JoSS Journal of Speech Sciences

DOI: 10.20396/joss.v14i00.20384

PROSODIC SYNCHRONY AND THE SEMANTIC AFFILIATION OF GESTURES IN THE MULTIMODAL UTTERANCE. A CASE STUDY

MONEGLIA, Massimo[1*]
CANTALINI, Giorgina[2]

[1] University of Florence – ORCID: https://orcid.org/0000-0002-7125-8653
[2] Civica Scuola di Teatro "Paolo Grassi", Milan – ORCID: https://orcid.org/0009-0002-8533-9865

Abstract: *In the last decade, much research in the Autosegmental framework (Loehr 2012; 2014; Rohrer et al. 2023; Rohrer 2022) has highlighted a strong synchronization between Strokes' apex and prosodic prominences (pitch accents and prosodic edges). The paper addresses whether the synchrony with prosodic prominence signals the gesture's semantic affiliation. The question is studied through the fine-grained analysis of an interview given by an actor where gestures were previously annotated (Cantalini 2022). The detection of prominences is achieved following the Language into Act Theory (Cresti 2000; Moneglia & Raso 2004), which connects Perceptively Relevant Prosodic Contours ('t hart et al. 1990) to the pragmatic functions of the Information Units (IUs). Even assuming this framework, the synchronization of gestures to prosodic prominence is systematic. The paper shows that the affiliation of gestures must concern the IU where the prominence falls (neither backward nor forward). Prosodic synchrony, however, does not determine the type of semantic affiliation, which can be lexical, modal, or pragmatic; it does, however, impose substantial restrictions. The stance of modal gestures can scope on one word or one IU. Pragmatic gestures correlate with the pragmatic functions foreseen by L-AcT and synchronize with prominences bearing the corresponding value. IUs not bearing a functional prominence (Scanning IU) can affiliate lexical or modal but never pragmatic gestures.*

Keywords: Multi-Modal Utterance, Gestures / Prosody Synchronization, Gesture Affiliation, Information Structure

[*] Corresponding author: massimo.moneglia@gmail.com

1 Introduction

In his seminal work, McNeill states that there are three synchrony rules governing co-speech gestures: semantic, pragmatic, and phonological. When speech and gesture co-occur, they are expected to present the same semantic information or perform the same pragmatic function, and the Stroke of the gesture precedes or ends, but does not follow, the phonological peak syllable of speech (McNeil 1992). Lexical affiliate, however, does not automatically correspond to the co-expressive speech segment. Gestures are 'windows onto thinking' (McNeil & Duncan 2000) and can refer to the underlying concept rather than the emerging speech: ‘conceptual affiliate’ instead of ‘lexical affiliate’ (Kirchhof 2011). Therefore, the notion of ‘lexical affiliate’ is insufficient to specify the semantic relations that a gesture may find in speech (Cienki 2022).

Gestures capture both visual and speech aspects of the Growth Point simultaneously (McNeill 1992; 2005; 2015; McNeill & Duncan 2000). They can complement the reference to objects or eventualities that are analytically determined in speech, but also act pragmatically or supplement linguistic information, conveying meanings in a synthetic manner. In this paper, we will propose that identifying gesture affiliation in speech is a way to disclose how speech and gesture contribute to the meaning of the multimodal utterance.

In parallel, a substantial body of research on gesture/prosody synchronization has been conducted in the last twenty years, particularly in the Autosegmental frame (Ladd 2004; 2008; Pierrehumbert & Hirshberg 1990) using the ToBI representation System (Beckman et al. 2005). This research, starting from Loehr's (2004; 2012; 2014) work, shows a strong synchronization between prominent pitch accents and Strokes' apex. More recently, the importance of prosodic edges has also been considered, as well as the information structure in connection to gesture functions (Cantalini & Moneglia 2020; Moneglia 2024; Rohrer et al. 2023; Rohrer 2022).
Considering the two sides of the speech/gesture relationship, this paper investigates to what extent prosodic synchrony and semantic affiliation overlap and whether the gesture’s synchrony to the prosodic prominence signals the gesture’s semantic relationship. To this end, a specific methodology has been settled.

Prosodic prominence will be defined within the configurational frame of the Language-into-act theory (L-AcT) (Cresti 2000; Cresti & Moneglia 2018) that assumes a one-to-one correspondence Prosodic-Unit (PU) / Information-Unit (IU) and considers that the prosodic cues of the IUs specify their function within the information structure of the utterance.

The functional values of IUs, such as, for instance, the expression of the Illocutionary force, the Topic function, and the Parentheses, are conveyed by Perceptively Relevant Prosodic Contours, which, by definition, constitute the main prosodic prominence in the utterance since they are intentionally performed (’t Hart et al. 1990).

Looking at these functional values, the linguistic information encoded by IUs will be highlighted beyond the sole lexical information. This approach will be crucial to identifying the affiliation of pragmatic and modal gestures (Kendon 2004; 2017) and their relation to prosody.

The dataset used for this case study is an interview given by an actor from the Cantalini Corpus (Cantalini 2022), where gesture structure and information structure have already been annotated, resulting in 214 Information units and 128 Gestures in 5:10 minutes of Italian spontaneous speech. To this dataset, we added two independent annotations: a) the prosodic prominences found in each PU; b) the linguistic affiliation of each gesture (if any). The affiliation, which necessarily relies on gesture interpretation in the local linguistic context of its occurrence, will be based on an operative method intended to make it as objective as possible.

The paper's structure is complex and follows a series of research steps. Section 2.1 will provide a brief introduction to L-AcT and the methodology for annotating prosodic prominence.

In Section 2.2, the methodology for defining gesture affiliations is sketched. In Section 2.3, the dataset and the annotation strategy in ELAN will be presented.

Section 3 describes the findings concerning the relation among Strokes / Prosodic prominences / Linguistic affiliations derived from the annotation. We will first present a fine-grained analysis of prototypical examples for each type of gesture's affiliation found in the dataset, detailing the empirical process followed to identify the affiliation of the Stroke and its relationship with the prosodic prominence. In section 3.1. we will deal with gestures affiliated with the meaning of the locutive content (word or phrase). In Section 3.2, we will go through the pragmatic functions that constitute the affiliation of a pragmatic gesture, specifically the Topic (3.2.1), the Illocutionary information (3.2.2), the Listing (3.2.3), and the Parenthesis (3.2.4). In Section 3.3, the linguistic contexts that do not constitute a possible affiliate of a pragmatic gesture will be considered. In Sections 3.4 and 3.5, we will identify gestures with Modal values visually added to the verbal content, as well as the peculiar case of gestures that find their affiliation in silent pauses. In Section 3.6, we will compile the quantitative and qualitative results obtained and attempt to draw generalizations regarding the overall question raised by the paper. We will claim that gestures find their affiliation in the IU to which they are synchronous, neither backward nor forward. On this basis, we hypothesize that the IU constitutes the linguistic counterpart of the gesture information, unpackaging the Growth Point from which the utterance originates.

Prosodic prominence attracts the gesture and signals a meaningful relationship with the linguistic information conveyed by the IU. Still, it does not select the specific affiliation among the lexical, pragmatic, or modal meanings the gesture may convey. This choice, however, turns out to be constrained by distributional restrictions. The few cases of asynchrony found in the dataset are discussed in 3.7 and framed within general prosodic and gestural strategies. Finally, the findings are summarized in the conclusions as a takeaway message for future research.

2 Methodology

2.1 Information Structure and the Annotation of Prosodic Prominence according to the Language into Act Theory

2.1.1 The utterance and its information structure

L-AcT applies the traditional speech act theory (Austin 1962) to the analysis of spontaneous speech by looking at prosodic performance. As a whole, it considers that, in speech, prosody is the necessary means to perform speech acts and behaves as an embodied interface with the semantic and morpho-syntactic constituents (locution). No illocutionary interpretation can be assigned to speech if the prosodic performance is not considered (Cresti 2020).

The utterance is the primary reference unit for speech analysis since it is the minimal entity that can be pragmatically interpreted (Biber et al. 1999; Cresti 2000; Izre'el et al. 2020). Prosody is crucial to segment the speech flow (Izre'el et al. 2020; Barbosa & Raso 2018)). The L-AcT methodology is based on recognizing prosodic breaks relevant to perception (Swerts 1997; Collier et al. 1993). The utterance is identified as a prosodically terminated sequence, i.e., a sequence marked by perceptively relevant prosodic breaks to which the competent speakers assign a terminal value (// in the transcripts).

Once a terminated sequence is identified, it is segmented into Prosodic units (PUs) when non-terminal prosodic breaks occur (marked with /)[1]. According to tradition, these prosodic boundaries are assumed to segment the utterance into units corresponding to IUs (Chafe, 1994).

