sweet, in spite of tasting [**its**] *bitterness*.

(3)  Me **ne** vado.
I'm leaving.

Note that for both partitive and non-partitive uses, in order to interpret the *ne*, the antecedent must be identified ("players of that U17-U19 [team]" in (1) and "medicine" for (2)). While there has been a recent effort to highlight the anaphoric properties of real occurrences of "ne" (Nissim and Perboni, 2008), there isn't as yet a comprehensive picture of this particle. In this paper, we contribute a series of annotation schemes that capture the anaphoric nature of "ne", and account for the different kinds of relations it establishes with its antecedent. We also contribute an annotated dataset that can be used for training automatic resolution systems, and that as of now provides us with a picture of this particle which is the most comprehensive to date.

## 2 Annotation schemes

Considering the examples above, we can see that the resolution of "ne" can be summarised as obeying the scheme in (4), where capturing the function of *ne* ($f(ne)$) is part of the resolution process. Figure 1 shows an example and its resolution.

(4)  $predicate + [f(ne) + antecedent]$

---

[1] Unless otherwise specified, all examples are from "Paisà" (Lyding et al., 2014), a corpus of about 250 mio tokens of Italian web data. The *ne* is bold-faced, the antecedent is underlined, and the predicate is in italics. Note that *ne* is often used as an enclitic, such as in (2). This can be the case with any of the three uses described.

Example: "Sto pensando a un modo per <u>staccare la spina</u>, **ne** *ho veramente bisogno*."
       'I'm thinking of a way to take I break, I really need to.'
$f(ne)$: "di + x" 'of + x'
**antecedent**: "staccare la spina" 'take a break' (lit. 'unplug')
**predicate**: "ho veramente bisogno" 'I really need to'

---

**resolution**: "ho veramente bisogno di staccare la spina" 'I really need to take a break'

Figure 1: Example of "ne" resolution components and procedure.

Table 1: Annotation scheme for the element "ne". The most used classes are highlighted in boldface.

| ne | anaphoric | partitive | **type** **token** |
|----|-----------|-----------|---------------------|
| | | ¬ partitive | **internal** **external** **prepobj** |
| | | cataphoric vague | |
| | ¬ anaphoric | | |

In order to account for all the entities involved in 4, we developed a set of annotation schemes that define three elements (ne, antecedent, and predicate) and their respective attributes. The schemes mainly build on on our own analysis of a random selection of corpus occurrences and on the only existing corpus-based work on "ne" (Nissim and Perboni, 2008), over which we introduce three substantial improvements:

**(i)** we distinguish between *type* and *token* for partitive uses to account for the difference between Example (7) and Example (8) below.

**(ii)** we add the values *internal*, *external*, and *prepobj* for non-partitive cases (see Section 2.1).

**(iii)** we mark explicit links between anaphor and antecedent, and anaphor and predicate.

Both (i) and (ii) are quite crucial conceptual distinctions, as we will see both in the scheme description as well as in the analysis of annotated data, while (iii) is important in the implementation of an automated resolution process. Additionally, the annotation is performed by means of a different, more appropriate annotation tool.

## 2.1 Scheme for *ne*

The scheme is summarised in Table 1. The primary branch for "ne" is its anaphoricity, which is

a binary feature. All cases of non-anaphoricity are basically idiomatic uses, especially with pronominal verbs (see also Example 3). These cases won't be further specified in the annotation.

Differently, anaphoric occurrences are classified as one of four types: *partitive*, *non-partitive*, *cataphoric*, *vague*. The rare cataphoric cases (6) are annotated as such without additional features. The value vague is used when the instance is anaphoric but with an unclear or unspecified antecedent (5), with no further annotation.

(5) L'aggettivo puoi anche metterlo dopo il sostantivo, a questo modo potresti continuare: "l'alba rugiadosa **ne** *trae prestigio ed eleganza*".

You can even place the adjective after the noun, and so you could continue: "the dewy dawn *gains prestige and elegance* **from it**

(6) *Ce* **ne** *fossero ancora molti* di <u>preti</u> nello scautismo . . .

