**On the distribution and availability of clefts: The case of regional Italian**

**Abstract**. Following the establishment of Lambrecht’s (2001) framework, and much work on the morphosyntax of (Italo)Romance clefts, it has become clear that these structures are in complementary distribution with informational focus fronting from a pan-Romance perspective. This paper investigates the availability of declarative and interrogative clefts in regional Italian, and their distribution with respect to their non-cleft equivalents, using experimental data. My claim is that, despite a well-known north-south divide in the availability of focus and clefts in Italo-Romance dialects (Cruschina 2012, Cruschina & Remberger 2017, a.o.), the *substratum* does not influence the licensing of these structure in regional Italian.

**1. Introduction**

Lambrecht (2001) observed a complementary distribution of varieties with informational focus fronting and varieties with (productive) nominal clefts. Informational focus-fronting is a syntactic strategy that shifts a constituent to the clause-initial left periphery where it receives prosodic prominence (Bianchi et al. 2015, Cruschina 2010;2012;2016, Cruschina & Remberger 2016). Unlike contrastive focus-fronting, which is used to correct a previous assertion, as in the standard Italian example in (1), informational fronting is pragmatically felicitous when used as an answer to a question that scopes over the focused element, as in the regional southern Italian example in (2).[[1]](#footnote-1)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (1) | Standard Italian (Bonan & Ledgeway in prep: 2(2)) | | | | | | | | |
|  | A: | Ho | sentito | che | hai | incontrato | Gianni… |  |  |
|  |  | have1PS | heard | that | have2PS | met | Gianni |  |  |
|  |  | ‘I heard you met Gianni…’ | | | | | | | |
|  | B: | Marco | ho | Incontrato, | (non | Gianni). |  |  |  |
|  |  | marco | have1PS | met | neg | Gianni |  |  |  |
|  |  | ‘I met Marco (not Gianni)/It was Marco I met (not Gianni).’ | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (2) | Regional southern Italian (Bonan & Ledgeway in prep: 2(3)) | | | | | | | | |
|  | A: | Mi | chiedo | chi | ha | incontrato… |  |  |  |
|  |  | refl | wonder1PS | who | has2PS | met |  |  |  |
|  |  | ‘I wonder who s/he’s met’ | | | | | | | |
|  | B: | *Gianni* | ha | incontrato! |  |  |  |  |  |
|  |  | Gianni | has3PS | met |  |  |  |  |  |
|  |  | ‘(It’s) Gianni (that) s/he met!’ | | | | | | | |

Informational focus fronting is however not felicitous cross-linguistically. Languages like standard Italian, for instance, will prefer the marked VS order in answers to questions bearing on the subject, and exclude SV (Belletti 2004, Belletti & Rizzi 2017). This is illustrated in (3).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (3) | Standard Italian | | | | | | | | |
|  | A: | Mi | chiedo | chi | abbia | incontrato… |  |  |  |
|  |  | refl | wonder1PS | who | has2PS | met |  |  |  |
|  |  | ‘I wonder who s/he’s met’ | | | | | | | |
|  | B: ?? | *Gianni* | ha | incontrato. |  |  |  |  |  |
|  |  | Gianni | has3PS | met |  |  |  |  |  |
|  | C: | Ha | incontrato | *Gianni*. |  |  |  |  |  |
|  |  | has3PS | met | Gianni |  |  |  |  |  |
|  |  | ‘(It’s) Gianni (that) s/he met!’ | | | | | | | |

