

Neapolitan language documentation: a transcription model

Rodolfo Basile

Abstract

This paper discusses the challenges faced when documenting Neapolitan, a Romance language spoken in Southern and Central Italy, and proposes a model for efficiently and coherently transcribing spontaneous speech and elicited sentences, in particular solving the problem of centralized vowels. Given its status, Neapolitan is not used in official settings, which contributes to the difficulty in finding a common ground on how it is/should be written. Most native speakers of Neapolitan do not adopt a coherent spelling, leading to the appearance of several writings, which oftentimes conform phonetically to the Neapolitan dialect spoken in the speakers' area of origin. This paper aims at contributing with a reference system upon which Neapolitan language documentation can be carried out in the framework of annotated linguistic repositories such as Universal Dependencies (Nivre et al., 2016).

1 Introduction

The term Neapolitan [ISO 639-3: nap] indicates the various dialects of the language known as Napoletano-Calabrese (Ethnologue, 2021), also defined as Continental Southern Italian by Glottolog (Nordhoff and Hammarström, 2011), and called *napulitano* [napuli'ta:nə] by its native speakers. In the strict sense, the term is used to indicate the dialect of Napoletano-Calabrese namely spoken in the Southern Italian city of Naples, the cultural and political center of the Kingdom of Naples, which existed between the 13th and the 19th century. For its historical significance, the dialect spoken in Naples is often considered the correct way of speaking and writing in Neapolitan. Despite purists defending the literary Neapolitan orthography (which I will henceforth call *classical Neapolitan*, c1N), essentially based on the work of literary authors from the past century (such as Eduardo de Filippo, Salvatore di Giacomo or Ferdinando Russo to name a few) and perceived as prestigious, no real standardization of this language has ever occurred. Neapolitan rather remains, to a large extent, a “vigorously spoken” (Ryan, 2018) language. The literary language is gradually losing importance, especially because of an increase in the general level of education of the Southern Italian population, resulting in *italianization* (Cerruti, 2016: 66); in addition, many native speakers may perceive classical Neapolitan as unnatural and, sometimes, pretentious.

This does not, however, mean that nowadays the Neapolitan language is only spoken. On the contrary, the internet has contributed to the revitalization of written Neapolitan: in social media, different alternative spellings of Neapolitan words have spread, contributing to the richness and vitality of this language, and often shedding light on each dialect's phonological system. This large amount of internet folklore (Blank, 2009; Tangherlini, 2018; Kibby, 2005) is already helping to preserve Neapolitan as a dynamic linguistic entity: written sources of spontaneous internet texts can help linguists understand the sociolinguistic situation of Southern Italy in more detail, especially because of phonological variation (Eisenstein, 2013). On the other hand, this very variation, alongside the absence of a codified standard, makes it difficult for

linguists to find a coherent way to transcribe Neapolitan spontaneous speech in research and in language repositories such as Universal Dependencies (Nivre et al., 2016).

2 The sociolinguistic background

The sociolinguistic situation in Italy is of outmost complexity. As maintained by Parry (2009: 329), “an oral Italian standard is a myth”, and it is crucial to make a distinction between oral and written Italian. Spoken Italian is usually influenced, at least to an extent, by dialectal variation and quite distinguishable regional features, especially regarding pronunciation (see also Crocco, 2017). In formal situations and in writing, Italian is the only socially accepted variety (Cerruti, 2016: 64), with regional languages maintaining a lower status and being used in informal contexts: for this reason, regional languages are often inappropriately referred to as *dialects*, both by speakers and linguists. In everyday conversation, code-switching and code-mixing often occur, and the Italian ‘high’ standard is used fluidly alongside regional languages. Because of this and given the fundamentally structural differences between the two linguistic systems, Berruto (1987) talks about *dilalia* rather than *diglossia*, which would exclude the literary standard from everyday settings.

3 Neapolitan in social media

The situation of Neapolitan is indeed similar to the one described above. From what I have observed, Neapolitan is being massively used in social media, where it tends to conform to the following characteristics: I. it is mainly used in informal contexts and often restricted to humorous content (such as internet memes; see Miltner, 2018); II. favors *dilalia*, mixing with Italian just like in natural face-to-face conversational settings; III. does not conform to a specific written standard, allowing different alternative spellings to arise. These claims are, however, based on a preliminary qualitative description, and would benefit from a more thorough quantitative corpus analysis that considers variables such as age, gender, location, type of content, and educational attainment (less educated speakers might use Neapolitan also in situations where Italian would be required, or might alternatively use hypercorrected Italian constructions).