L-AcT states that IUs convey a specific information function defining the information structure of the utterance (Cresti 2000; Cresti & Moneglia 2018; Moneglia & Raso 2014). Information functions have a pragmatic definition and are framed in a prosodic pattern ('t Hart et al. 1990). The core IU of an utterance is the Comment, whose function is to specify how the locutive content must be interpreted in the world, i.e., the illocution of the utterance, such as assertion, question, order, and several illocutionary activities that are performed in ordinary speech (Cresti 2020). The Comment is the only necessary IU type in the utterance and can be interpreted independently of any other IU.

In this frame, information structure is not limited to the Topic/Comment or Topic/Focus relation as commonly assumed (Krifka 2008; Krifka & Musan 2012). The pragmatic definition of the Comment unit and the idea of a strict correspondence among Prosodic-unit, Information-function, and Prosodic-cues allow us to treat concepts considered separately in the linguistic literature under the same principles and to define them for their pragmatic function.

In particular, according to the standard L-AcT definitions (Moneglia & Raso 2014), the Topic (TOP) specifies to the addressee the pragmatic domain that is relevant to the Comment interpretation; the Appendix integrates the locution of the Comment or the Topic (APC/APT), the Parenthesis (PAR), adds information or a modal evaluation to the utterance on a different plane; the discourse Connector (DCT) serve to relate one information patterns with the preceding ones. Finally, various IU types are dedicated to managing the interaction with the addressee, working as Discourse Markers (Schriffin 1987; Raso 2014)[2].

All the above units depend on the Comment and have a one-to-one correspondence with a dedicated PU shaped by a Perceptively Relevant Prosodic Contour with differential prosodic properties. However, the strict correspondence PU / Information function is weakened considering that the locutive content of the Topic, Comment, Appendix, and Parenthesis can be scanned into prosodic units (SCA), which do not bear an autonomous information function.

The information patterns around a COM defined at the utterance level can be, in principle, projected[3] in higher-level reference units that characterize prosodically terminated sequences in speech, as sketched in the following paragraph.

2.1.2 Illocutionary Patterns and Stanzas

Considering prosodically terminated sequences, L-AcT identifies two types of reference units beyond the utterance: *illocutionary patterns* and *stanzas*, which feature more than one Comment unit within the same reference unit.

An Illocutionary pattern is a terminated sequence compounded by a chain of Comments called Multiple Comments (CMM). It is performed in one shot in a single prosodic and pragmatic program, giving rise to a natural rhetoric model where two or more pragmatic units are linked to

[1] Their identification has been proven consistent in corpus annotation across languages, reaching a high interrater agreement in different frames (Du Bois 1993; Buhmann et al. 2002; Amir et al. 2004; Moneglia et al. 2005; Moneglia et al. 2010; Izre'el & Mettouchi 2015; Panunzi et al. 2020). See Barbosa & Raso (2018) for an overview of the speech segmentation issue.

[2] Unfortunately, only one occurrence of these IU types (Dialogic Units) in this monologic dataset. For this reason, Dialogic units are not discussed here despite their interest in identifying pragmatic gestures (Lopez-Ozieblo 2020).

[3] There are restrictions on this projection that cannot be treated within this paper.

achieve a goal (Listing, Reinforcement, Comparison, and Tag questions are the main patterns) (Cresti 2000; Panunzi & Saccone 2018; Saccone & Panunzi 2020).

Stanzas are sequences of Comment units (Bound Comments - COB), when turns are longer than in informal dialogic exchanges and a greater degree of textual complexity than a single utterance is needed, such as life stories, narratives, lectures, and conferences (Cresti 2010; Saccone, 2020). COBs are linked within the Stanza till a terminated prosodic break through successive additions following the flow of thought; they are prosodically "linked" together (Panunzi & Saccone 2018) and bear homogeneous and weak illocutionary forces. From a prosodic perspective, COB units generally exhibit prosodic contours that are concluded by a *final rise*, perceived as a continuation mark. A COM with a terminal profile concludes Stanzas. Stanzas can be internally structured into information patterns that develop around each COB, typically including Topic (TOP), Appendix (APC), and Parenthesis (PAR).

2.1.3 Operative method for the annotation of prosodic prominence in the dataset

The prosodic cues conveying the functional value of an IU are perceptively prominent because the speaker intentionally uses them in his speech activity ('t Hart et al. 1990) for packaging the information. Therefore, the main prosodic prominence of an IU is defined as the Minimal Perceptively Relevant Prosodic Contour that characterizes its functional value: we use the acronym MPR to refer to this contour. To the end of this paper, let us sketch the main properties of the *root* unit shaping the Comment, the *prefix* unit shaping the Topic, the *suffix* unit shaping the Appendix, and the prosodic properties of the Parenthesis (summarized in Table 1).

The *root* unit ('t Hart et al. 1990) shows a variety of prosodic contours that select the specific illocution conveyed by the IU (Firenzuoli 2003; Moneglia 2011; Rocha 2016; Cresti 2020; Cresti & Moneglia 2023) and can comprehend a *preparation,* a *nucleus*, and a *tail*. The *nucleus* is necessary and sufficient to convey illocutionary information.

The *prefix* IU must precede the Comment and, independently of its length, record highly prominent language-specific f0 contours aligned to the last syllables of the units (Firenzuoli & Signorini 2003; Signorini 2005; Raso et al. 2017; Cavalcante 2020).

The *suffix* is a holistic contour characterized by an overall slightly descending f0 profile to the baseline, following a TOP or a COM (Cresti 2012; Cresti 2021).

Parenthesis shows a *platform* contour marked by f0 slope with respect to the other IUs of the information pattern (Cresti 2000; Moneglia & Raso 2014; Saccone & Trombetta 2021 Tucci 2010; Tucci 2012; Saccone & Panunzi 2023)

These contours constitute the core prosodic prominence within the utterance and are annotated in the dataset. SCA units, which lack autonomous information functions, do not have an MPR. Beyond this crucial information, we also find Perceptively Relevant Prosodic Movement in the utterance that signals lexical emphasis but does not play a functional role. Typically, they correspond to a sudden pitch rise or vowel lengthening. We use the acronym *MPR to refer to this type of prosodic prominence, which is systematically annotated in the dataset.

To clarify the annotation workflow, let us consider how MPRs and *MPRs are annotated in one utterance of the interview.

(1a) l'azione /^{TOP} che sta sotto /^{SCA} la volontà che sta sotto quella battuta/^{APT} ti fa trovare la sua sonorità //^{COM}

'The action /^{TOP} which is under /^{SCA} the will underlying that line /^{APT} it makes you find its own sound //^{COM}'

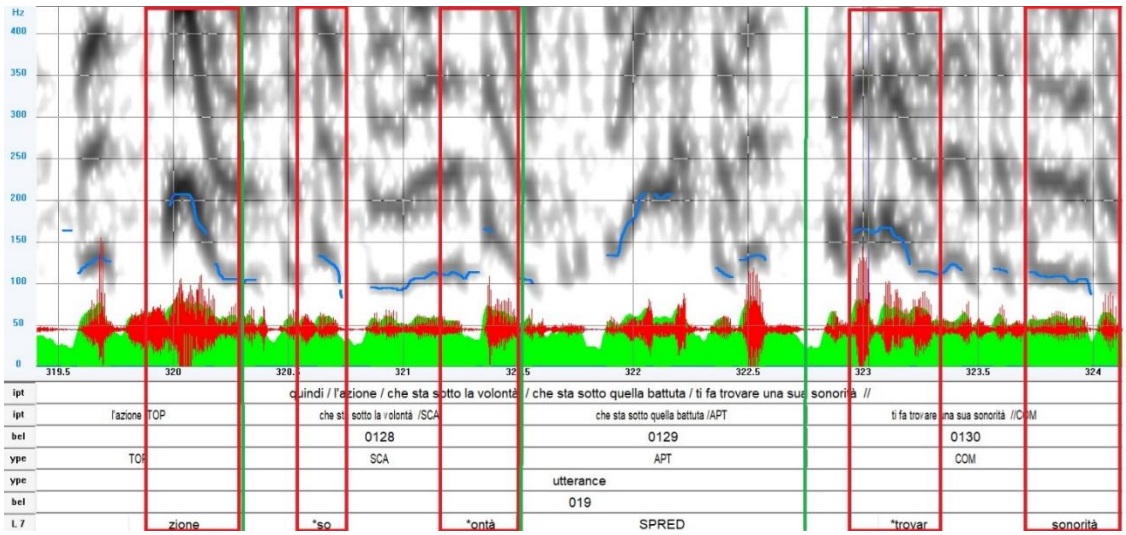

Figure 1a: Annotation of prominences on the f0 track of (1a)

(1) is divided into four IUs (boundaries are marked by the green lines in the f0 track). The segmentation and the assignment of each IU to an IU type were achieved in previous work. The accuracy of this previous annotation has been verified by a second coder when marking the prosodic prominences of each unit and amended when necessary. The comparison of the two annotations shows a substantial agreement (Kappa Fleiss = 0.779).

