La linguistica vista dalle Alpi
Linguistic views from the Alps

STUDIA ROMANICA ET LINGUISTICA

condita a Peter Wunderli et Hans-Martin Gauger

curant Daniel Jacob, Elmar Schafroth, Edeltraud Werner,
Araceli López Serena, André Thibault et Manuela Caterina Moroni

SRL 57

PETER LANG

Ermenegildo Bidese / Jan Casalicchio /
Manuela Caterina Moroni (a cura di / edited by)

La linguistica vista dalle Alpi
Linguistic views from the Alps

Teoria, lessicografia e multilinguismo
Language Theory, Lexicography and Multilingualism

Studi in onore di Patrizia Cordin
Studies in Honor of Patrizia Cordin

PETER LANG

Contenuto/Content

Emanuela Cresti and Massimo Moneglia

LABLITA – University of Firenze

The Discourse Connector according to the Language into Act Theory: data from IPIC Italian

Abstract: The paper introduces the notion of Discourse Connector information unit (DCT) as a Discourse marker function. According to the Language into Act Theory, DCT is dedicated to signal the addressee a link between the preceding *discourse* and the *pragmatic unit* that is about to be uttered. The DCT can be identified since it is performed through a dedicated prosodic unit with specific modulation features. Research carried out on the Data-Base IPIC shows that in Italian the lexical repertory of the DCT consists of conjunctions and connectives/adverbials. The semantics of the DCT is vague and for this reason its occurrences can be deleted or substituted with other arbitrary connectives keeping the pragmatic interpretability of the utterance. The frequency of DCTs in relation to the number of utterances might appear rather low (2.81%), however its occurrences constitutes 12.2% of discourse marker units. The DCT occurs mainly at the onset of utterances, but when it occurs in a *stanza*, it signals that the spoken text goes on.

Keywords: Discourse Connector, Spontaneous Speech, Language into Act Theory, Corpus driven research, Prosody, Information Structure.

1. Introduction

This paper discusses the definition of the Discourse Connector information unit (DCT), as proposed within the Language into Act Theory (L-AcT, Cresti 2000).[1]

Generally speaking, a DCT connects different parts of a given discourse and indicates a continuation. Specifically, it signals to the addressee of a link between a part of the *discourse* that has already occurred and a *pragmatic unit* that is about to be accomplished by the speaker. The lexical repertory of the DCT consists of invariable morphemes which are traditionally classified as conjunctions and connectives/adverbials, for instance *perché* ('because'), *ma* ('but'), *però* ('but'), and *quindi* ('then'). A DCT does not enrich the semantics of the utterance because

---

[1] L-AcT is part of a pragmatic tradition of linguistic studies (Austin 1962; Biber et al. 1999; Leech 2014). See recent references for a detailed illustration (Cresti 2012; Cresti 2014; Moneglia & Raso 2014; Cresti 2018; Cresti & Moneglia 2018).

the semantic content of the connective fulfilling the DCT function is vague and may easily be deleted or substituted with another arbitrary connective without risking the pragmatic interpretability of the utterance. The DCT occurs mainly at the onset of a reference unit (see below) and, prosodically speaking, is clearly identifiable as it is always performed through a dedicated prosodic unit with specific modulation features.

The research presented in this paper is framed within the pragmatic conception of information structure that informs L-AcT and is based on the systematic analysis of the informal Italian section of the C-ORAL-ROM corpus. The latter is archived in the IPIC database (Panunzi & Gregori 2012)[2] and corresponds to 74 texts, 124,735 words, and 21,007 reference units. The issue of identifying reference units in spoken language is still topical since the majority of specialists working on spoken corpora agree that the sentence, according to its syntactic definition, cannot represent a proper basic unit. The reference units, necessary for the speech analysis, are *utterances* and *stanzas* according to L-AcT and have pragmatic definitions: the *utterance* is the counterpart to the speech act and the *stanza* corresponds to speech activity which expresses a flow of thought (Chafe 1994; Cresti 2010; Cresti 2014; Cresti & Moneglia 2018a). Both utterances and stanzas are identifiable in the speech flow through marked prosodic breaks that are performed intentionally by the speaker, and as a result are picked up in human perception. These reference units are called terminated sequences.

In the distribution examined the DCT appears to primarily occur in standalone prosodic units found at the beginning of speech reference units, referred to as *opening position*. This position is comparable with what is called the 'left periphery' of the sentence in the generative linguistics literature. Following important articles by Rizzi (1997; 2013) a theoretical debate opened around the syntactic explanation of expressions occurring at the beginning of sentences, and in recent years researchers have wondered how left periphery syntax interacts with discourse information, speech acts, and modal adverbs (Rizzi & Bocci 2017; Haegeman 2012). Corpus-based studies of the particular properties constituting the 'left periphery' in speech may thus be relevant to other domains such as the generativist and the French tradition derived from the macro-syntactic approach.

---

2   IPIC is a multilingual database composed of mini-corpora for spontaneous speech (American English, Brazilian Portuguese, Italian, Spanish) and is aimed at the comparison of information structure for these languages. IPIC presents text/sound aligned reference units that are tagged extensively in regard to their information structure (following the L-AcT model).

A central aspect in the generative literature concerns the assumption that the left periphery of the sentence can be filled not only by Topics and functionally independent syntactic heads (such as 'peripheral' adverbials), but also by Focus units. A Focus is considered to be in the left periphery of a sentence when, for instance, it depends on a Contrast unit. But given that Focus is the necessary core of a sentence, if it is considered to be in the left periphery, then we must ask in the left periphery of what entity should the Focus be evaluated?

The problem of left periphery assignment extends beyond generativist research into corpus-based approaches also. For instance, looking at the point of view of the macro-syntactic tradition (Blanche-Benveniste et al. 1990), Degand and colleagues (2014) consider the so-called left periphery a means of anchorage between the Basic unit of discourse (BDU) and the context, where units developing different functions may be enacted to signal the "macro-structure of the discourse". The authors classified the constituents occurring in the left periphery into 4 types,[3] which develop topical or meta-discursive functions and may be filled by either SV or SN or SAd. Thus, given that the main predicate (SV) of the BDU often occupies the starting position, it is not evident in what consists its left distribution.

It must be noted that, according to the C-ORAL-ROM data (Cresti & Moneglia 2005), nearly 40% of utterances are composed of simple core constituents, without any preceding Topic or Adverbial. Thus, the starting position of what can be considered a Focus represents a normal and spread distribution, that seems to exceed special cases of contrast or meta-discursive functions.

Of course, the generative approach isn't based on spontaneous speech data and considerations regarding the prosodic performance have been carried out on laboratory examples (see Rizzi & Bocci 2017), so to make any comparison possible some basic assumptions need to be made. According to L-AcT, the core of the utterance corresponds to a Comment information unit, which accomplishes the illocutionary force.[4] The *information pattern* (in accordance with L-AcT terminology) is grounded in the occurrence of the Comment, which is necessary and sufficient for performing an utterance. Then the high frequency of utterances

---

3   The criterion for identifying the basic units of the discourse (BDU) relies on the combination of syntactic and prosodic features that have been detailed in the analysis of the LOCAS-F corpus (Degand & Simon 2009).

4   On the meaning of illocution, there is, in our opinion, a fundamental misunderstanding. Generally speaking, illocution is equated and reduced to the grammaticalized types of sentences (assertive, interrogative, imperative) and not to linguistic action schemata which constitute a rich and complex repertory of conventionally defined illocutionary types (Cresti 2005; 2017; 2019; forthcoming).

composed of a single Comment – therefore with a Focus in a starting position –
cannot be surprising, and any kind of illocutionary acts can be expected. Only
Topic or other optional information units as Incipit, Allocutive, Discourse Con-
nector, could be said being on the "left-periphery of Comment", but the necessary
accomplishment of illocution conveyed by Comment cannot be said being in the
periphery of anything else.