Another similar example is the analysis carried out by Jørgensen et al. (2015) about how African American Vernacular English (AAVE) is written in social media and the challenges arising. However, their case study is different in that it deals with *diglossia*, since speakers do not use Standard American English, but instead adopt a specific dialect of it, which has unique characteristics especially on the phonetic level. Moreover, AAVE is entirely based on its standard counterpart, so the more phonetic spelling (e.g. *dat* instead of *that*) can nonetheless be traced back to standardized dialects of English. This is not the case with Neapolitan.

Having developed from a different branch of Romance, Neapolitan has evolved parallelly to Italian, and not as a dialect of it (Ledgeway, 2009: 5–13; Ryan, 2018). For this reason, more phonetic transcriptions of Neapolitan cannot be traced back to their Italian counterparts, but rather compared to their old literary form, which does not have a proper standardization. Although lacking quantitative evidence, Maturi (2009) perfectly summarizes the technical differences between the literary, phonological Neapolitan orthography and the *popular* one, largely based on inconsistent phonetic spellings.

4 Language policy

At this point, the standardization of a new literary standard for Neapolitan seems unlikely to happen for essentially two reasons. First, the education system in Italy works fully and exclusively in Italian (and, in rare cases, in other major European languages, like the German schools in Rome, Milan and Genova): regional languages and dialects are not even taught as L2 and are certainly not used officially as languages of instruction. Second, Neapolitan is still perceived as ‘low’, and the italianization process is at its peak: native speakers of Neapolitan increasingly advocate for writing in ‘clean’ Italian, perceived as the ‘useful’ language. Although many southerners take pride in their Neapolitan origins and language, a number of them still sees the emerging of their identity as wrong in certain contexts, and the use of Neapolitan as impolite

or generally avoidable with strangers. The culture of the *parlare bene* (It. ‘to speak well/correctly’, in other words ‘to speak in Italian’) is still unfortunately widespread as the correct way to raise children.

Writing Neapolitan in the classical way means complying with a literary pseudo-standard which does not benefit from political status and is obviously difficult for native speakers to accept today. Revitalizing the written Neapolitan language in everyday usage should not mean introducing a prescriptive orthographic model and expecting its success, like purists tend to do, but rather letting the language live through the people who use it regularly and according to their own native speaker intuition. But what about transcribing spontaneous speech and elicited utterances? In the next part, I will provide some essential guidelines that linguists can use for these purposes. The guidelines are not intended to be an attempt to standardize Neapolitan, but rather to respect the phonetic characteristics of the major Neapolitan dialects, without having to compromise on intelligibility.

5 Transcribing Neapolitan

The transcription should be as transparent as possible. This means that it should respect dialectal variation without undermining the very essence of the language. It is also clear, for the reasons discussed above, that conforming too strictly to classical Neapolitan would reduce every dialect to a single standard that often does not mirror reality. Maturi (2009) will be used as a reference point in that it lists all the major problems of this task. Particular attention will be given to the ones that have phonological implication. It should be noted that, even though the list of problems encountered is not extensive, the primary principle one should follow is that of variation: in case a word allows for more than one spelling, the researcher should use one that mirrors the speaker’s dialectal background as transparently as possible. Examples of this purely phonetic problem are given in 5.4, 5.5 and 5.6.

5.1 Centralized final vowels and unstressed word-internal vowels

In classical Neapolitan, word-final vowels, called by Maturi (2009) ‘etymological’, essentially determine the grammatical gender of the noun or adjective or the person in verbal conjugation, and often coincide with their Italian counterparts, e.g. *chillo* that.M, *chella* that.F, *ghiamme* go.we ‘we go/let’s go’, *core* heart.M, and so on. This does not mirror the spoken language at all, where final vowels are basically always reduced to a schwa (except for Noun-Phrase internal *liaison*, see 5.2). To conform to this phonetic tendency, a solution could be that of omitting the final vowels, since it is after all also very common in contemporary spellings on social network. The aforementioned words become then *chill* that.M, *chell* that.F, *ghiamm/jamm* go.we ‘we go/let’s go’, *cor* heart.M. It is worth noting that, like in many other words, in *chill* and *chell* gender is marked by the word-internal vowel already, and that the final schwa is not always compulsory: [kil:] and [kil:ə] are both acceptable. For ambiguous cases like *bell* [b:el:ə] beautiful.M/F ‘beautiful’, information about the grammatical gender can be understood by looking at the context, just like in spontaneous speech. This solution is, however, not optimal, because semantical information is lost.