Lambrecht’s (2001) observation that the varieties with focus shifting do not display productive clefting while the varieties with in situ informational foci do, which I refer to as ‘Lambrecht’s generalisation’ throughout, has been further supported in recent works on Italo-Romance. Among these are works on southern Italian dialects such as Cruschina’s (2012, 2016) investigations of the variety of Sicilian spoken in Mussomeli, and Bonan’s (2023) work on Eastern Trevisan, a dialect spoken in the continental part of the Veneto region. As predicted by Lambrecht’s generalisation, the former has informational focus fronting and very limited clefts (at least those with the regular ‘copula>focus’ ordering, cf. Bonan 2023 for additional details), while Eastern Trevisan has low informational foci, and productive clefts. Overall, there appears to be a divide between northern and southern Italo-Romance dialects, with only the former displaying productive clefting. Other noteworthy pieces of research on norther Italian and Romance clefts are Benincà 1978, Sornicola 1988, Parry 1997, Poletto & Vanelli 1997, Mathieu 1999, Valentini 2012, Poletto & Pollock 2006, Kato & Ribeiro 2009, Lobo et al. 2019, Berretta 1994, Kato & Raposo 1996, Kato & Ribeiro 2009, Dufter 2009, Roggia 2009, Cardoso & Alexandre 2013, De Cesare 2014, Garassino 2014, Belletti 2009, Valentini 2016, Sánchez Candela 2017, Destruel et al. 2019. For focus fronting, cf. Cruschina 2011;2012, Remberger 2014, Bianchi et al. 2015, Cruschina & Remberger 2017, a.o.

The ASIt, Atlante Sintattico d’Italia[[2]](#footnote-2), indirectly supports Lambrecht’s generalisation further. The atlas indeed features 1147 sentences tagged as clefts from the dialects of 225 different Italian localities. Interestingly, only 33 of these localities are from Central-Southern Italy, and 3 from Island Romance territories[[3]](#footnote-3). The disparity between the total amounts of attested clefts per area can be observed in the histogram in (4).

(4) **Geographic distribution of clefts in the ASIt**

(4) supports the prediction, based on Lambrecht’s generalisation and on the aforementioned works on clefting and focus fronting in Italo-Romance, that clefting should be more productive in northern Italian varieties than in the rest of the territory.

Noteworthily, 43 out of the total 1147 occurrences tagged as clefts do not feature in the histogram in (4) because they were actually void, while 17 had to be excluded on grounds of being instances of focus fronting wrongly classified as clefts. Interestingly, and quite predictably, these were all productions from Central-Southern varieties. Representative examples are given in (5)-(6). Note that the ASIt was compiled using questionnaires that required the informants to translate Italian sentences (listed here as ‘Target’) into a counterpart in their native dialect (‘Production’):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Instances of focus fronting classified as clefts in the ASIt** | | | | | | | | | |
| (5) | a. | Target (Italian) | | | | | | | |
|  |  | Sei | tu | che | non | vuoi | capire. |  |  |
|  |  | is | you | that | neg | want | understand |  |  |
|  |  | “It’s you who don’t want to understand” | | | | | | | |
|  | b. | Production (Troina, Sicily) | | | | | | | |
|  |  | Tu | no | vo | capiri. |  |  |  |  |
|  |  | you | neg | want | understand |  |  |  |  |
|  |  | “You don’t want to understand” | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (6) | a. | Target (Italian) | | | | | | | |
|  |  | Sei | tu | (che) | avrai | qualcosa | da | raccontarmi. |  |
|  |  | is | you | that | will.have | something | to | tell=me |  |
|  |  | “You’re the one who will have something to tell me” | | | | | | | |
|  | b. | Production (Molfetta, Apulia) | | | | | | | |
|  |  | Taw | ada | disc’ | qualche | cos’. |  |  |  |
|  |  | you | have.to | say | some | thing |  |  |  |
|  |  | “You’ll have to say something” | | | | | | | |