In effect, the function of the DCT may be evaluated as occurring in the so-
called left periphery of the Comment since it must precede it. Furthermore, the
DCT's performance only has sense if it develops a specific function directed to-
ward the addressee and toward the continuation of the utterance, which exists
only if there is an illocutionary core.

Every information unit type is classified within L-AcT in terms of its informa-
tion function, prosodic performance, and distribution with respect to the neces-
sary Comment unit. Beyond the Comment, other optional information units may
participate in the information pattern and their distribution depends on that of the
Comment. These units develop two types of functions: textual, in the case where
they participate in the composition of the utterance's semantics (Topic, Parenthesis,
Appendix, Locutive Introducer), or dialogical, if they support the exchange with
the addressee and the continuation of the spoken text (Incipit, Phatic, Conative,
Allocutive, Expressive, Discourse Connector). Each information unit is performed
via a dedicated prosodic unit and the set deployed forms the *prosodic pattern* of
the utterance. The information unit types and their tags are given in Table 1. The
definition provided in L-AcT for the DCT (in bold) stresses its dialogical function
rather than any possible contribution to the content of the utterance.

Table 1: Tag set of information unit types (Moneglia & Raso 2014).

Type of Unit	Name	Tag	Definition
Textual	Comment	COM	Accomplishes the illocutionary force of the utterance.
	Topic	TOP	Identifies the domain of application for the illocutionary act expressed by the Comment.
	Appendix of Comment	APC	Integrates the text of the Comment and concludes the utterance, indicating agreement with the addressee.
	Appendix of Topic	APT	Yields a delayed integration of the information given in the Topic.

Type of Unit	Name	Tag	Definition
	Parenthesis	PAR	Inserts information into the utterance with a meta-linguistic value.
	Locutive Introducer	INT	Expresses the evidence status of the subsequent locutive space, marking a shift in the coordinates for its interpretation.
	Multiple Comment	CMM	Constitutes a chain of Comments which form an *illocutionary pattern* i.e. an action model which allows the linking of at least two illocutionary acts, for the performance of a single, conventional rhetorical effect.
	Bound Comment	COB	A sequence of weak Comments, which are produced by progressive adjunctions following the flow of thought (*Stanza*).
Dialogic	Incipit	INP	Opens the communicative channel, bearing a contrastive value and initiating a dialogic turn or an utterance.
	Conative	CNT	Pushes the listener to take part in the Dialogue or stop his uncollaborative behavior.
	Phatic	PHA	Controls the communicative channel, maintaining it. Stimulates the listener toward social cohesion.
	Allocutive	ALL	Specifies to whom the message is directed while holding their attention and forming a cohesive, empathic function.
	Expressive	EXP	Works as an emotional support, stressing the sharing of a social affiliation.
	Discourse Connector	DCT	**Connects different parts of the discourse, indicating their continuation to the addressee.**

Dialogical units may be compared to what have commonly been called Discourse Markers in the literature (Schiffrin 1987; Bazzanella 2006; Frosali 2008; Pons Borderia 2008; Raso 2014). Discourse Markers play a central role in the analysis of speech in Discourse Analysis and Interactionalist approaches, and their function may also be considered illocutionary in some cases (Barth-Weingarten et al. 2010; Couper-Kuhlen & Selting 2017). Conversely, in L-AcT each dialogical unit must be clearly distinguished from the Comment. Given that it is the latter that decides the pragmatic interpretability of the utterance, the other information units can be erased while still maintaining the utterance's pragmatic and semantic interpretability.

   Another distinction for dialogical units (and information units in general) is that their lexical fulfilment cannot be considered a condition for their functional

identification. Corpus-driven investigation allowed us to discover trends and preferences in the linguistic fulfilment of each type of information unit, however it is impossible to link a lexical entry to the accomplishment of a specific information function. Only generic morpho-syntactic preferences can be envisaged. For instance, Topics are mainly filled by NPs and PPs (Signorini 2005; Mittmann-Malvessi 2014), Comments by a VP, but also by Adverbs (Cresti 2005), Locutive Introducers by *verba dicendi*, but also proper names (Giani 2005), and so on. Moreover, the lexical entries of textual units such as the Topic and Parenthesis or of an Introducer unit can neither be foreseen nor substituted by another.

However, for dialogical units it may be observed that they show a common, recurring lexicon. For instance, proper names are employed in large part to produce Allocutive types, fixed expressions of stopping or pushing for the Conative, interjections for the Expressive, and connectives for Discourse Connectors. Each of these lexical entries remains detached from the utterance content and is used "operationally", thus we will see that the morpho-syntactic role and semantics they are foreseen to convey in writing are lost or weakened. For example, if a proper name is used as an Allocutive within an utterance, it might be substituted by a pronoun, or an adjective, or even a vocalization sufficient in drawing the attention of the addressee and bearing an emphatic effect, meaning the information function is not connected to the specific semantics or PoS of the involved lexical entry.

DCTs are coherent with this overall property of dialogic units, and, in favor of this conception, we will provide arguments based on actual spoken performances. However, examples have been found in which connectives complying with prosodic properties of DCT maintain their full semantic value, implementing the text of the utterance as textual units do. This paper will address this question and will provide detailed analyses based on actual corpus data.

2. presents the overall features of DCTs, its prosodic characteristics, and its distribution in Italian spoken corpora, as derived from IPIC Italian. In 3. the lexical repertory of the connectives fulfilling its function is presented. 4. deals with the semantic and functional definition of the DCT and presents the specific language contexts in which it maintains its full semantic content.

2. The Discourse Connector

2.1 The features of the DCT

For Discourse Connectors, it must be observed that expressions fulfilling it mostly correspond to morphemes that are traditionally classified as conjunctions and connectives e.g. *perché, però, ma, e, o, quindi*, (respectively: 'because/

why', 'however', 'but/anyway', 'and', 'or', 'then').[5] Ordinarily these morphemes can coordinate, subordinate and compose subsequent clauses or phrases within a syntactic configuration. Their common Italian usage is demonstrable through sentences such as the following:

(1a)   non so perché lo ha fatto
      'I don't know why he did it'

(1b)   è caro ma funziona
      'it's expensive but works'

(1c)   insegna matematica e aiuta gli studenti
      'he teaches mathematics and helps the students'

(1d)   è andato a piedi, quindi arriverà tardi
      'he went on foot, therefore he will arrive late'

In (1a)-(1d) the morpheme mainly links a main clause to a subsequent clause, specifying a logical relation developed within compositional phrasing. If the conjunctions and connectives in the examples are deleted, the derived phrasing is meaningless or unacceptable as a sentence:

(1a')   * non so lo ha fatto
      'I don't know he did it'

(1b')   * è caro funziona
      'it's expensive it works'

(1c')   * insegna matematica aiuta gli studenti
      'he teaches mathematics he helps the students'

(1d')   * è andato a piedi arriverà tardi
      'he went on foot he will arrive late'

It may be verified, however, that if the same morphemes are realized in speech with the specific distribution and the prosodic properties of a DCT they show different behavior: they are not syntactically compositional with respect to any subsequent clause or phrase.[6] Let us look at some spoken examples:

---

5   A more complete list will be given in section 3.

6   The compositional nature of the DCT is assumed in the linguistic literature (Bazzanella 2006), however this is not corroborated by corpus-based investigations or prosodic correlations. Ferrari, exemplifying in detail the set of expressions that are considered as connectives in writing, assumes that their functions exceed the syntactic dependency and the simple linkage of clauses, in favour of logic and argumentative relations participating in the text's architecture (Ferrari 2015).