In (1), the first IU is a Topic, and the last is the Comment. The Comment can be interpreted in isolation and is an assertion with a conclusive value. If the other IUs are erased, the utterance will work fine as a Topic Comment Pattern, both from a semantic and prosodic point of view. The Topic and the Comment bear a prosodic prominence, which is the nuclear parts of the units bearing the respective functional values (MPR marked in red boxes). The Topic's prominent *rising/falling* contour corresponds to one of the forms foreseen in L-AcT for this unit (Signorini 2005; Cavalcante 2020). The *gently falling contour* of the Comment is the typical profile of a Conclusive illocution (Cresti 2020).

The functional value of the two contours is verified perceptually, ensuring that we get a well-formed Topic/Comment prosodic pattern by erasing all segmental material except the two nuclei. This empirical process challenges the annotator's perception and requires maintaining a fixed interpretation of the segmental material of the IU, despite its modification. For instance, the phonological word in the Topic unit of (1) records three syllables (la-zio ne'). If the first syllable is erased, the sequence 'zione / ti fa trovare la sua sonorità' is perceived as a well-formed Topic / Comment utterance from a prosodic point of view, allowing the interpretation as a Topic of the pseudoword 'zione'. On the contrary, if even the second syllable is erased, the pseudoword 'one' within the pattern 'one / ti fa trovare la sua sonorità' is not perceived as a well-formed Topic / Comment pattern; i.e., 'one' cannot be interpreted as a Topic. The reason for this judgment can be only prosodic, since there can be any semantic differences between two pseudowords. Therefore, we have perceptual evidence that the *rising-falling* contour in the last box on the right of Figure1a is the MPR that has pragmatic value since it is necessary and sufficient to perform the Topic function.

The process is repeated considering the Comment. In short, if all segmental material is erased in the unit 'ti fa trovare la sua sonorità' except the last word 'sonorità', the sequence 'zione / sonorità' is perceived as a well-formed Topic / Comment Pattern, and the word 'sonorità' can be interpreted in isolation, keeping the illocutionary value of Conclusion of the utterance. Therefore,

the *gently falling* contour on the word 'sonorità' in the last box of Figure 1a constitutes the MPR of the Comment.

The processing of prominences that do not bear a functional value follows the previous steps. For instance, in the Comment unit, in correspondence to the word 'trovare', a prominent *rising-falling* movement (first box of the IU, on in Figure 1a) is evident to perception and on the f0 track. However, if the word is isolated, erasing all the other syllables of the IU, the sequence 'zione/trovare' is not perceived as a well-formed Topic Comment pattern, and 'trovare' cannot be interpreted at all. Therefore, we verify that the *rising-falling* contour on the word 'trovare' constitutes one emphasis with no functional value (*MPR).

Beyond the Topic / Comment pattern, the two units following the Topic in example (1) (SCA / APT) also present a typical annotation case. It is verified that the sequence does not constitute a valid information pattern with the Comment if the Topic is erased. The scanning unit, however, bears two prosodic prominences, giving emphasis (*MPR) to the words 'sotto' (under) and 'volontà (will). The APT has no nuclear part and exhibits an overall *falling* profile until *the baseline*. In this case, we do not find any emphasis. Therefore, no specific segment of the IU constitutes a prominence, and the contour is marked 'SPRED,' meaning that the perceptually relevant movement with functional value is up to the holistic profile of the unit.

Once each prominence in the utterance is identified, each movement participating in the prosodic contour is manually annotated on the f0 track, and the syllables concerned are transcribed in a layer.[4] Their perceptual relevance is then confirmed by observing whether the annotation fits the Glissando threshold provided by WINPITCH (Martin 2015). [5]

[4] The reader can download the audio files and the video file referred to in the paper from https://doi.org/10.5281/zenodo.17079664. A second expert annotator replicated the perceptual judgements, and the annotation was delivered after consensus agreement. Being manually annotated on the f0 track, the edges of the prominences suffer from some under-determinacy (+/-10 ms) that is accepted in this practice.

[5] Glissando is the rate of f0 change above which a melodic change is supposed to be perceived and determines the perceptual boundary between a static pitch and a melodic variation. If the variation is less than the threshold, the perception will correspond to a static tone; if it is higher, it will be perceived as a melodic variation. The threshold was established for synthetic vowels using a semitone scale (Rossi 1971; Rossi 1978; 't Hart 1975).

Table 1: Specification for the annotation of prosodic prominence according to L-AcT

IU Type (TAG)	**IU Information Function**	**Prosodic Contour with functional value (MPR)**	**Prosodic emphasis on words (*MPR)**
COMMENT (COM; COB; CMM)	Express the illocutionary information within Utterances, Illocutionary patterns, and Stanzas.	Root unit with a Nucleus MPR: Open repertoire of nuclear forms	Possible
TOPIC (TOP; TPL)	Specify the domain of relevance of the illocutionary activity expressed in the COM	Prefix unit with a Nucleus MPR: Closed repertoire of nuclear forms aligned to the right	Possible
APPENDIX (APC; APT)	Integrate the information conveyed by COM or TOP	Suffix unit with no Nucleus MPR: Falling profile of the IU according to the declination line (SPRED)	Possible
PARENTHESIS (PAR)	Metalinguistic insertion in the utterance	No Nucleus MPR: Platform profile of the IU at a lower f0 range (SPRED)	Possible
DISCOURSE CONNECTOR (DCT)	Specify the relation of the subsequent Information pattern to the preceding one.	Holistic form of a single word MPR: SPRED	Not possible
SCANNING (SCA)	Scans the IU and does not specify any information function	Platform profile - No nucleus No MPR	Possible
Time taking & Interrupted units	No information function	No MPR	Not possible

2.2 Operative method to identify the affiliation of a gesture

The linguistic affiliate is the linguistic information allowing the interpretation of a gesture in the context in which it is performed, i.e., the information allowing the disclosure of the metaphor that the gesture embodies in the context of the utterance (Cienki 2023; Cienki & Muller 2008). The following is an example of semantic affiliation in the lexical material of an information unit.

Consider first the description of the gesture phrase in Figure1b. In the preparation phase, the subject raises both hands mid-air, then, in the Stroke, keeps the hand's position (*Vertical Palm Up Open Hand, Molding a rounding shape*), moves them down on his right side, and finally returns to rest position.[6]

[6] For 'Molding' as a mode of representation, see Müller / Bressem / Ladewig 2013: 711-712; Müller 2014: 1691.

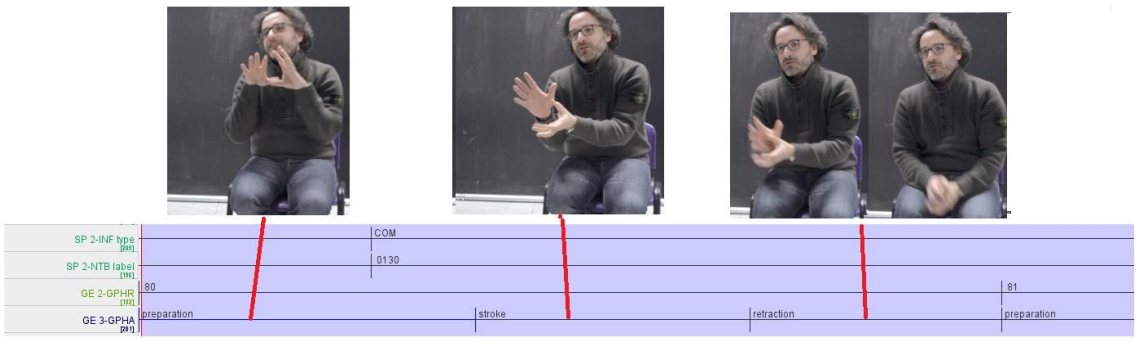

Figure 1b: Phases of one gesture, mute vs with audio source

Gesture's semantic value is almost underdetermined. We might say that the subject depicts something round that moves in space. However, the meaning of the gesture may be better approximated considering how it is connected to the linguistic content of the IU synchronous to the Stroke in (1b), that is, the COM of the utterance just presented in (1a):

(1b) ti fa trovare una sua sonorità //^{COM}
'it makes you find its sound //^{COM}'

The metaphor underlying the gesture can be disclosed by considering the wording of (1b). The subject makes a round object appear in his proximal space as it happens when something is found ('trovare'), and the round shape of the object refers to the qualities of the sound ('sonorità'), which is metaphorically considered 'well-rounded'. Therefore, the phrase 'trovare la sua sonorità' (find its own sound) is the locutive affiliation of the gesture, which synthetically expresses a meaning complementing that phrase.

We used this method to define the affiliation of each Stroke in our dataset, where prosodic prominences were not previously annotated.