It is impossible to know for sure whether the reason behind the scarcity of data from Central-Southern and Island Romance varieties in the ASIt is a by-product of a sample choice made by the coordinators of the Atlas (i.e., a translation of clefts was not asked to Central and Southern speakers in the first place), or of the actual unavailability of clefts in certain varieties. Nonetheless, it is tempting to attribute the diatopic distribution of cleft data observed in (4) to Lambrecht’s generalisation. On this matter, it must be noted that only a total of 9 target sentences were submitted to the informants by the authors of the ASIt questionnaire, and of these only 1 was a DO cleft, while all other were S clefts. Since none of the translations provided for the DO cleft were from a Central, Southern or Island Romance variety, it is possible that the combined 12% of total occurrences of translated clefts from these varieties constitutes a biased result due to a sample inaccuracy, namely the fact that not enough cleft types were tested. The prediction based on Lambrecht’s generalisation and what we know about Italo-Romance clefting and focus fronting is that a larger and more composite dataset would decrease the percentage of clefts attested in Southern, Central and Island areas even further. Therefore, while the ASIt data for clefts are incomplete for the sake of this investigation, and plausibly not immune to contact phenomena that compromise the genuineness of the output of translation tasks, these nonetheless support the existence of a Pan-Romance trend in the distribution of clefts that appears quite robust.

In this paper, I study the distribution and availability of declarative and interrogative clefts in regional Italian, and their distribution with respect to prosodically-marked focalisations. My goal is to observe whether the dialectal *substratum* has an influence in the availability of clefts in different varieties of regional Italian spoken throughout the Peninsula, and in the Ticino region of Switzerland (which is Italian speaking). To do so, I present and discuss data gathered in October 2021 by means of a substantial on-line survey delivered through the Typeform platform.[[4]](#footnote-4) The working hypothesis is as follows: either there will not be any language variation in the way Italian is spoken regionally, which is highly unlikely, or the Central and Southern speakers will favour fronted prosodically-marked foci to clefts whenever possible, as they do in their dialectal *substratum*.

The paper is organised as follows. In §2, I discuss the general outline of the survey, the targeted structures, and the speakers’ profiles. In §3, I discuss the outline of Part I of the survey, which preliminarily assessed the availability of focus fronting and clefting in all varieties, and its results. In §4, I investigate the topic further by means of more complex linguistic contexts and structures.

**2. The Survey: General outline and speakers’ profiles**

The survey on which this paper is based was designed and delivered using Typeform, a user-friendly online survey tool. The choice of survey tool was made to ensure that the informants would be able to navigate the survey easily, and Typeform provides uncomplicated and UX informed templates, and ensures readability across device types (PCs, mobile phones, etc.).

The survey was distributed using social media including Facebook, Facebook Messenger, WhatsApp, and the author’s research project profile.[[5]](#footnote-5) The survey was launched at the beginning of October 2021 and discontinued about three weeks later. A total of 64 informants completed the survey in its entirety and an additional 30 informants submitted incomplete answers. Given the satisfying general participation, the data gathered from the latter were discarded to avoid possible overcomplications of the analysis.[[6]](#footnote-6)

**2.1 Targeted structures**

The survey was divided into two parts, which for convenience I shall call Part I and Part II throughout. Part I constituted a preliminary assessment of the informants’ linguistic profile. More precisely, three separate forced-choice tasks assessed whether the informants had a preference for non-cleft over cleft questions in a variety of contexts and for certain word orderings in *out-of-the-blue* informational contexts (i.e., when a new piece of information is provided in an utterance that does not constitute the answer to a question), and tested the availability of focus fronting in contrastive and corrective environments. Part II checked the preferred word orderings in answers to cleft and non-cleft questions, in *out-of-the-context* informational utterances (i.e., when the shared knowledge to which the utterance adds a piece of information can be inferred from the ongoing linguistic context), and in contrastive and corrective environments. I discuss these in greater detail in §3 and §4.

2.1.1 Macro-queries

The macro-queries of the survey were as follows:

1. whether the informants have a clear preference between cleft and non-cleft questions;
2. whether marked orderings can be observed in *out-of-the-blue* informational contexts;
3. whether marked orderings are observed in *out-of-the-context* informational contexts;
4. whether marked orderings are observed in narrow scope informational contexts, i.e., in answers to questions that bear on a precise constituent, and whether differences can be observed when the question asked is a cleft vs when the question is non-clefted;
5. whether marked orderings are observed in contrastive utterances, i.e., when the speaker corrects a previous declarative utterance;
6. whether marked orderings are observed in corrective utterances, i.e., when the speaker corrects a previous interrogative utterance.