(2) * GIU: eh /^{PHA} quelle eran < belle > //^{COM}
'those were beautiful'
* PAL: [<] < **ma** > /^{DCT} l' eran belle proprio /^{COM} eh //^{PHA} [ifamcv19,7]
'but, they were really beautiful'
%ill: confirmation[7]

(3) MAR: lui deve indovinare /^{COM} capito ?^{PHA} capito?^{COM}
'He must guess // (have you) understood, understood ?'
*VIT: mh //^{COM}
*MAR: **quindi** /^{DCT} metti tre neri //^{COM} [ifamcv09,168]
'so, put three black (pieces)'
%ill: instruction

(4) *GIU: quando gli se ne dà /^{TOP} segnare a tale //^{COM} dopo quando arriva un pezzetto /^{TOP}
tu vai a richiedegnene …^{COM} **perché** /^{DCT} tu te ne ricordi dopo < quando tu vai a
cercarlo /^{SCA} anche anni > //^{COM} [ifamcv19,69]
'when you give some to them / you assign this to that guy // when one piece will arrive /
you go and ask for it … in fact / you figure it out later when you try to find it //'
%ill: [1] instruction; [2] expression of obviousness; [3] explanation

From a grammatical perspective these expressions must generally be evaluated as
"connectives", since they do not have syntactic scope such as when participating in
propositions like those of (1a) – (1d). In principle, these connectives could even
be omitted without affecting the accomplishment of the utterance. To illustrate:

(2a) * GIU: eh /^{PHA} quelle eran < belle > //^{COM8}
'those were beautifull'
* PAL: [<] < ma > /^{DCT} l' eran belle proprio /^{COM} eh //^{PHA} [ifamcv19,7]
'But they were really beautiful'
%ill: confirmation

(3a) *MAR: lui deve indovinare //^{COM} capito ?^{PHA} capito?^{COM}
'He must guess/ (have you) understood, understood ?'
*VIT: mh //^{COM}
*MAR: **quindi** /^{DCT} metti tre neri //^{COM} [ifamcv09,168]
'So, put the three black (pieces)'
%ill: instruction

(4a) *GIU: quando gli se ne dà /[TOP] segnare a tale //[COM] dopo quando arriva un pezzetto /[TOP] tu vai a richiedegnene …[COM] **perché** /[DCT] tu te ne ricordi dopo < quando tu vai a cercarlo /[SCA] anche anni > //[COM] [ifamcv19,69]

'when you give some to them / you assign this to that guy // when one piece will arrive / you go and ask for it … in fact / you figure it out later when you try to find it //'

%ill: [1] instruction; [2] expression of obviousness; [3] explanation

In the previous cases the pragmatic value of the utterances remains unchanged, even though the DCT has been removed. Furthermore, the specific grammatical meaning of each lexical entry, which is fundamental to a syntactic constituent, is vague when the entry is left in and to some extent each expression might be substituted with one of the others, causing little or no change in the semantic value of the utterance. This is clear in examples (2), (3), and (4). The DCTs show low semantic specificity and in both (2) (*ma* 'but') and (4) (*perché* 'because') the expressions could be swapped for one another or for any generic connective and still maintain the same information.

With respect to information flow, the DCTs in examples (2) through (4) serve to connect 'that which is about to be' (usually a Comment unit accomplishing an illocutionary act) with that preceding ('that which already exists'). We call the latter *discourse,* and it can be interpreted as both *context* and *co-text*, which is to say the shared situation, the previous addressee's turn, the previous speaker's utterance, or other specific cases which we shall see.[9]

The DCT is traditionally considered to take place at the starting point of a reference unit (*utterance* or *stanza*), which we call the *opening position,* demonstrated in the previous examples. However, within stanzas the most significant position is the 'internal' one. A systematic investigation of data indeed shows that DCTs often occur within a stanza, opening and connecting a sub-unit, a simple Bound Comment, or an entire sub-pattern[10] with an anaphorical relation to the previous one/s. (5) is an example, with a DCT connecting sub-patterns within a *Stanza*:

---

9   From a syntactic point of view, the chunk or word developing the DCT function behaves no differently to the other information units within L-AcT i.e. it is a semantic/syntactic island (Cresti 2014).

10 A *Stanza* is composed of one or many *Bound Comments*, each of which in turn may record a local information pattern (sub-pattern).

(5) *TIZ: < lei c' ha > da tenergli l'amministrazione /[COB] **quindi** /[DCT] dice /[INT] a casa /[TOP_r]
     tanto con Federico non fo' niente //[COM-r] [ifamdl08,29,]
     'she must keep the administration for him / then / she goes / home / anyway with
     Federico I can't do anything //'
     %ill: [1] description; [2] reported speech

In terms of prosodic performance, the DCT corresponds to a modulated prosodic
unit, that may be longer than the average duration of components syllables, and it
may occur in formal variants. The DCT's prosodic execution makes it perceptually
prominent, supporting the information function of a bridge. The prosody of the
DCT corresponds to intentional prosodic movements and must be distinguished
from hesitations (which have no functional value) and scanning units (which form
a composition together with subsequent elements by definition).[11]

In conclusion, the DCT neither logically binds clauses and phrases in a syntac-
tic structure nor serves to segment expressions. It works as an information unit
with a specific dialogic function, signaling to the addressee that what has occurred
beforehand in the discourse will relate to pragmatic units that are about to occur.

2.2 The Prosody of the DCT

DCT units are performed via a dedicated prosodic unit called the *connector* in
L-AcT.[12] Its primary prosodic characteristic is the *modulated* performance, as in
the following example:

(3') *MAR: quindi /[DCT] metti i tre neri //[COM]
      'So, put the three black (pieces)'
      %ill: instruction   [ifamcv09,168]

---

11 A chunk of speech between two prosodic breaks may not develop an information
    function in cases where the prosody simply divides into parts, 'scans', an information
    unit that is too long to be performed as one prosodic unit. These *Scanning units* (SCA)
    fall into the list of textual information units, since only textual units such as the Com-
    ment, Topic, Parenthesis, and (rarely) the Appendix may be scanned. In the Romance
    languages, scanning is almost always on the left, i.e. only the last part of the scanned
    unit conveys the information function in question.

12 The f0 track has been processed through WinPitch.

Fig. 1: f0 track of example 3'. Modulated falling.
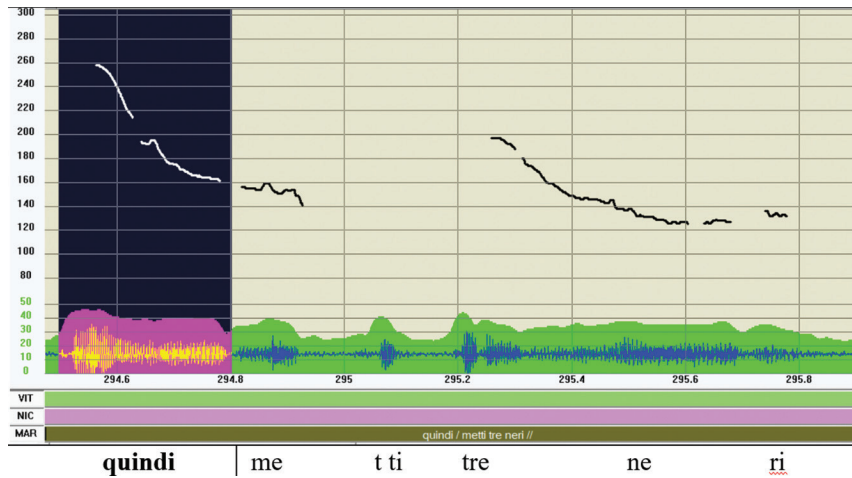
The DCT can be further described according to the L-AcT system (Cresti and Moneglia 2018) via the following prosodic parameters:

– Composition: no composition (-preparation, – tail)
– Duration: medium, with possible lengthening
– Speed: medium or high speed
– Perceptual value: high
– Intensity: standard
– Spectrogram: medium phonetic performance

Let's look at example (6), which demonstrates a lengthening of the connector unit (in bold), with three syllables lasting 1400 ms.