Notice that the description of the gesture's meaning is highly idiosyncratic. Different annotators may use different descriptors; however, the affiliation selection is more intersubjective: people agree that the gesture has something to do with 'trovare' (find), and that what was found is sound.[7]

Crucially, we considered the various linguistic levels encoded in the IU as a possible object of affiliation beyond the locutive content (words or phrases). As the Language into Act theory highlights, the IU can bear an information function within the utterance; it can be a COM specifying an illocution, a Topic specifying a domain of relevance, a Parenthesis, inserting a different discourse plane, or it can be a CMM framed in an illocutionary pattern, etc. This annotation was already present in ELAN when searching for affiliation, and it is very relevant in identifying pragmatic and modal gestures (Kendon 2004:158-159; Bressem & Müller 2014; Kendon 2017).

[7] The selection of affiliates in the dataset, given in the first shot by the first author, was verified by the second author, and a consensus agreement was reached in the case of disagreement.

2.3 The structure of the data set

2.3.1 The annotation of Gesture / Information structure relation

The dataset is a monologue from an informal interview with an actor about his work, taken from the Cantalini Corpus (Cantalini 2022). The annotation of gesture structure, prosodic boundaries of the utterance, information unit boundaries, and information function played by each information unit, achieved in previous work, shows the following layer hierarchy in ELAN:[8]

- Gesture hierarchy
    - G-Units
        - GPhrases
            - G-Phases: Preparation, Pre-Stroke hold, Stroke, Post-Stroke hold, Retraction

- Terminated sequence hierarchy
    - Terminated sequence type: Utterance, Stanza, Illocutionary pattern
        - Prosodic boundaries identifying PUs: NTB / TB
            - Information function of PUs: (according to the L-AcT tagset)

The monologue has a 5:10-minute duration. It comprises 128 GPhrases and 214 IUs aligned to the speech signal. 192 IUs belong to the main types foreseen in L-AcT. However, the annotation also highlights time-taking units (TMT), some isolated units working as discourse markers, and some interrupted units. The following table illustrates the IU types and their amount.

Table 2: Information Unit Types in the dataset

COM	CMM	TOP	TPL	PAR	APC APT	SCA	DCT	INT IMP-r PHA	i-COM i-TOP	TMT	UNC	
58	8	28	5	24	8	49	11	4	3	14	2	214

Based on this annotation, we added two independent layers to the dataset, following the methodology presented in sections 2.1.3. and 2.2:

a. The selection of the linguistic affiliation by each Stroke
b. The timestamp of the Perceptively Relevant Contours found in each information unit (MPR / *MPR).

2.2.2 The annotation of prosodic prominence

Given the above units aligned to speech, in this research, we annotated 201 prosodic prominences in a dedicated layer according to the L-AcT methodology, 73 of which emphasized a word (*MPR), and 128 movements conveyed informational value (MPR).

[8] Gesture annotation follows the model in Kita et al. (1998); with integration from Bressem, Ladewig, Müller (2013) and Ladwing & Bressem (2013). See Cantalini et al. (2020) for the validation measures on the original corpus annotation.

The annotation was achieved through WINPITCH, then saved in PRAAT and imported into the original ELAN files. The visual information did not influence the annotation of prominences, as it was not available to the annotator, who worked solely on the audio source.

To make clear the possible correspondence between the prosodic prominence and the semantic information in speech, a layer was added specifying the syllables involved in the prominence, according to the following hierarchy:

SP3-MPR: MPR or *MPR

SPR-MPR-OV: Syllables involved in prominence or SPREAD

The following is a quantitative representation of the dataset, derived from the annotation of prosodic prominence and the previous annotations of gestures and information units.

Table 3: Gestures, information units, and prosodic prominences in the dataset

Duration	Information units	Prominence of MPR type	Prominence of *MPR type	GPhrs
5:10	214	128	73	128

Prosodic prominences have a temporal duration and are not an apex. Therefore, we could compare prosodic movement and Stroke duration, observing whether they overlap.[9]

2.2.3 The annotation of gesture affiliates

From a theoretical point of view, the linguistic levels that convey meaning in an utterance and can affiliate a gesture can be found at the following levels, each headed by the respective tag:

- LEX: the locutive level, in which semantic information is distributed (word, phrase, or proposition);
- INF: the information function of the IU (COM/COB/CMM; TOP/TPL; PAR; APC/APT / SCA; etc.);
- MOD (L/-U): the modality/stance expressed by the speaker on the IU content: a) about a single lexical choice (MOD-L); b) about one IU (MOD-U).

The selection of the possible gesture's affiliation in the stretch of speech according to the above possibilities was determined for each GPhrase in the dataset. The prosodic prominences in IUs were unavailable to annotators when the interpretation was given. The affiliation was defined in ELAN in four dependent layers, parent of the Stroke (or Independent Hold) of each GPhrase, according to the following hierarchy:

G2 GPHA (Stroke / Independent Hold)

GE3: Description of the gesture feature[10]

GE3-Gloss: gesture interpretation in the given context

AFFILIATION OV: semantic information that allows interpretation

[9] It is worth noticing that the relationship between Prosodic prominences and IUs is not a one-to-one correspondence. Sixty-six units in the previous table never bear an MPR because of their nature, which does not specify an information function (SCA, interrupted IUs, and filled pauses). Moreover, the MPR of PAR and APC/T units (SPRED) is not marked when the PAR or the APC/T shows lexical emphasis. On the other hand, one IU can also present one *MPR beyond the MPR that characterizes its information value, so scoring two Prominences.

[10] Gestures are described by applying the descriptors derived from Bressem & Müller (2014), Muller/ Bressem / Ladewig, 2013), and other standard reference literature (Kendon, 2004; Bressem, 2013; Muller, 2014.

AFFILIATION TYPE according to the following closed vocabulary: LEX (comprising the word and Phrase levels); INF; MOD-L; MOD-U

Based on the annotation of Affiliations, Prosodic prominences, and Strokes, we investigated whether and how the prominent prosodic information of IUs is predictive of the gesture's linguistic affiliation. In particular, we verify:

a) the type of affiliation found by each Stroke;
b) the possible relation between the affiliate and the prominences in the stretch of speech;
c) the synchrony between each Stroke and the prosodic contours constituting prominences in the utterance.

Given that the gesture systematically starts before affiliation and prosodic prominence, we considered a Stroke synchronous to the prosodic cue or to the affiliate if the time span of the Stroke reaches the contour and/or its affiliation.

3 Discussion

3.1 Locutive affiliations

Example (2) and Figure 2a present a sequence of two parenthetic units, each accompanied by a gesture. Considering the content of the two units, each Stroke can be interpreted considering its affiliation to the locutive content of each IU and, in this case, specifically to one lexical item.

(2) nella **vita** non lo so /^{PAR} ma a teatro **sicuro**[11] /^{PAR}
'in **real life** I do not know /^{PAR} but in a piece for **sure** /^{PAR}'

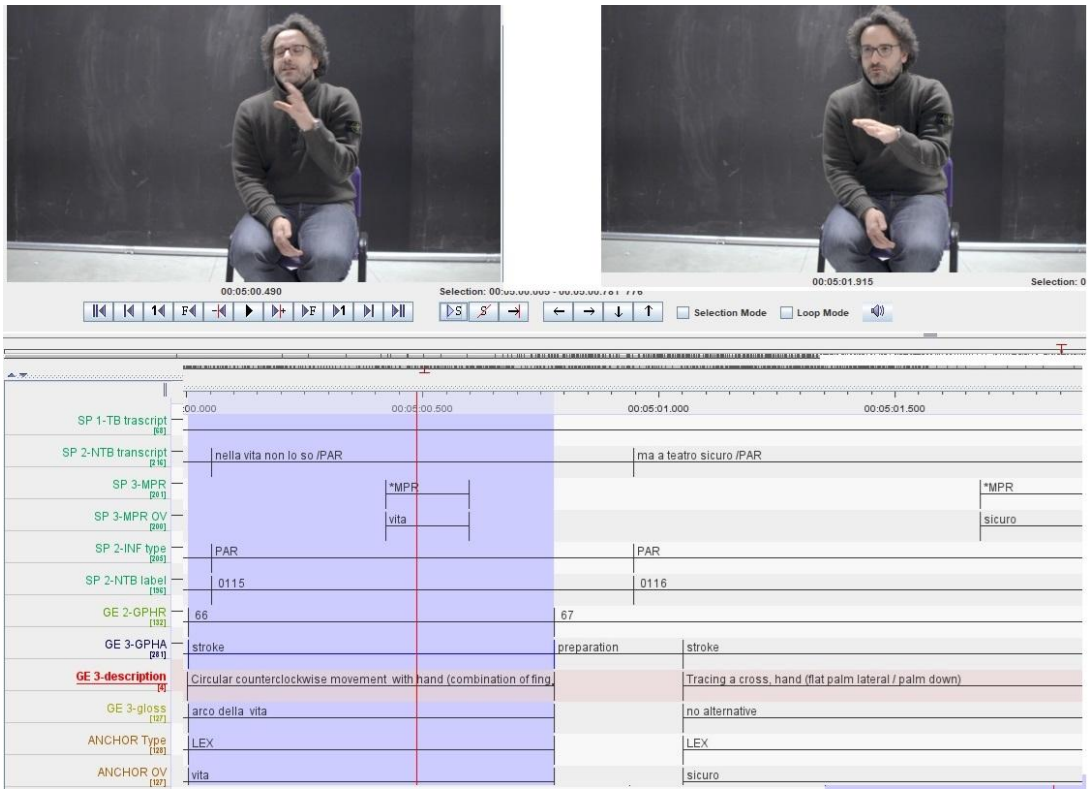

Figure 2a: Strokes affiliated at the word level in (2)

[11] For a better reading, the affiliation is marked in bold in the examples here and below, and the syllables involved in the prominence are underlined (not in the English translation). The Information functions of the IU are superscripted at IU boundaries and are in bold when the gesture is affiliated with that function.