For all these queries, I was specifically interested in discovering whether relevant variations could be observed in different varieties of regional Italian.

2.1.2 Micro-queries

Minor research questions were also investigated, which I shall call micro-queries throughout. These were meant to assess whether the following had an impact on the availability of the structures and orderings under investigation:

1. type of verb (I tested all structures with unergative, unaccusative, transitive and ditransitive verbs);
2. constituent type (subject, DO, IO, adverbial);
3. animacy of the referent (±human, ±animated). In what follows, I shall never discuss animacy, as it turned out to not play any significant role in the survey results.

It must be noted that all participants were volunteers (rf. §2.2 for relevant metadata). Since the survey was exceedingly long to be taken in full, I distributed the questions randomly among three random groups of speakers, making sure that everyone was tested for all macro-queries, but not necessarily for all micro-queries.

**2.2 Participants: Metadata and discussion**

A total of 64 complete surveys were submitted. Of all speakers, 67% reported speaking their local dialect fluently, while the remaining 33% only had a passive competence. The speakers’ distribution throughout the different regions of Italy and Switzerland were as in Table 2:

Chart, bar chart

Description automatically generated

Although, at a first glance, the Veneto speakers seem to outnumber all other regions, when we group together the speakers of the Western Romance macro-area and those of the South Romance macro-area (the last including Central Italy, Tuscany and the Sassarese speakers) we can see that the informants were actually conveniently well-distributed between the two groups, with a total of 31 speakers for Western Romance, and 33 for South Romance. This can be appreciated in the image in (7), in which I further divided the speakers into micro-areas, whose interest will become clearer in the upcoming sections:

(7) **Informants’ origins (micro-areas)**

**A picture containing text, dome, kitchen appliance

Description automatically generated**

The Gallo-Italian micro-area, which also includes the Ticino speakers, and the Venetan micro-area are the two subparts of Western Romance represented in this survey. Conversely, the Deep Southern, Neapolitan, Transitional and Tuscan and Sassarese micro-areas are considered parts of the Southern Romance macro-area. I have used the Deep Southern label to include Sicily, the southern part of Calabria, and the southern part of the Apulia region.

Given the composite origins of the informants, the variety of Italian used in the survey was neutralised to a maximum to avoid rejection of otherwise acceptable structures on the grounds of regional preferences for unrelated structures not investigated in the survey. Accordingly, the following were excluded from the investigated tokens entirely:

1. use of ‘passato prossimo’ (overwhelmingly preferred in northern Italy to refer to past events that require the present perfect or the simple past in Standard English, with no distinction made on the basis of temporal proximity) and ‘passato remoto’ (widespread in Central and Southern regions);
2. use of surnames and first names, given that the speakers of certain regions require these to be preceded by a definite article (e.g. *la Carmen* and *la Bertè*, ‘the Carmen (first name)’ and ‘the Bertè (family name)’ in regions such as the Veneto; *il Gianni* and *il Berlusconi*, ‘the Gianni’ and ‘the Berlusconi’ in regions such as Lombardy) while others and especially Southern speakers reject these entirely.
3. for the same reason as in 2, use of relationship names, therefore avoiding contrasts such as *dare a nonna* vs *dare alla nonna* (‘give to grandmother’ vs ‘give to the grandmother’), the former being unacceptable in northern Italy.

Accordingly, all tokens either utilised the simple present (*presente indicativo*) or the simple future (*futuro semplice*), featured animals or well-known characters such as *Babbo Natale* (‘Father Christmas’) or *la Befana* (a witch who brings sweeties to children at the beginning of the New Year), and avoided relationship names with the exclusion of *mia madre* (‘my mother’) and *mio padre* (‘my father’), which to the best of my knowledge are never preceded by a definite article in any variety under investigation.