(6) *SIM: **inoltre** /[DCT] mi dovresti togliere una curiosità //[COM]
'moreover / you should explain something to me //'
%ill: request

Fig. 2: f0 tracks for example (6). Lengthened modulated falling pattern.

inoltre | mi dovresti togliere una curiosità
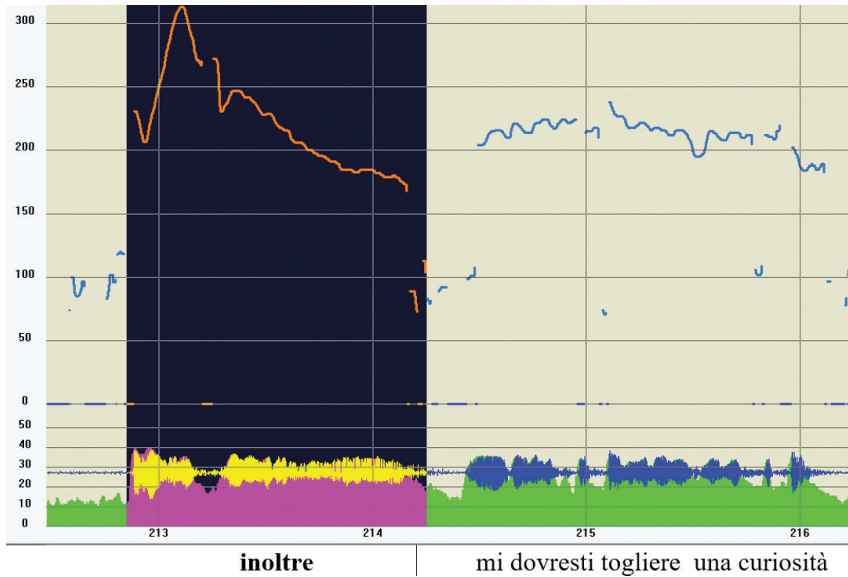
Our corpus-based investigations revealed essentially three possible variations in the prosodic form of the DCT, all of which maintain its modulated aspect:

1: **flat modulated** f0 movement, with mid/low f0 value[13]

(7) *MAR: (per)ché allora /[DCT] l'aspetto /[TOP]   dico /[PAR] ma com'è ?[COM] [ifamdl20, 36]
    'cause then / her appearance / I say / but how is it?'
    %ill: reported speech

---

13 The flat form of the connector can occur in both the utterance's opening and within
    stanzas.

Fig. 3: f0 track for example (7). Flat modulated.

	ché	**allora**		l' as	pet to	/	dico	/	co s'é
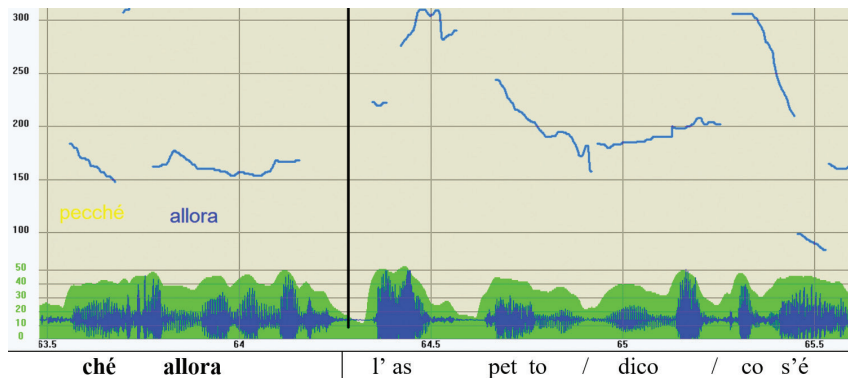
2: **modulated raising** f0 movement, ending in a mid f0 value[14]

> (8)   *DAN: e l'ha prese anche qui /[CMM] **però** /[DCT] voglio dire… [CMM]
> 'and he had taken (them) also from here / but / I mean …'
> %ill: evidentiality assertion + expression of obviousness   [ifamcv15]

Fig. 4: f0 track for example (8). Modulated raising.

e l'ha	prese	an	che	qui	**però**	voglio	dire

---
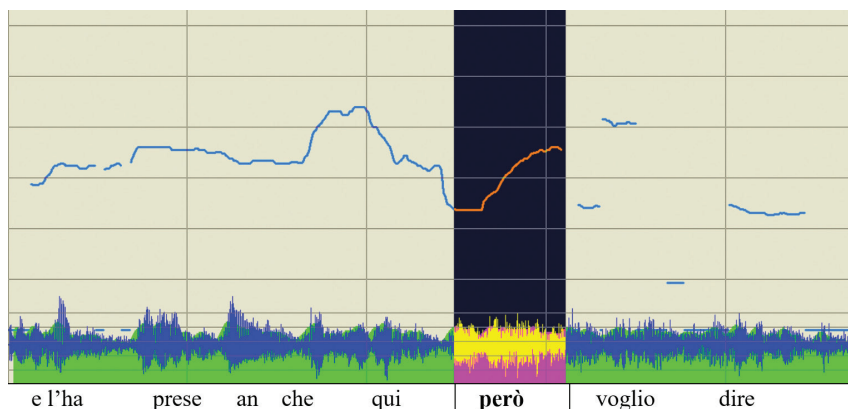
14 Typically found within utterances and connecting illocutionary patterns (Adversative
pattern), as in (8).

3: **modulated falling** f0 movement, starting with a high f0 value and ending in low f0 value.

Falling contours have been found in opening position in both utterances and stanzas. See examples (9) and (10), containing two occurrences for each distribution type.

(9) *ELA: **quindi** /^{DCT} scusi /^{CNT} intanto i [/1]^{EMP} i tre milioni l' anno / ^{TOP} **quindi** /^{DCT} mensili /^{TOP} quanto < vengano > ?^{COM}
'then / excuse me / in the meantime / three million a year / then / monthly / how much it is?'
%ill: question   [ipubdl02]

Fig. 5: f0 track for example (9). Modulated falling.

| **q**uindi | scusi intanto i tre   milio ni l'anno | **quindi** | mensili quanto vengo no |
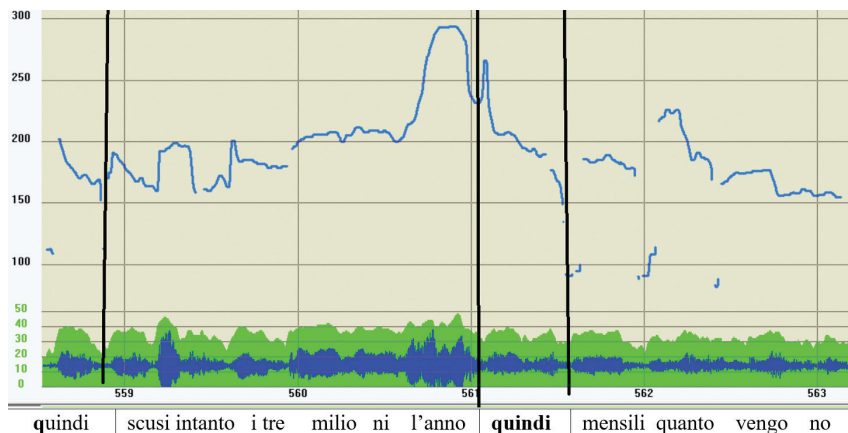
(10) *CLA: sa' /^{INP} **comunque** /^{DCT} si sentì /^{SCA} il colpo di fucile /^{COB} **sicché** /^{DCT} voglio dire /^{PAR} tanto lontano un'era //^{COM}   [ifammn03]
'you know / however / we heard / a rifle shot / so / I mean / it couldn't be so far //'
%ill: narration