The first Stroke is a *circular counterclockwise movement* with the hand (*combination of fingers one-five-bent*) affiliated with the word 'vita' (life). Establishing this relation, the metaphor underlying the gesture becomes clear: 'the arch of life.' The second Stroke traces a cross *(flat hand palm lateral / palm down*) and can be interpreted when affiliated with the word 'sicuro' (for sure). The underlying metaphor in this context might be 'there is no alternative'.

Two prosodic prominences are reported in ELAN on the words 'vita' (life) and 'sicuro' (sure). This corresponds to the emphasis marked in red boxes on the f0 track in Figure 2b

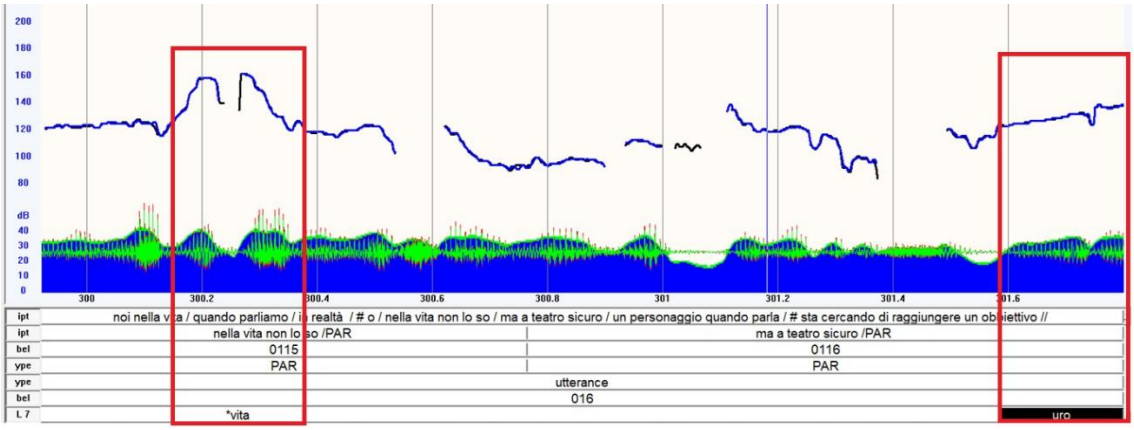

Figure 2b: Annotation of prominences on the f0 track of (2)

Therefore, we verify that, in this case, we have a complete synchrony of the three variables:

- The two Strokes overlap almost completely with the IUs and end when reaching the prominence
- The lexical affiliate overlaps with the prosodic prominence

Affiliation in a word was found in all IU types except the units with no lexical content, like time-taking units (TMT). Synchrony with the prosodic prominence (*MPR or MPR) occurs in almost all cases. The few exceptions are discussed in 3.7.

The affiliation of gestures to the locutive content is not limited to single words. Example (3), again in a Parenthesis IU, illustrates a prototypic case of gesture affiliation regarding a phrase (NP), i.e., a gesture affiliated at the compositional level of meaning.

(3) se è un regista /^{TOP} con cui tu non hai /^{SCA} un **rapporto consolidato** /^{PAR}
'if he is a director /^{TOP} you don't have /^{SCA} a **solid relationship** with /^{PAR}

The PAR unit in (3) is synchronous with a *rapid up-down movement at the speaker's center and center-center space*[12]*, palm down open hand to palm up*.

[12] See McNeill (1992: 89) for the notion of 'gesture location'.

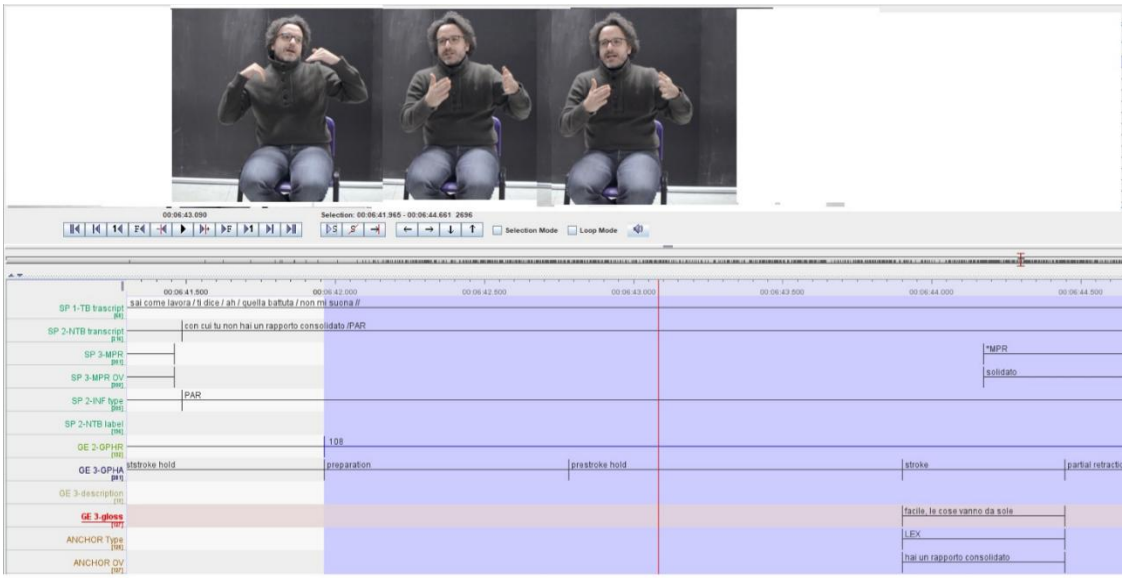

Figure 3a: Stroke affiliated at the level of Phrase in (2)

This gesture cannot be interpreted in connection with 'rapporto' (relationship) or 'consolidato' (solid). Still, it can be understood when considering the NP (a solid relationship). In this relation, the gesture may refer to an 'easy development of events', a quality that characterizes solid relationships. In other words, the connection to the phrase discloses the underlying metaphor, 'an easy and fluid deal as happens when the relation is solid'. So, the gesture complements the meaning of the locution, not at the word level but at the compositional level.

The Stroke is synchronous with one *MPR reaching the lengthening and the rise of intensity on the word 'consolidato'(solid), characterizing the f0 track in Figure 3b.

From the two examples regarding the affiliation in the locutive content, we can draw a simple conclusion. Admitting that the synchrony with the prominence predicts the affiliation, this cannot be the case when the affiliate is a Phrase. Nothing in the prosodic prominence in the word 'consolidato' allows us to foresee that the affiliate is a Phrase instead of the word on which the prominence falls.

Therefore, considering the simple case of the gesture affiliation found in a phrase, we verify that the gesture's synchrony with a prosodic prominence cannot predict the affiliation type. This conclusion will become even more evident when looking at pragmatic and modal gestures.

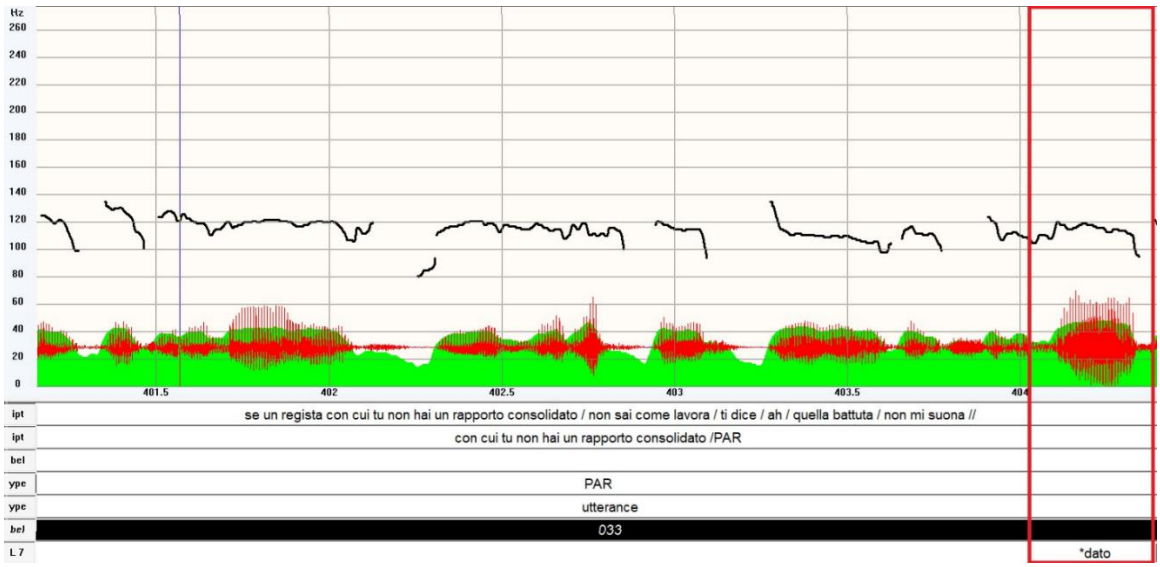

Figure 3b: Annotation of prominences on the f0 track of (3)

3.2 Pragmatic affiliations

3.2.1 Topic Function

The possibility of affiliating a gesture in the functional information expressed by IUs may shed light on the concept of "pragmatic gestures" beyond the three functions foreseen by Kendon: a) adding interaction with the interlocutor (performative function); stressing parts of the utterance (parsing function); c) changing the interpretation of the utterance or commenting on it (modal function); (Kendon 2004:159).