**3. The Survey: PART 1**

Part I of the survey was composed of forced-choice tasks. Accordingly, the informants were given a linguistic context and 2-4 possible utterances. The informants were required to choose the utterance that they found the most natural in the provided linguistic context. I discuss the tasks and the results in what follows.

**3.1 Cleft vs non-cleft question**

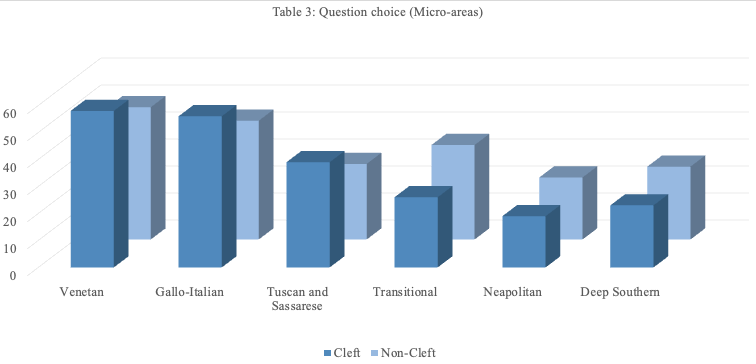
In this task, the informants were provided with 7-8 contexts, and for each of these they were asked to choose the interrogative question that sounded the most natural for them in the given context. In all cases, the informants were presented a cleft and a non-cleft interrogative. The choice and order of presentation of the contexts were randomised, as was the relative order between interrogatives. Here and throughout, the speakers were carefully instructed to always choose the option that sounded better to them, regardless of what they thought the grammatically correct option was.

A representative example for this task is reported in (8):

|  |  |  |
| --- | --- | --- |
| (8) | Context: | Guardi dalla finestra. Qualcuno sta parcheggiando la macchina in giardino. Chiedi: |
|  |  | ‘You look out of the window. Someone is parking their car in your garden. You ask:’ |
|  | Choice 1: | Chi è che arriva? |
|  |  | ‘Who’s arrived?’ Lit: ‘Who is it that has arrived?’ |
|  | Choice 2: | Chi arriva? |
|  |  | ‘Who’s arrived?’ |

3.1.1 Results for the macro-query

The general picture shows an equal distribution of cleft and non-cleft questions between macro-areas, with 114 chosen clefts and 93 non-clefts in Western Romance, and 112/95 in Southern Romance. When looking at micro-areas, a slight increase in the preference for non-clefts can be observed in Transitional, Neapolitan and Deep Southern varieties. In this respect, Tuscan and Sassarese therefore pattern with Western Romance varieties, as illustrated in Table 3:



3.1.2 Results for the micro-queries

An interesting effect in the choice between clefts and non-clefts can be observed in subject and adverbial questions, for both of which the speakers overwhelmingly prefer clefting. In the case of questions bearing on the subject, clefts were preferred 100 times out of 129. The distribution is again homogeneous between macro-areas, with 51 clefts and 11 non-clefts chosen in Western Romance, and 49/18 in Southern Romance.[[7]](#footnote-7)

Moving to micro-areas though, we observe a clear choice for clefts in Venetan and Gallo-Italian (chosen in 84% and 80% of cases, respectively), and a decrease in the use of the structure moving south (Wider Sicilian speakers preferring it to non-clefts only in 67% of total cases). This can be seen quite clearly in the histogram in Table 4:

Chart

Description automatically generated

A similar distribution is observed in adverbial questions, which showed twice the number of choices of clefts with respect to non-clefts. Again, the distribution was quite similar among macro-areas, but showed a clear preference for clefting in West Romance (especially in Gallo-Italian), and in Tuscan and Sassarese. Quite surprisingly, a balanced choice between structures was observed for the speakers from Transitional, Neapolitan and Deep Southern areas. As to verb types, no clear effect could be observed, suggesting that the first-merge position of the constituent (and particularly that of the subject) is not relevant in the choice between clefting and non-clefting.