Fig. 6: f0 track for example (10). Modulated falling.

| sa | comunque | si sentì il colpo di fucile | sicché | voglio dire tanto lontano un' era |
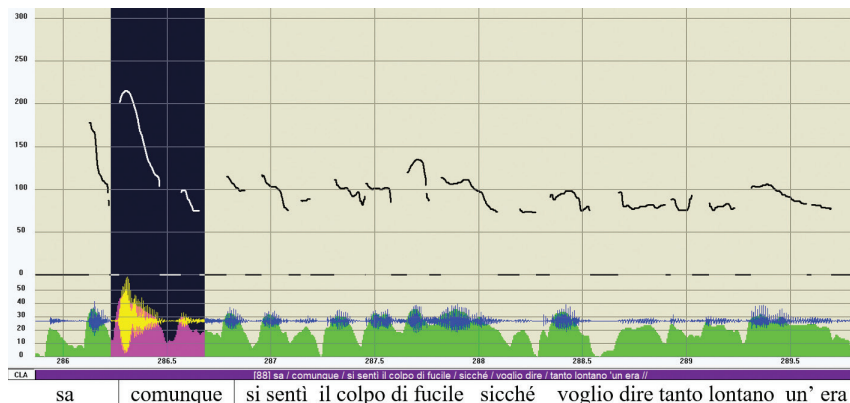
In conclusion, beyond the previous formal variants, the DCT is clearly identifiable from a prosodic point of view since it is always performed by a dedicated prosodic unit of the *connector* type and it demonstrates specific features of modulation and possible lengthening. The connector unit is perceptually marked from a prosodic point of view.

2.3 Quantitative data and the distribution of DCTs

IPIC Italian shows that the frequency of DCTs in relation to the number of reference units (587) might appear rather low: only 2.81% of reference units contain a DCT. However, this value must be considered in relation to the overall number of dialogical units, which come to 5071, among which DCTs represent 12.2%. Given the reduced number of stanzas with respect to utterances, 53% of DCTs appear within utterances and 47% in Stanzas (see Table 3). But, what is more interesting in the frame of the present research is the proportion of DCTs in relation to the different types of reference units: less than 2% of utterances have a DCT, compared to 13% for stanzas (see Table 2).[15] So, DCTs may be considered a typical strategy in the continuation of stanzas.

---

15   Table 2 and Table 3 also detail the absolute number of DCTs in IPIC (619), not just the number of reference units presenting a DCT (587); some reference units (typically stanzas) may record more than one DCT. This datum is coherent with the trend of greater dialogical unit usage in stanzas (27%) than in utterances (8.38%).

Table 2: Reference Units in IPIC with a DCT

IPIC Reference units 20844	Utterances 18810	Stanzas 2034
Reference units with DCT 587 (2.81%)	323 (1.71%)	264 (13%)

Table 3: DCTs for Utterances and Stanzas

	Utterances 18810	Stanzas 2034
Number of DCTs 619	328 (53%)	291 (47%)

Looking at the data more in-depth, a rather complex distributional situation is visible. It is summarized as follows:

a) Opening a turn:
Turns may begin with a DCT, as examples (2') and (3') show (rare);

b) Opening a reference unit (utterance, stanza):
DCTs mostly occur at the "opening" of reference units (59.5%), including, in this distribution, cases in which the DCT occurs after another dialogical unit such as the Incipit or Phatic (e.g. the first DCT in (10));

c) Opening and iterated:
Although rare, DCTs may be iterated (as in (7));

d) Opening a Bound Comment within a stanza:
With regard to stanzas, the DCTs mostly open a COB (sub-pattern) within the stanza (e.g. the second DCT in example 10);

e) Introducing a CMM within specific illocutionary patterns:
A DCT may connect the second or third CMM composing an illocutionary pattern (such as in example (8));

f) Internal distribution:
Within an utterance a DCT may connect a long TOP, or a TOP followed by a PAR, to an entire and new Topic-Comment information pattern (as in example (9)). In this case the first Topic works as a kind of co-text.

In conclusion, the distribution of DCTs is coherent with its dialogical definition, since it never connects simple textual information units (such as two Topics), but can only introduce units with pragmatic value, connecting those units to the discourse. Table 4 gives a summary of the quantitative data.

Table 4: Position of the DCT in Utterances and Stanzas

	UTTERANCES with DCT 328	**STANZAS with DCT 291**
OPENING **59.5%**	292 (89%)	76 (27%)
WITHIN **40.5%**	36 (11%) (Illoc pattern + Topic)	215 (73%) (+ Illoc pattern)
TOTAL DCT	619	

Beyond the 'opening' DCTs at the beginning of an utterance or stanza (59.5%), which is in the relative majority, the 'internal' DCTs seem to reveal interesting qualities. They demonstrate primarily the connection of a sub-pattern within a stanza but may also connect CMMs within an illocutionary pattern and, rarely, heavy Topics to a Topic-Comment pattern. We will see in section 4 that 'internal' DCTs may, in some cases, maintain a semantic value and, as a result, have the dialogical character of their information function questioned.

3. DCT Lexicon

Within L-AcT, a prosodic pattern is assumed to perform an information pattern and constitutes the interface between the illocutionary and the locutionary act. Thus, the prosodic performance of a DCT allows us to assign a functional role to a lexical entry. The lexical entries for DCTs are invariable morphemes that may be classified according to grammatical tradition as:

– conjunctions /logic operator: *e, o, ma, che*
– connectives /adverbials: *perché, però, per cui, quindi, allora, sicché, anche se, poi*

More specifically, the following is the list of all expressions fulfilling the DCT function in IPIC Italian, identified according to prosodic, distributional, and functional criteria. We classified the lexical repertory, distinguishing the entries based on their 'opening' or 'internal' distribution and their occurrence in an utterance or stanza, and tried to discover possibly significant lexical correlations. The list is as follows:

   Utterance, opening distribution: *Allora*, *almeno*, *anche perché*, *anche se*, *anzi*, *appunto*, *bensì*, *che*, *comunque*, *dunque*, *e*, *e allora*, *e comunque*, *e così*, *e dopo*, *e invece*, *e più*, *e poi*, *e quindi*, *in quanto*, *infatti*, *inoltre*, *intanto*, *invece*, *ma*, *mentre*, *o*, *oppure*, *ossia*, *per cui*, *per cui ecco*, *perché*, *perché allora*, *perché comunque*, *perché tanto*, *però*, *poi*, *quindi*, *sennò*, *sicché*, *tanto*.

Other distributions within the utterance: *Allora, comunque, e, e allora, ma, o, o sennò, perché, per cui, però, oppure, poi, quindi, sicché.*

Stanza, opening distribution: *Allora, anche perché, che, comunque, e, e allora, e infatti, e invece, e quindi, in quanto, ma, o, oppure, perché, per cui, perché invece, però, poi, quindi, sicché, tant' è che.*

Other distributions within the stanza: *Allora, che, che appunto, che poi, cioè, comunque, e, e allora, e poi, e quindi, in quanto, invece, ma, ma poi, o, oppure, per cui, perché, perché insomma, però, poi, quindi, sicché, sicché dice.*

Table 5 gives a lexical summary focusing on the lexical entries with at least 4 occurrences.