Let's see the prototypical cases of gestures whose affiliation cannot be found in the locutive meaning but can be interpreted if the IU's informational value is considered.

(4) noi <u>nella vita</u> /$^{\mathbf{TOP}}$ (quando parliamo) /$^{\mathrm{TOP}}$
'in our life /$^{\mathbf{TOP}}$ (when we talk)/$^{\mathrm{TOP}}$'

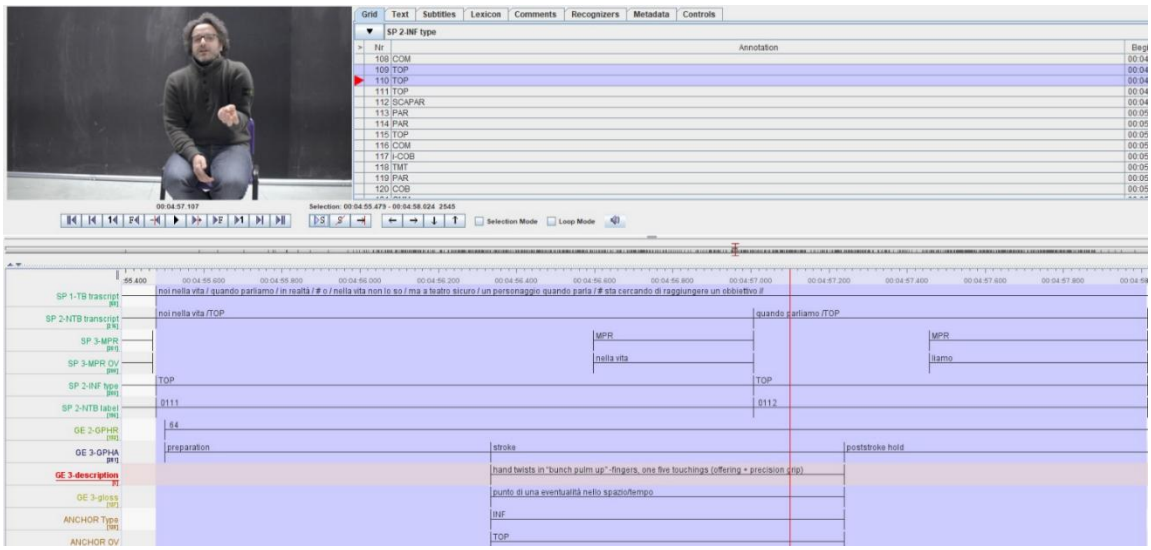

Figure 4a: Stroke affiliated with the Topic function in (4)

The gesture in Figure 4a goes *from the rest position to mid-air toward the interlocutor*; *the hand twists in "bunch pulm up" - the combination of fingers*, *one-five touching, like in offering + precision grip*.

Contrary to the previous cases, in the context of this utterance, the Stroke cannot be interpreted in connection with any lexical item of the IU 'vita' (life), nor in connection with the subsequent IU 'quando parliamo' (when we talk).

As Muller (2014) noted in a different frame, this kind of gesture has a pragmatic value, creating a point in time and space for the interlocutor. In so doing, the speaker refers to a possible eventuality. Gesture affiliation becomes clear when considering that the unit's function is to set up a Topic. The gesture refers to an eventuality that constitutes the aboutness of the illocutionary activity, and by consequence, finds its affiliation in the unit's functional value.

Looking at the synchrony with prosodic cues, the gesture extends over the IU and reaches the MPR (and is kept in post-Stroke hold over the following Unit, which is still a Topic). As Figure 4b shows, the prosodic contour is one of the typical movements signaling the Topic; i.e "falling /platform /rising" placed on the final syllables of the IU (Cavalcante 2020; Raso Cavalcante & Mittmann 2017; Firenzuoli & Signorini 2003; Cresti & Moneglia 2018).

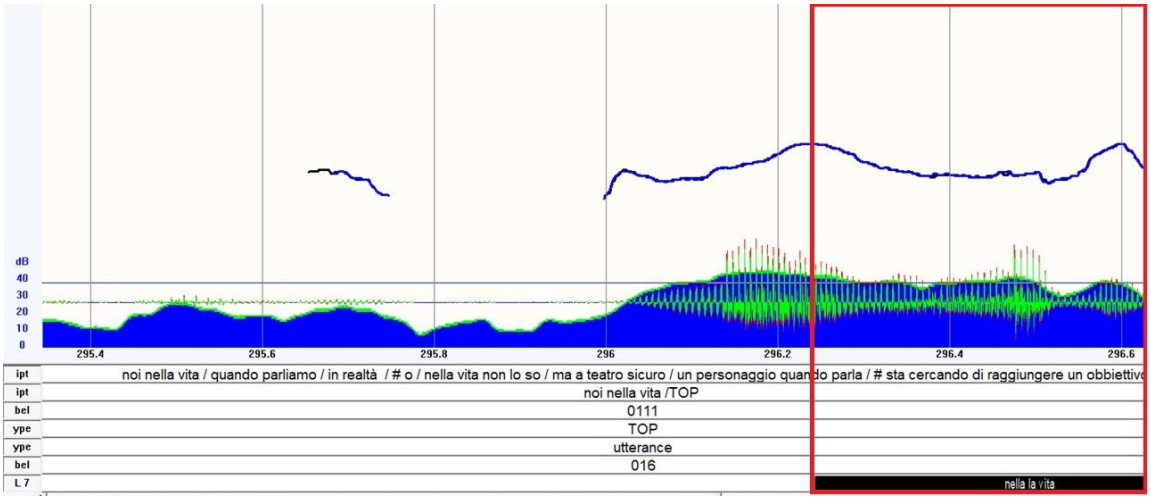

Figure 4b: Annotation of prominences on the f0 track of (4)

Therefore, considering that the MPR marks the topic function of the IU, we can focus on the fact that the gesture is synchronized with the formal prosodic cue of its pragmatic affiliation.

Gesture's affiliation with the information function of Topic is frequent. Almost half of the gestures occurring in a Topic Unit have a pragmatic interpretation that matches the Topic function. According to the type of entity selected as a Topic in the utterance, these gestures refer to one eventuality, one object, a subject, the speaker itself, or one distal or proximal point or period in the space/time coordinates. As in (4), the gesture synchronizes to the Topic IU and reaches the MPR in its Stroke phase.[13]

We can notice again that the relation of the Stroke with the MPR marking the Topic function does not allow us to foresee that the gesture's affiliation is with the IU's pragmatic function. For instance, in the following example of the Topic unit, the gesture depicts *a forward-*

[13] The possibility of reaching the MPR in the post-Stroke-hold phase is also verified in the dataset, so synchronizing in the full expressive phase (multi-segment phase as opposed to single-segment phase) according to Kita et al. (1998).

twisting movement. It is interpreted in connection with the word 'azione'(action): 'the action done when interpreting a text is something that finds its way'.