It seems therefore reasonable to conclude that, although Western Romance varieties and Tuscan and Sassarese display a preference for cleft over non-cleft questions, in interrogatives what really influences the choice between the two structures is the type of the constituent on which the question bears: subjects and adverbials indeed trigger clefting more than DOs and IOs across the Italo-Romance territories.

**3.2 Out-of-the-blue informational utterances**

Out-of-the-blue informational utterances systematically exclude the use of clefting, whence the choice to merely test non-cleft word orderings to assess whether certain speakers may accept shifting relevant constituents in this type of contexts. Normally, shifting should not be present here unless a speaker deliberately chooses to focus on one constituent over the others.

A representative task instance is reported in (9):

|  |  |  |
| --- | --- | --- |
| (9) | Context: | Come ogni giorno, telefoni a tua madre. Le devi annunciare l'ultima novità, della quale non sa nulla. Quale delle seguenti frasi usi per farlo? |
|  |  | ‘Like every day, you call your mother. You want to tell her the latest piece of news, which she knows nothing about. Which of the following sentences do you use to do so?’ |
|  | Choice 1: | Il nostro cane abbaia in continuo. Lo dobbiamo portare da un istruttore! |
|  |  | ‘Our dog barks non-stop. We need to bring it to an instructor!’ |
|  | Choice 2: | Abbaia in continuo il nostro cane. Lo dobbiamo portare da un istruttore! |
|  |  | ‘Our dog barks non-stop. We need to bring it to an instructor!’  Lit: ‘Barks non-stop our dog.’ |

3.2.1 Results for this task

The data gathered from this task showed such an intimate link between our macro- and micro-queries that I shall discuss them together here. Overall, the data confirmed that the base ordering of constituents is the preferred one in *out-of-the-blue* contexts, with the only exception of unaccusative verbs which I discuss below.

When it comes to subject placement, the informants showed a sheer preference for SV orderings with unergative verbs, with only 1/64 of total occurrences being VS. Conversely, unaccusative verbs showed a preference for SV only in 58% of cases. Given the argument structure of unaccusative verbs, whose subject is widely acknowledged to start out as an internal argument (i.e., in the structural position normally occupied by direct objects, Burzio 1986), it is unsurprising that some speakers chose the VS ordering here. No geographic effect could be detected and, given that there are no nominal constituents other than the subject in the sentences under investigation, feature relativized minimality (Rizzi 2016, Villata et al. XXXX) effects caused by animacy were not expected either, as confirmed by our data.

**3.3 Correcting and contrastive focus shifting**

[…]

**4. The Survey: PART 2**

**5. Conclusions**

**References**

1. Following standard practice, I use small capitals to signal contrastively-focused constituents and small caps with italics to indicate informationally-focused constituents. [↑](#footnote-ref-1)
2. <http://asit.maldura.unipd.it/>, all data retrieved on 17.03.2021. [↑](#footnote-ref-2)
3. <https://www.google.com/maps/d/u/0/edit?mid=1Hy7nodW3lpHNNVSX90W5gnMwXQc6fhw6&ll=41.979953103466514%2C12.686880899999998&z=6> for an overview of the attested varieties. [↑](#footnote-ref-3)
4. <https://www.typeform.com>. [↑](#footnote-ref-4)
5. Insert link here in non-anonymous version. [↑](#footnote-ref-5)
6. All data in .xlsx format, visuals, and the various versions of this paper are publicly available in a dedicated GitHub repository: insert link here in the non-anonymous version of the paper. [↑](#footnote-ref-6)
7. Note that the total count of observed structures is not always necessarily equal in different areas as an effect of randomisation. [↑](#footnote-ref-7)