Table 5: Distribution of lexical entries with at least 4 occurrences

Opening utterance	Within utterance (+ Illoc Pattern)	Opening stanza	Within stanza (+ Illoc Pattern)
74 <u>perché</u>	8 ma	27 <u>perché</u>	60 <u>perché</u>
54 <u>però</u>	7 però	7 quindi	35 e
18 e	5 quindi	4 che	25 <u>però</u>
23 quindi /e quindi	3 e allora	4 sicché	25 quindi / e quindi
25 allora / e allora		4 per cui	6 poi
13 comunque		4 <u>però</u>	6 per cui
12 ma			5 sicché
6 sicché			5 allora
6 per cui			14 ma
4 che			4 o

The lexical homogeneity of DCTs is clear, since the unit is filled with entries belonging to connectives and conjunctions only. No other part of speech occurs that satisfies the prosodic and informational criteria.

No relevant differences are envisaged among the connectives occurring in utterances and stanzas beyond the fact that in the utterance's opening position (which has the higher frequency) a wider choice of entries may be found. For instance, *perché* ('because') is the most frequent in both.[16] For opening position within illocutionary patterns the most common are the two Italian variants of the adversative connective, *però* and *ma* ('but'), probably connected to the high frequency of the Adversative illocutionary pattern.

---

16 For detailed research on the usage of *perché* in spoken Italian see Acciardi (2010).

As already seen, the DCT bears little semantic content and its lexical filling may be easily substituted by other connectives. Moreover, just like all information units other than the Comment, their omission does not affect the pragmatic interpretability of the utterance. We would like to underline that the locutive fulfilment of a DCT constitutes an *island*, which is to say it is syntactically isolated and doesn't develop any kind of regency or coordination with the next linguistic chunk (Cresti 2014). Connectives developing a DCT function have no syntactic scope and so are neither coordinating nor subordinating conjunctions. In contrast, as we have seen in examples (1a')-(1d'), conjunctions and connectives cannot be deleted without compromising textual compositional satisfaction.

4. The dialogical function of DCTs

4.1 DCT as a dialogical function

In accordance with its general definition the DCT connects different parts of the discourse, indicating continuity to the addressee. As we have seen in earlier discussion, the unit specifically connects the discourse to a subsequent unit with a pragmatic value. The relation's distribution in the corpus was found to be as follows:

– in the opening of a turn, a DCT connects an utterance or a stanza to the discourse (context and dialogue)
– in opening an utterance or a stanza within a turn, it connects these reference units to the previous ones (co-text)
– within a stanza it connects a Bound Comment (or an entire sub-pattern) to the previous ones (early part of the stanza's text)
– introducing the second or third CMM within an illocutionary pattern, it connects the CMM to the previous one(s) (early part of an illocutionary model)
– connecting a long Topic to a Topic-Comment pattern (early part of an information pattern)

Thus, the units that are connected by a DCT are pragmatically characterized (utterance, stanza, Bound Comment, CMM, Topic-Comment) while their connection leads to the continuation of the spoken text at different levels (dialogue, co-text, text of the stanza, illocutionary model, information pattern of the utterance).

Within L-AcT, the DCT is found among the dialogical functions, but this assumption may be questioned. In principle, the Discourse Connector could also be considered a Textual function. The reasons for keeping the original choice follow.

Beyond the Comment, Textual units (Topic, Parenthesis and Locutive Introducer)[17] contribute to the information pattern of the utterance, combining information units with specific roles relative to the Comment (field of the force's application, speaker's modal appreciation, introduction of reported speech). Conversely, dialogical units are aimed at the addressee and their goal is in ensuring the continuation of the spoken exchange. DCTs rather than combining information units within an utterance, as Textual units do, seek to signal to the addressee that there is a link between some part of the discourse and a pragmatic unit that is about to be accomplished by the speaker. From this point of view, DCTs can be placed among the dialogical functions.

Furthermore, as we saw in 3., DCTs correspond to a restricted set of lexical items and this property mirrors the dialogical units, which have their own specific lexical repertories (Frosali 2008; Raso 2014). More specifically, it's worth underlining from a semantic perspective that each Textual unit adds content to the Comment and is characterized by an independent modality (Cresti 2014). Neither of these features appears to be present for DCTs:

a) connectives fulfilling the DCT function have only vague semantic content and can be substituted or erased; these lexical entries don't enrich the utterance's semantics;

b) DCTs, being connectives with vague content, don't correspond to semantic *scenes* and by consequence do not have modality.

In examples (2), (3), (4), and (5) we have already witnessed the ability to substitute a connective with another or to even delete it, when the latter occurs in opening position for an utterance or even when introducing a COB within a stanza. In conclusion, we have functional, prosodic, and semantic evidence that the DCT belongs to the dialogical type.

4.2 Problematic cases

The dialogical function of the DCT is grounded in the fact that the semantics of connectives with the DCT function appear vague, and the lexical entry could be substituted with another or even suppressed without affecting the pragmatic interpretability of the reference unit. Corpus-based investigations have shown that this is not the case in only two distributional contexts:

---

17 Appendixes of Topic and Comment are not considered since they are fundamentally locutive integrations.

– when, in specific illocutionary patterns, the connective links a second CMM;
– when the connective is introducing a Bound Comment characterized by a strong change of modality within a stanza.

In these two cases the information function of the connectives, which in effect resemble DCTs in terms of lexicon, prosody, and distribution, may be questioned.

Illocutionary patterns (*reinforcement, list, adversative, comparative, alternative, necessary relation*) have a high probability of occurrence in speech (IPIC records more than one thousand instances). Many of them are adversative illocutionary patterns, such as example (8) and in this case the DCT cannot be erased since its presence is necessary to express the adversative value, which would otherwise be lost.

Let's look at (8), connecting two Multiple Comments (CMM) within an *adversative* illocutionary pattern, in detail.

(8) *DAN: e l'ha prese anche qui /^{CMM} però /^{DCT} voglio dire… ^{CMM}
'and he had taken (them) also from here / but / I want to say…'
%ill: assertion of evidence + expression of obviousness   [ifamcv15]

In (8), the adversative connection between the first CMM (accomplishing an assertion of evidence) and what the speaker is about to introduce (an expression of obviousness) is supported by the connective *però* ('but'). The latter cannot be substituted by any other connective, such as *and*, *or*, *because,* since the substitution would cause a change in meaning while the adversative pattern would be lost.

In principle, the connective could be suppressed, as in (8a). However, the deletion of the connective will cause the onset of two independent utterances with a different pragmatic interpretation (assertion of evidence, expression of reproach) missing the adversative effect:

(8a) *DAN: e l'ha prese anche qui //^{COM} ~~però /^{DCT}~~ voglio dire… ^{COM}
'and he had taken (them) also from here // ~~but /~~ I would object…'
%ill: [1] evidentiality assertion; [2] expression of reproach   [ifamcv15]

Therefore, in this case, the semantic contribution of the connective is not vague. Moreover, we may also argue that it conveys an epistemic attitude.

A comparable scenario is offered by other illocutionary patterns that are linked by "not easily erasable" connectives. Let's take (11), an example of an illocutionary pattern of *necessary relation* bound by a long and flat *perché*.

(11) *CLA: è stato un disastro tornare /^{CMM} **perché** /^{DCT} gli si è rotto la pila //^{CMM}
It was a real disaster coming back / reason why / the stack broke //
[ifammn03]
%ill: expression of negative evaluation + explanation

In (11) the deletion of the connective will cause the creation of two independent utterances and the loss of the illocutionary pattern of necessary relation. Consequently, the semantic contribution of the connective is significant.