I wish there still were many priests in the scouting movement . . .

The main distinction is thus between partitive and non-partitive uses. Both values are then further detailed, and the resulting five categories – boldfaced in Table 1 – are the core of the scheme. Below we explain the opposition between *type* and *token* references for partitive cases, and the difference between *internal*, *external*, and *prepobj* for non-partitive cases.

**partitive**    Consider Examples (7)-(8).

(7) **type** – Ho comprato quindici <u>paste</u>, e **ne** avrei prese *ancora*!

I bought fifteen *pastries*, and I would have bought even *more*!

(8) **token** – Ho comprato <u>quindici paste</u>, **ne** ho mangiate cinque.

I bought *fifteen pastries*, and I ate five [**of those**].

While in (8) the antecedent of "ne" is the whole NP "quindici paste" (fifteen pastries) and the predicate thus selects a subset ("cinque", five) of those

fifteen, the antecedent of "ne" in (7) is not a specific set of pastries, rather the class "pastries". Indeed, the predication is not about a portion or set of the aforementioned "fifteen pastries", but rather on instances of "pastries" in general. The *type* cases are akin to nominal ellipsis (Lobeck, 2006), and the type/token distinction is reflected in there being a direct or indirect anaphoric link.

The above are examples we made up for the sake of clarity: the contrast is explicit thanks to the occurrence of the same head noun. The opposition can anyway be observed in actual extracted data, too, and we report two cases in (9) and (10).

(9) **type** – non solo non risolve [. . . ] i problemi che l'hanno scatenata, ma li aggrava e **ne** *crea di nuovi ancora più gravi.*

(10) **token** – È uno spettacolo grandioso costato 150 milioni di dollari; e probabilmente **ne** *incasserà il triplo o il quadruplo.*

**non partitive** We introduce three new distinctive values which capture both semantic and syntactic aspects. First, syntactically, we specify whether the function of "ne" is resolved within the predicate's argument structure, and it is thus annotated as a "prepositional object" (*prepobj*), as the example in Figure 1. Second, semantically, we distinguish between what we call "internal" and "external" references. "Internal" is a reference made – through the predicate of "ne" – to a feature which is already part of the antecedent, as in (12). Another case of "internal" is (2), where the bitter taste is an internal feature of the medicine. With "external" we mark references which introduce some feature – again, via the predicate – which is external to the whole represented by the antecedent, as in (11). This distinction, neglected in the literature, has semantic implications on the part-whole relation which gets established, or even created, between anaphor and antecedent.

(11) **external** – [. . . ] il possesso dell'oggetto del nostro desiderio, posticipando**ne** *la soddisfazione.* [. . . ] the possession of the object of our desire, procrastinating its satisfaction.

(12) **internal**[2] – [. . . ] il possesso dell'oggetto del nostro desiderio, posticipando**ne** *l'intensità.* [. . . ] the possession of the object of our desire, procrastinating its intensity.

---

[2]This example is made up on the basis of the *external* one above for easing the comparison.

## 2.2 Scheme for *antecedent*

The antecedent is what resolves the anaphoric interpretation of "ne". In the examples concerning the distinction between type and token, the antecedent is "paste" ('parties') in (7) and "quindici paste" ('fifteen pastries') in (8). While in both examples the antecedent is an NP, it is possible for a VP (as in Figure 1) and even a full sentence (S) to serve as antecedent. The annotators are asked to mark as *antecedent* the whole linguistic expression which they identify as the antecedent, and to assign some syntactic features to it. The annotation scheme is summarised in Table 2.

Table 2: Annotation scheme for antecedent

| antecedent | NP | subject | modified ¬ modified |
| | | object | modified ¬ modified |
| | | other | |
| | VP | | |
| | S | | |

For each antecedent we thus specify its syntactic category, its grammatical role distinguishing just between *subject*, *object*, and any other role (*other*). For antecedents featuring as subject or object we also specify whether they have any sort of modification (adjectives, relative clause, and so on).