(5) **l'a<u>zione</u>** /^{TOP} […] ti fa trovare una sua sonorità //^{COM}
'**the action** /^{TOP} […] makes you find its own sound //^{COM},

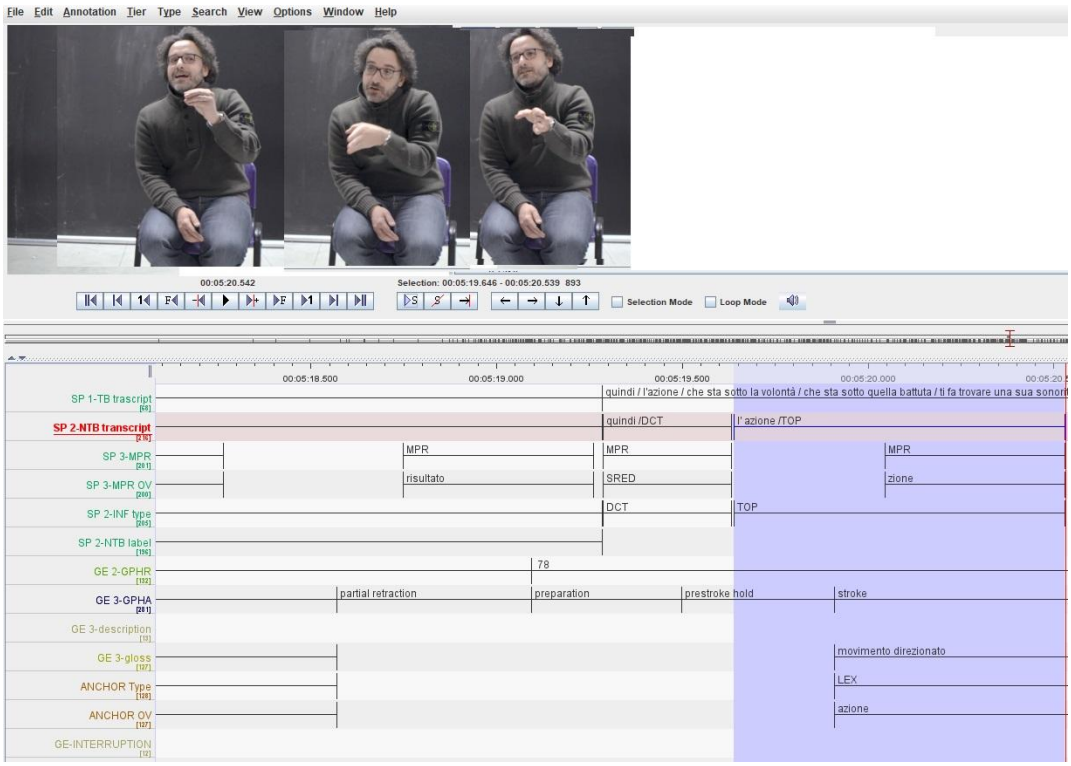

Figure 5a: Stroke affiliated at the word level in the Topic unit of (5)

The gesture almost overlaps the functional movement of the Topic highlighted in Figure 5b, which also marks the word that constitutes the actual gesture's affiliation. So, synchrony with a functional MPR does not necessarily predict a pragmatic affiliation.[14]

[14] Notice that if the pragmatic gesture in Fig 4a is substituted in this context, it works fine and still signals the Topic function of the Unit, but the reverse doesn't hold i.e. the gesture in Fig 5a is meaningless in the context of Fig 4a

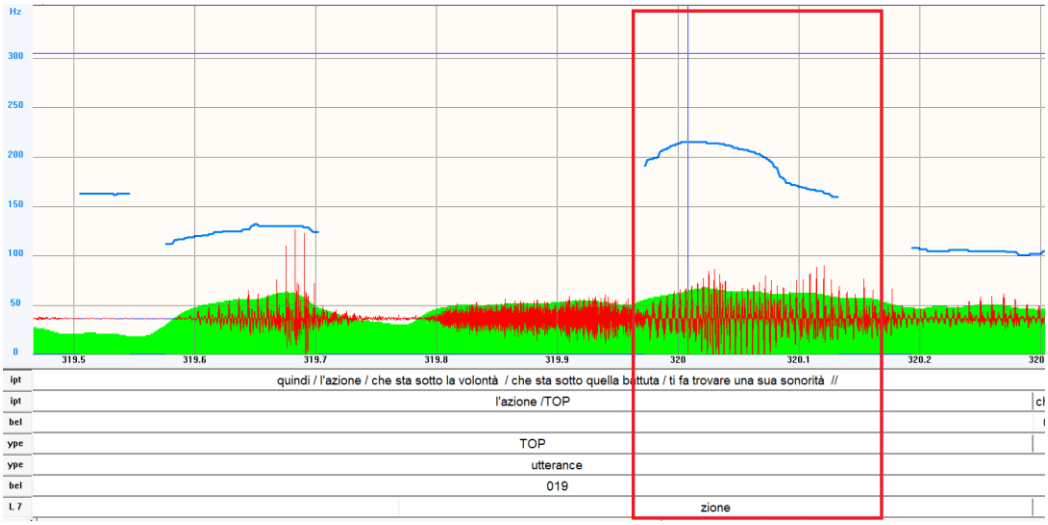

Figure 5b: Annotation of prominences on the f0 track of (5)

3.2.2 The Comment function: the illocutionary information as the gesture's affiliation

The synchrony between the gesture and the Perceptively Relevant Contours with functional values characterizes all types of pragmatic gestures found in the dataset. Figure 6a shows a gesture affiliated with the illocutionary information conveyed by the Comment in (6).

(6) sono molto diversi /^{COM} l'uno dall'altro /^{APC}
'they are very different/ ^{COM} from each other /^{APC}

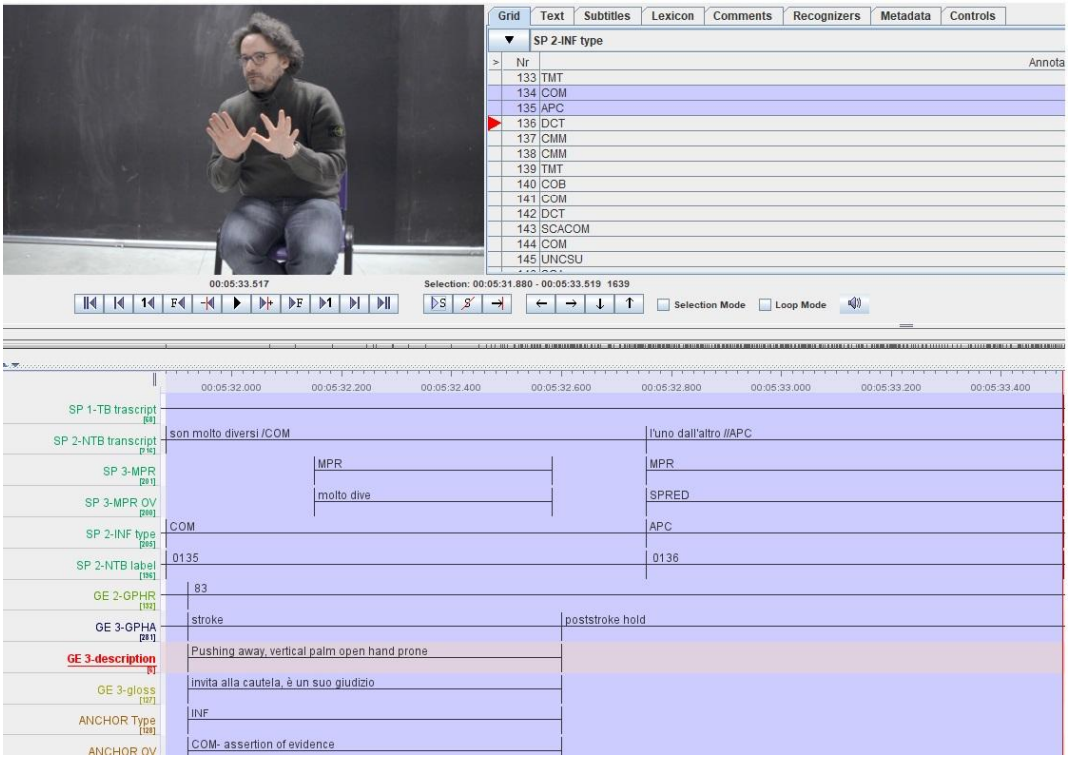

Figure 6a: Stroke affiliated with the illocutionary information in the Comment unit in (6)

The gesture depicted in Figure 6a can be described as *a movement directed toward the interlocutor, vertical palm, open hand prone* (stopping gesture). Again, we cannot interpret the gesture in affiliation with the lexical concepts of the IU ("different", "very", or "each other," but it can be interpreted when considering the illocutionary value of the utterance, which is an *assertion expressing a personal evaluation by the speaker* (Evident Assertion in the L-AcT taxonomy by Cresti, 2020). Roughly speaking, the gesture may mean something like: 'Caution, it is my judgment.' In this interpretation, the gesture supplements the speech-act performance with a precautionary sign, and its pragmatic function could be classified as 'interactive stance-taking', whose scope is specifically the illocutionary activity.

Figure 6b shows the rapidly falling nuclear part on the last tonic syllable of the Comment, which characterizes this type of assertion.