In cases as (8) and (11) the morpheme seems not to develop anymore its dialogical information function of Discourse Connector. The speaker conceives the first and the second CMM in a unified manner following an illocutionary pattern to create a certain "rhetorical" effect: that adversative or that of a necessary relation. Within this logical level the morpheme seems to carry on its grammatical function of coordinating or subordinating conjunction.

Comparing (8) and (11) with other illocutionary patterns connected by a DCT in which the connective conveys a vague explicative relation, the DCT shows different properties. (12) is an example of an illocutive pattern of reinforcement:

(12) *MAX: dai più gas //COM
      'speed up the car'
      %ill: order
      *MAR: eh/PHA più di così no /CMM perché /DCT siamo al massimo // CMM
      'eh/ not more than this / cause / it is the maximum (speed) //
      %ill: refusal + protest   [ifamdl19]

The recording was carried out during a driving lesson which presented a dangerous situation, and the illocutionary value of (12) cannot be fully appreciated without listening to the prosodic performance. However, the connective – which also has a short duration – may be erased without losing the semantic relation of reinforcement between the two CMMs.

With regard to connectives in the opening position of a Bound Comment, it has been observed that when a long stanza is made up of many Bound Comments, each one corresponds to a semantic island (ideas, according to Chafe 1994). Even if they belong to a single flow of thought and are enacted through the same mental activity, a Bound Comment may be characterized by a change in modality. For instance, in (13) the occurrence of the modal change in the stanza corresponds to a sub-pattern composed of a DCT followed by the final COM:

(13) *ANG: a differenza invece delle murature piene /COB che isola le piastre museale /COB
      perché insomma /DCT bisogna avere un po' di tranquillità // COM   [ifamcv16]
      'unlikely of walls made of bricks / that isolate the museal platform / because in short / we must stay secure'
      %ill: [1] description; [2] explication; [3] conclusion

In (13), the DCT clarifies that the need of 'staying secure' is the specific reason for having "walls made of bricks". In this case, the relation between the second

COB and the final COM cannot be recovered if the DCT is erased. If the DCT is deleted, the text of the stanza will be lost with the onset of a meaningless discourse.

Conversely, in example (14) even if the same connective *perché insomma* ('because in short') connects one COB to the final COM, it does not really specify a causal relation. The Comment, indeed, adds only supplementary information to the situation reported by the speaker, which is a narrative composed of COBs with a common modality, and the DCT may be erased without losing any information.

(14) *EMI: un po' < sbandatello /^{COB} si presentò davanti a tutti quei riccastri /^{COB} perché insomma /^{DCT} la cena costava ottanta novantamilalire //^{COM} [ifamcv06]

'(he was) a dropout guy / he showed up in front of all that new rich people / 'cause in short / the dinner cost eighty ninety thousand lire //'

'%ill: [1] narration; [2] narration; [3] narration

Thus, for stanzas the semantic value of the connective fulfilling a DCT must be verified on a case-by-case basis with respect to the onset of a modal change. Finally, the question regarding the nature of morphemes, that open certain illocutionary models (adversative, necessary relation) or represent a point of modal change within stanzas, remains open.

5. Conclusions

The definition and identification of the Discourse Connector information unit (DCT) is confirmed on the basis of Italian DB-PIC data. Generally speaking, a DCT connects different parts of the discourse and indicates its continuation, signaling to the addressee that there is a link between part of the discourse that has already occurred and a pragmatic unit that is about to be accomplished by the speaker.

The lexical repertory of the DCT consists of invariable morphemes that can be classified according to grammatical tradition as conjunctions and connectives/adverbials. However, the DCT cannot be confused with the role of any conjunction or connective, since it doesn't develop any compositional relation.

Furthermore, the DCT must also be distinguished from textual units, since it does not combine information units within an utterance as textual units do and does not enrich the utterance's semantics. The semantic content of the connectives fulfilling the DCT function is vague, and they may easily be deleted or substituted with another arbitrary connective without risking the pragmatic interpretability of the utterance.

The DCT's frequency is similar to that of other information units, representing 2.81% of all dialogical units. It's worth noting that even if its absolute quantity

concerns more utterances (328) than stanzas (291), their percentage across utterances comes to 8.38% while for stanzas it is 27%. Thus, DCTs appear to be a typical strategy employed in the continuation of stanzas.

The DCT occurs mainly in the opening position of a reference unit (utterance or stanza), however, for stanzas the more significant position is the 'internal' one concerning the connection of sub-patterns. It is worth noting that DCTs may also introduce a CMM within an illocutionary pattern.

Only in rare cases does a DCT connect a long topicalization to a Comment within an utterance.

Prosodically speaking, the DCT is clearly identifiable since it is always performed by a dedicated prosodic unit of the *connector* type with specific modulation features and may present three formal variations. Furthermore, the DCT's execution together with its distribution make it perceptually marked, supporting the information function of a bridge with respect to what occurs beforehand.

Specific cases that contradict the above description and characteristics have been found. Connectives in the opening position of a CMM within specific illocutionary patterns, and before Bound Comments that signal a significant change of modality within a stanza, may behave differently from a semantic perspective and are of semantic and modal relevance.

References

Acciardi, Daniela (2010). *Funzioni connettivali nella lingua italiana scritta e parlata. Analisi corpus-based del connettivo 'perché'*. PhD Thesis, University of Florence.

Austin, John Langshaw (1962). *How to Do Things with Words*. Oxford: Oxford University Press.

Barth-Weingarten, Dagmar, Elisabeth Reber & Margret Selting (eds.), (2010). *Prosody in Interaction*. Amsterdam: Benjamins.

Bazzanella, Carla (2006). Discourse Markers in Italian. Toward a 'compositional' meaning', in Kerstin Fischer (ed.), *Approaches to discourse particles*. Amsterdam: Elsevier, 504–524.

Blanche-Benveniste, Claire, Mireille Bilger, Christine Rouget, Karel van den Eynde & Piet Mertens (eds.), (1990). *Le Français Parlé: Études Grammaticales.* Paris: Éditions du C.N.R.S.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.

Cavalcante, Frederico A. (2015). *The topic unit in spontaneous American English: a corpus-based study*, Phd Thesis. UFMG, Belo Horizonte.

Chafe, Wallace (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago.

Couper-Kuhlen Elisabeth & Margret Selting (2017). *Interactional Linguistics: Studing Language in Social Interaction*. Cambridge: Cambridge University Press.

Cresti, Emanuela (2000). *Corpus di italiano parlato*. Firenze: Accademia della Crusca.

Cresti, Emanuela (2005). Per una nuova classificazione dell'illocuzione a partire da un corpus di parlato (LABLITA), in Elisabeth Burr (ed.), *Tradizione e innovazione: il parlato. Atti del VI Convegno internazionale SILFI*. Pisa: Cesati, 233–246.

Cresti, Emanuela (2006). Some comparisons between UBLI and C-ORAL-ROM, in Yuji Kawaguchi, Susumu Zaima & Toshiro Takagaki (eds.), S*poken Language Corpus and Linguistics Informatics.* Amsterdam: Benjamins, 125–152.

Cresti, Emanuela (2010). La Stanza: un'unità di costruzione testuale del parlato, in Angela Ferrari (ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione.* Firenze: Cesati, 713–732.

Cresti, Emanuela (2014). Syntactic properties of spontaneous speech in the Language into Act Theory: data on Italian complements and relative clauses, in Tommaso Raso & Heliana Mello (eds.), *Spoken corpora and linguistics studies.* Amsterdam: Benjamins, 365–410.