## 2.3 Scheme for *predicate*

The predicate of "ne" is what provides the completion to the interpretation of the anaphoric relation. In terms of annotation we specify only whether the predicate is a noun phrase or a verb phrase, and in the former case whether it is a modified NP or not.

## 3 Data selection and annotation

We collected 500 random occurrences from PAISÀ (Lyding et al., 2014), a web corpus of Italian which contains however good quality data. Instances of "ne" were extracted in a context of two preceding and one following sentence with respect to the matching sentence. In a given paragraph, then, possibly more than one occurrence of "ne" was included, but only one was occurrence per paragraph was highlighted to be annotated.

To perform the annotation we customised MMAX2, an XML-based annotation tool specifically devised to mark up coreference links (Müller

and Strube, 2006). We introduced two different links to connect each instance of "ne" to its predicate and to its antecedent. To ease the annotation process, these are visualised with different colours. A screenshot is given in Appendix A. We implemented the annotation categories described above, i.e. "ne", "antecedent", and "predicate", and made available, for each of them, all relevant attributes. MMAX2 lets developers create attributes in dependence of certain values assigned to other attributes so that, for instance, the attribute "class", whose values are "type" or "token", is activated only if the instance is annotated as "partitive". MMAX2 also lets annotate discontinuous material as part of the same entity, which came useful when annotating predicates and antecedents. The output is standoff XML.

The authors of this paper independently annotated the data, achieving a score of $K = .78$ on the classification of "ne". This is considerably lower than the agreement reported in (Nissim and Perboni, 2008), but the classification categories in our scheme are higher in number, and finer-grained.

## 4 Corpus Analysis

### 4.1 Distribution

Out of the 500 extracted instances, only two were not annotated because they were not actual occurrences of "ne". An overview of the distribution of annotated categories is given in Table 3.

Table 3: Distribution of categories in the dataset

| | | | |
|---|---|---|---|
| anaphoric | partitive | token | 19 |
| | | type | 52 |
| | | total partitive | 71 |
| | ¬partitive | external | 107 |
| | | internal | 33 |
| | | prepobj | 127 |
| | | total ¬partitive | 272 |
| | cataphoric/vague | | 45 |
| | total anaphoric | | 388 |
| non-anaphoric | | | 110 |
| invalid | | | 2 |
| total | | | 500 |

As we can see, only about one-fifth of the cases are non-anaphoric. Also, among anaphoric instances, we can observe that most commonly, "ne" is used in a purely anaphoric, *non-partitive* way, suggesting a behaviour different than the one described in the theoretical literature.

### 4.2 Anaphoric aspects

In terms of specific anaphoric relations, we observed both direct and indirect links. The observed combinations are summarised in Appendix B.

In coreference, we observe that the relationship between "ne" and its antecedent can be of all kinds apart from purely partitive, as it isn't a specific object that "ne" refers to, rather the class that the antecedent introduces (such as "pastries" in Example 7). Syntactically, the antecedent occurs as a direct object only in the partitive_token case, otherwise is always an indirect complement, usually introduced by "di" ('of') or "da" ('from').

Cases of indirect anaphora that we observed are of two sorts: (i) the reference to classes rather than objects, which is found with partitive *types*, and is syntactically akin to nominal ellipsis, as in (7); and (ii) instances of bridging (Clark, 1975; Kleiber, 1999). Within (i) we observe also cases of *other-anaphora* (Modjeska, 2002), such as (13) below.

(13) Possiamo tenere soltanto un versetto che ci accompagna però durante tutta la giornata e lo memorizziamo, lo ruminiamo, e domani **ne** prendiamo *un altro* [...]

We can only keep a verse with us during the day, and we memorise it, we think it over, and tomorrow we will take *another one*

Bridging is observed with cases of non-partitive external and especially internal features, as indeed bridging anaphors usually convey a (widely speaking) meronymic relation to their antecedent (e.g."intensità, 'intensity', in (12)). For the sake of space we cannot go into the details of the meronymic relations observed, but they have been classified according to (Cruse, 1985).