A better solution, and one that I hereby choose to adopt, is that of using a breve diacritic to indicate centralized vowels. This is used for example in Romanian, where the grapheme /ă/ indicates a schwa. The aforementioned words become hence *chillǎ* that.M, *chellǎ* that.F, *ghiammǎ/jammǎ* go.we ‘we go/let’s go’, *corǎ* heart.M.

Final vowels should be maintained in writing especially because of words containing, in word-final position, ambiguous graphemes like /g/, representing both [g] and [dʒ], or /c/, representing both [k] and [tʃ]. Example: *raggiǎ* [radʒ:ə] ‘rage’; *cacciǎ* [katʃ:ə] ‘hunt’. This is also a way to efficiently distinguish their counterparts traditionally spelled with /e/, (*ar*)*raggǎ* and *caccǎ*, which indicate the singular persons of the verbs *arraggiassǎ* ‘to get angry/to rage’ and *cacciǎ* ‘to hunt’.

The breve diacritic can be also used for unstressed sounds represented by word-internal /e/ graphemes, like in *pecché* [pə'k:e] ‘why/because’ or *védé/véré* [və're] ‘to see’.

5.2 Noun-Phrase internal liaison

Inside a NP, linking vowels are usually pronounced in Neapolitan. A NP made up of the adjective *bbellō/bbellā* ‘beautiful’ and *uaglionē/uaglionā* ‘boy’/‘girl’ can be written like *bbellu uaglionē* ‘handsome boy’ or *bbella uaglionā* ‘beautiful girl’. Note that *uaglionē* comes from the cIN versions *guaglione* ‘boy’, where the word-initial consonant has been deleted because usually not pronounced and the final vowel is reduced (although some speakers tend to mark the final reduced vowel for female, in which case the grapheme /a/ should be written). In the examples above, the final non-reduced vowel of the adjective is reinstated by liaison. Because it is not stressed, the cIN masculine vowel [o] in *bbellō* [b:el:ə] ‘beautiful’ rises to [u]. In unstressed position, Neapolitan allows only for [a], [u], [i] and [ə] (Maturi 2009).

Liaison does not, however, occur in all NPs. While reduplication of adjectives/adverbs allows for its appearance in certain cases, like *chianu chianō* ‘slowly and carefully’ (lit. ‘slow slow’), in some dialects it does not for some other ones, as in the phrasal construction *bbuonō bbuonō* ‘even if’ (lit. ‘good good’ or ‘well well’).

5.3 Word-initial gemination

Gemination is very common in word-initial position and should be maintained whenever possible. It is often the case of *intrinsically long consonants* (Maturi 2009), which sometimes even distinguish minimal pairs, e.g. *ca* (/cã) ‘which, that’ vs *ccà* ‘here’ (another contributor to this minimal pair is the fact that *ca* is never stressed, and hence its vowel is often also reduced to a schwa). Maturi (ibid.) also reminds that voiced bilabial plosives and voiced postalveolar affricates are always geminate regardless of where they occur, e.g. *bbruttō* ‘ugly’, *riebbētō* ‘debt’ (cf. cIN *debbeto*), *ggiallō* ‘yellow’ or *raggionē* ‘reason’.

Gemination can also be syntactic. For example, the consonant at the beginning of the word *votā* ‘time’ (as in ‘one time’) becomes geminated in expressions like *cierti vvotē* or ‘*e vvotē* (It. *alle volte*) ‘sometimes’. Maturi (ibid.) reminds that, sometimes, syntagmatic gemination even distinguishes minimal pairs, e.g. ‘*e figliē* vs ‘*e ffigliē* (‘the sons’ vs ‘the daughters’).

5.4 Elision

Elision is usually indicated by an apostrophe. In Neapolitan, examples are definite articles ‘*o*, ‘*a*, ‘*e* (from *lo*, *la*, *le*), as in ‘*o ppanē* ‘the bread’, ‘*a casā* ‘the house’, ‘*e ccriaturē* ‘the children’ and their correspondent preclitic direct objects, as in ‘*o ssacciō* ‘I know that’.