Cresti, Emanuela (2017). The empirical foundation of illocutionary classification, in Anna De Meo & Francesca Maria Dovetto (eds.), *La comunicazione parlata, Napoli, SLI – GSCP International Conference Napoli 2016.* Napoli: Aracne, 243–264.

Cresti, Emanuela (2018). The illocution-prosody relation and the information pattern in the spontaneous speech according to the Language into Act Theory (L-AcT) in Matthias Heinz & Manuela Caterina Moroni (eds.), *Prosody: Grammar, Information Structure, Interaction*. Special issue of *Linguistik online*, 33–62.

Cresti, Emanuela (2019). Per una classificazione empirica dell'illocuzione. Lo stato dell'arte, in Marco Biffi, Francesca Cialdini & Raffaella Setti (eds.), *"Acciò che'l nostro dire sia ben chiaro" Scritti per Nicoletta Maraschio*. Firenze: Accademia della Crusca, 261–279.

Cresti, Emanuela (forthcoming). The pragmatic analysis of speech and its illocutionary classification according to Language into Act Theory, in Shlomo Izre'el, Heliana Mello, Alessandro Panunzi & Tommaso Raso (eds.), *In search of the*

basic unit of speech: A corpus-driven cross-linguistic approach to spontaneous spoken communication. Amsterdam: Benjamins.

Cresti, Emanuela & Massimo Moneglia (2005) (eds.). *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*. DVD + vol. Amsterdam: Benjamins.

Cresti, Emanuela & Massimo Moneglia (2010). The Informational Patterning Theory and the Corpus-based description of Spoken language. The compositional issue in Topic-Comment pattern, in Massimo Moneglia & Alessandro Panunzi (eds.), *Bootstrapping Information From Corpora in a Cross Linguistic Perspective*. Firenze: Firenze University Press, 13–46.

Cresti, Emanuela & Massimo Moneglia (2018a). The illocutionary basis of Information Structure. Language into Act Theory (L-AcT), in Evangelia Adamou, Katharina Haude & Martine Vanhove (eds.), *Information structure in lesser-described languages: Studies in prosody and syntax*. Amsterdam: Benjamins, 359–401.

Cresti, Emanuela & Massimo Moneglia (2018b). The definition of the Topic within Language into Act Theory and its identification in spontaneous speech corpora, in Margarita Borreguero Zuloaga, Vahram Atayan & Sybille Grosse (eds.), *Models of Discourse Units*, Special issue of Revue Romane [RRO 53:1], 30–62.

DBIPIC http://www.lablita.it/app/dbipic/.

Degand, Liesebeth & Anne-Chaterine Simon (2009). Mapping prosody and syntax as discourse strategies: how Basic Discourse Units vary across genres, in Anne Wichmann, Dagmar Barth-Weingarten & Nicole Dehé (eds.), *Where prosody meets pragmatics: research at the interface*. Bingley: Emerald, 79–105.

Degand, Liesebeth, Laurence Martin & Anne-Chaterine Simon (2014). Unités discursives de base et leur périphérie gauche dans LOCAS-F, un corpus oral multignres annoté in *4e Congrès Mondial de Linguistique Française SHS Web of Conferences* 8, 2613–2625. https://doi.org/10.1051/shsconf/20140801211.

Ferrari, Angela (2015). Connettivi e gerarchie testuali in Margarita Borreguero Zuloaga & Sonia Gómez-Jordana Ferary (eds.), *Les marqueurs du discoursdans les langues romanes*. Limoges: Lambert-Lucas, 119–136.

Firenzuoli, Valentina (2003). *Le Forme Intonative di Valore Illocutivo dell'Italiano Parlato: Analisi Sperimentale di un Corpus di Parlato Spontaneo (LABLITA)*. PhD Thesis. Firenze: Università di Firenze.

Firenzuoli, Valentina & Signorini, Sabrina (2003). L'unità informativa di Topic: correlati intonativi, in Giovanna Marotta & Nadia Nocchi (eds.) *Atti delle XIII Giornate GFS*. ETS: Pisa, 177–184.

Frosali, Fabrizio (2008). Il Lessico degli ausili dialogici, in Emanuela Cresti (ed.), *Prospettive nello studio del lessico italiano, Atti del IX Congresso della Società Internazionale di Linguistica e Filologia Italiana.* Firenze: FUP. 417–424.

Giani, Daniela (2005). *Il discorso riportato nell'italiano parlato e letterario: confronto tra due corpora.* PhD Thesis. Firenze: Università di Firenze.

Haegeman, Liliane (2012). *Adverbial Clauses, Main Clause Phenomena, and Composition of the Left Periphery.* New York, NY: Oxford University Press.

Leech, Geoffrey (2014). *The Pragmatics of Politeness*. Oxford: Oxford University Press.

Mittmann-Malvesi, Maryualé (2012). *O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Topico no Portugues Brasileiro*. Phd Thesis UMFG: Belo Horizonte.

Moneglia, Massimo & Tommaso Raso (2014). Notes on the Language into Act Theory, in Tommaso Raso & Heliana Mello (eds.), *Spoken corpora and linguistics studies*. Amsterdam: Benjamins, 468–494.

Nicolas Martinez, Carlota (2012). *Cor-DiAL* (*Corpus oral didáctico anotado lingüísticament*e). Madrid: Liceus.

Panunzi, Alessandro & Lorento Gregori (2012). DB-IPIC. AN XML Database for the Representation of Information Structure in Spoken Language, in Heliana Mello, Alessandro Panunzi & Tommaso Raso (eds.), *Pragmatics and prosody. Illocution, modality, attitude, information patterning and speech annotation*. Firenze: Firenze University Press, 133–150.

Panunzi, Alessandro & Maryualê Malvessi-Mittmann (2014). The IPIC resource and a cross-linguistic analysis of information structure in Italian and Brazilian Portuguese, in Tommaso Raso, Heliana Mello & Massimo Pettorino (eds.), *Speech and Corpus linguistics*. Amsterdam: Benjamins, 29–151.

Pons Borderia, Salvador (2010). Unidades, Marcadores Discursivos y Posición, in Óscar Loureda Lamas & Esperanza Acín Villa (eds.), *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros, 327–358.

Raso, Tommaso (2014). Prosodic Constraints for Discourse Markers, in Tommaso Raso & Heliana Mello (eds.), *Spoken Corpora and Linguistic Studies*. Amsterdam/Philadelphia: Benjamins, 411–467.

Raso, Tommaso & Heliana Mello (eds.), (2012). *C-ORAL-BRASIL I: Corpus de referência de português brasileiro falado informal*. Belo Horizonte: Editora UFMA.

Rizzi, Luigi (1997). The Fine Structure of the Left Periphery. in Liliane Haegeman (ed.), *Elements of Grammar: A Handbook of Generative Syntax*, Kluwer: Dordrecht. 281–337.

Rizzi, Luigi (2013). Focus, Topic and the Cartography of the Left Periphery, in Silvia Luraghi & Claudia Parodi (eds.), *The Bloomsbury Companion to Syntax*. London: Bloomsbury, 436–451.

Rizzi, Luigi & Guliano Bocci (2017). The left periphery of the clause Primarily illustrated for Italian, in *Blackwell Companion to Syntax*, Blackwell Publishers. https://doi.org/10.1002/9781118358733.wbsyncom104.

Rocha, Bruno (2016). *Uma metodologia empírica para a identificação e descrição de ilocuções e a sua aplicação para o estudo da Ordem em PB e Italiano*. PhD Thesis, UFMG: Belo Horizonte.

Signorini, Sabrina (2005). *Topic e soggetto in corpora di italiano parlato*. PhD. Thesis, University of Florence.

Schiffrin, Deborah (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

WINPITCH <http://www.winpitch.com/>.