## 5 Conclusion

Actual corpus data, annotated thanks to the development of specific annotation schemes focused on the anaphoric potential of "ne", shows that the function of "ne" cannot be at all limited to a 'partitive' pronoun or as a test for certain syntactic types, as it is usually done in the theoretical literature. It also highlights several aspect of the anaphoric properties of "ne", both semantically and syntactically. We plan to exploit the dataset to develop an automatic resolution system.

## References

A. Belletti and L. Rizzi. 1981. The syntax of ne: some implications. *The Linguistic Review*, 1:117–154.

L. Burzio. 1986. *Italian Syntax: A Government-Binding Approach*. Reidel, Dordrecht.

Herbert H. Clark. 1975. Bridging. In Roger Schank and Bonnie Nash-Webber, editors, *Theoretical Issues in Natural Language Processing*. The MIT Press, Cambridge, MA.

Patrizia Cordin. 2001. Il clitico "ne". In Lorenzo Renzi, Giampaolo Salvi, and Anna Cardinaletti, editors, *Grande grammatica italiana di consultazione dell'Italiano*, vol. I. Il Mulino, Bologna.

Alan D. Cruse. 1985. *Lexical Semantics*. Cambridge University Press, Cambridge.

G. Kleiber. 1999. Associative anaphora and part-whole relationship: the condition of alienation and the principle of ontological congruence. *Journal of Pragmatics*, pages 339–362.

Anne Lobeck. 2006. Ellipsis in DP. In Martin Everaert and Henk van Riemsdijk, editors, *The Blackwell Companion to Syntax*, volume 2, pages 145–173. Blackwell, Oxford.

V. Lyding, E. Stemle, C. Borghetti, M. Brunello, S. Castagnoli, F. Dell'Orletta, H. Dittmann, A. Lenci, and V. Pirrelli. 2014. The PAISÀ corpus of italian web texts. In *Proceedings of the 9th Web as Corpus Workshop (WAC-9), in conjunction with EACL 2014*, Gothenburg, Sweden.

Natalia N. Modjeska. 2002. Lexical and grammatical role constraints in resolving other-anaphora. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Conference (DAARC 2002)*, pages 129–134, September.

Christoph Müller and Michael Strube. 2006. Multi-level annotation of linguistic data with MMAX2. In Sabine Braun et al., editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany.

Malvina Nissim and Sara Perboni. 2008. The italian particle ne: Corpus construction and analysis. In Nicoletta Calzolari et al., editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Cinzia Russi. 2008. *Italian Clitics. An empirical study*, volume 193 of *Trends in Linguistics Monograph*. Walter de Gruyter, Berlin/New York.

A. Sorace. 2000. Gradients in auxiliary selection with intransitive verbs. *Language*, 76:859–890.

## Appendix A:
## Screenshot of annotation in MMAX2

The instance of "ne" to be annotated is always enclosed in bold square brackets. Once annotated, the antecedent is in blue, the predicate in green. The red arc marks the link between anaphor and antecedent, while the yellow one links anaphor and predicate.
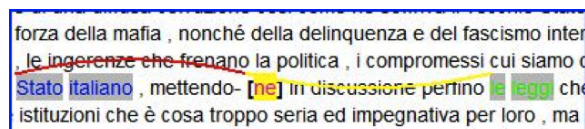


## Appendix B:
## "Ne" types and anaphoric relations

Observed configurations of types of "ne" and anaphoric relations. For a description, please refer to Section 4.2.

|  |  | direct anaphora | | |
| --- | --- | --- | --- | --- |
|  |  | d-obj | i-obj | encaps |
| partitive | type | – | – | – |
|  | token | ✓ | – | – |
| ¬partitive | prepobj | – | ✓ | ✓ |
|  | external | – | ✓ | ✓ |
|  | internal | – | ✓ | – |

|  |  | indirect anaphora | |
| --- | --- | --- | --- |
|  |  | nom ellipsis | bridging |
| partitive | type | ✓ | – |
|  | token | ✓ | – |
| ¬partitive | prepobj | – | – |
|  | external | – | ✓ |
|  | internal | – | ✓ |