5.5 Vowel alternations and diphthongization

A perfect example of Neapolitan dialectal variation is vowel alternation. Compared to the Naples area, in some other provinces of Campania semi-closed back and front vowels often rise, and the researcher should take this into account when documenting the language. Examples: ‘*o canē* vs ‘*u canē* ‘the dog’, ‘*e ffemmēnē* vs ‘*i ffemmēnē* ‘the women’, *d''o/r''o* vs *d''u/r''u* ‘of the.M’.

Diphthongization should also be taken into account, as in some areas of the Casertan and Neapolitan hinterland it is common to hear *siē* [siə] instead of *sì* [si] ‘yes’.

5.6 Consonant alternations

While in cIN the grapheme /d/ is used to represent both [d] and [r], in many dialects rhotacism is more present than in others, as well as betacism ([v] turns into [b]) and gammacism ([j] turns into [g]). These features should be maintained in writing, again for preserving dialectal variety. Examples: *védé/véré* ‘to see’, *vevērē/bevērē* ‘to drink’, *jí/ghí* ‘to go’.

More examples of consonant alternations: *dē/rē* ‘of’, *quacchē/cocchē* ‘some’, *dormē/rormē* ‘sleep.3SG’, *durmì/rummi/dormē* ‘to sleep’.

6 Conclusion

This paper shows the challenges encountered in transcribing Neapolitan and suggests a model to follow to carry out language documentation in a way that is consistent in its phonology and allows, at the same time, to observe single speakers' characteristics. Neapolitan is not standardized, and the present work is not an attempt to standardization. However, it contains one important contribution, which has been previously ignored mainly because of the italianization of Neapolitan orthography and pertains to reduced vowels. The reduced vowel [ə], which in cIN is represented by the allographs /a/, /o/ and /e/, is here rendered with respectively /ǎ/, /ǒ/ and /ě/, all featuring a brevis diacritic, in an attempt to both maintain their phonological asset and indicate their common qualitative feature.

References

- Gaetano Berruto. 1987. Lingua, dialetto, diglossia, dilalia. *Romania et Slavia Adriatica*, pages 57–81.
- Trevor J. Blank. 2009. *Folklore and the Internet: Vernacular expression in a digital world*. University Press of Colorado.
- Massimo Cerruti. 2016. L'italianizzazione dei dialetti: una rassegna. *Quaderns d'Italià* 21, pages 63–74.
- Claudia Crocco. 2017. Everyone has an accent. Standard Italian and regional pronunciation. In *Towards a New Standard*, pages 89–117, de Gruyter Mouton.
- Jacob Eisenstein. 2013. Phonological factors in social media writing. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 11–19.
- Ethnologue = David M. Eberhard, Gary F. Simons and Charles D. Fennig (eds.). 2021. *Ethnologue: Languages of the World. Twenty-fourth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- Anna Jørgensen, Dirk Hovy and Anders Søgaard. 2015. Challenges of studying and processing dialects in social media. In *Proceedings of the workshop on noisy user-generated text*.
- Marjorie D. Kibby. 2005. Email forwardables: folklore in the age of the internet. *New Media & Society* 7.6, pages 770–790.
- Adam Ledgeway. 2009. *Grammatica diacronica del napoletano*. Max Niemeyer Verlag.
- Pietro Maturi. 2009. Grafie del napoletano dai classici a Facebook. *Bollettino Linguistico Campano*, 16(2).
- Kate M. Miltner. 2018. Internet memes. In *The SAGE handbook of social media* 55, pages 412–428.
- Joakim Nivre et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*.
- Sebastian Nordhoff and Harald Hammarström. 2011. Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In *First International Workshop on Linked Science 2011-In conjunction with the International Semantic Web Conference (ISWC 2011)*.
- Mair Parry. 2009. Sociolinguistics in Italy. In *The Routledge handbook of sociolinguistics around the world*, pages 347–360, Routledge.
- John M. Ryan. 2018. Embracing Neapolitan as a language which is key to the reconstruction of Early Romance. *Theory and Practice in Language Studies* 8.11, pages 1377–1386.
- Timothy R. Tangherlini. 2018. Toward a generative model of legend: Pizzas, bridges, vaccines, and witches. *Humanities*, 7(1), 1.