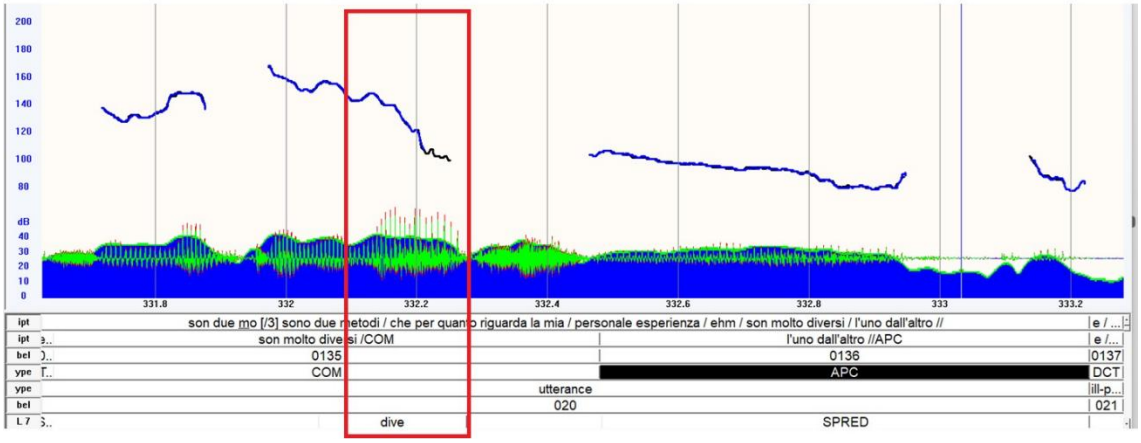

Figure 6b: Annotation of prominences on the f0 track of (6)

Looking at the gesture/prosody synchrony, the gesture extends over the unit, reaches the MPR, and extends in post-Stroke hold over the subsequent Appendix. So, we have full synchrony Stroke / Prominence / Affiliation, where the affiliate is the illocutionary information specified by the prosodic prominence[15].

We identified seven gestures that find affiliation in speech acts of different illocutionary classes and types recorded in the L-AcT taxonomy: Directives (Recall), Expressives (Contrast, Obviousness), and Representatives (Assent, Evidence assertion, Ascertainment, Conclusion). All of them almost overlap with a Comment IU and, like in this case, reach the MPR expressing the illocutionary activity.

The synchrony with the one MPR constituting the prosodic nucleus of the root unit appears as a compulsory condition for gestures affiliated with the illocutionary value. These gestures cannot be foreseen in synchrony with emphasis.

3.2.3 Listing and Illocutionary Patterns

Listing is one of the basic patterns of speech (Couper-Kulen & Selting 2018; Saccone & Panunzi 2020; Kahan et al. 2019). Various Comments or Topic (2 or 3 IUs) are gathered in a prosodic pattern whose function is to make these IUs a list, as in the list of Topics in (7) (Mittmann 2012).

(7) giocando con le pause /^{TPL} con i ritmi /^{TPL} con i toni /^{TPL}
'playing with pauses /^{TPL} with rhythms /^{TPL} with tones /^{TPL}

[15] Again, the synchrony with the MPR does not predict its nature; the speaker could have performed a gesture affiliated with the word 'diversi' (different), that is, the word bearing the MPR.

The MPRs, marked in the red boxes, played one after the other, are sufficient to give rise to the listing effect.

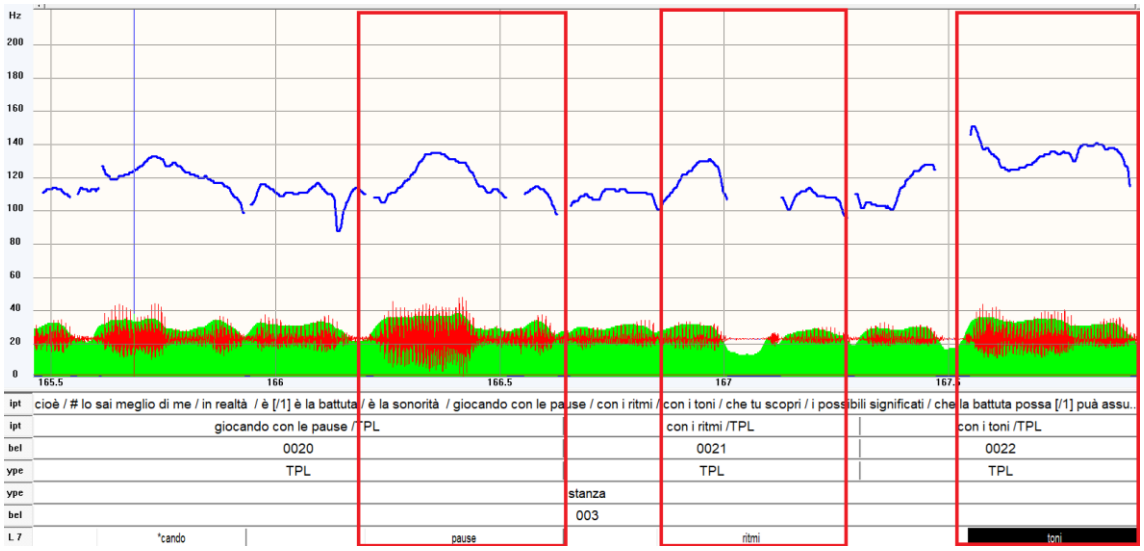

Figure 7b: Annotation of prominences on the f0 track of (7)

A variety of repeated movements can accompany listing (Dressel et al. 2023; Dankel & Satti 2019); here, as shown in Figure 7a, the gesture, annotated as a single Stroke, comprises three apexes through a multi-segment phase,[16] *alternative right/left/right movement of wrists relative to the torso changing the flat hand palm orientation,* each one in synchrony with the corresponding MPRs of the list.

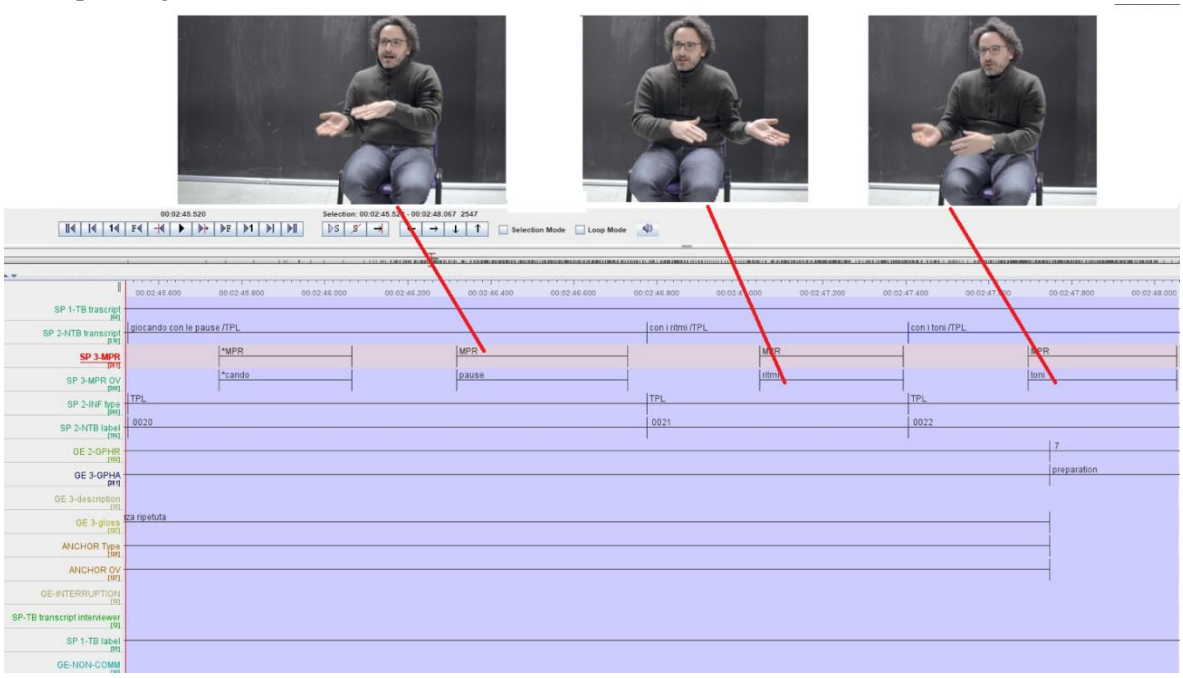

Figure 7a: Stroke series affiliated with the Topic List in (7)

[16] Multisegmented phases are characterized by abrupt direction changes, keeping the same velocity; Kita et- al.1998:29-30).

Looking at the affiliation of this gesture, we can only verify that it does not have its scope on the concepts that are listed, i.e 'pause' (pause),'ritmi'(rhythms), toni'(tones), but it is directly interpreted if affiliated with the activity of making a list.

The listing strategy does not regard Topic units only, but is better framed within Illocutionary pattern constructions, where Comment units are linked in one prosodic model.
Our dataset records four types of illocutionary patterns: listing, reinforcement, alternation, and combination. In (8), the subject lists two components that characterize the choice of an actor when interpreting a text:

(8) è la battuta /^{CMM} è la sonorità /^{CMM}
' It is the line /^{CMM} it's sound /^{CMM}'

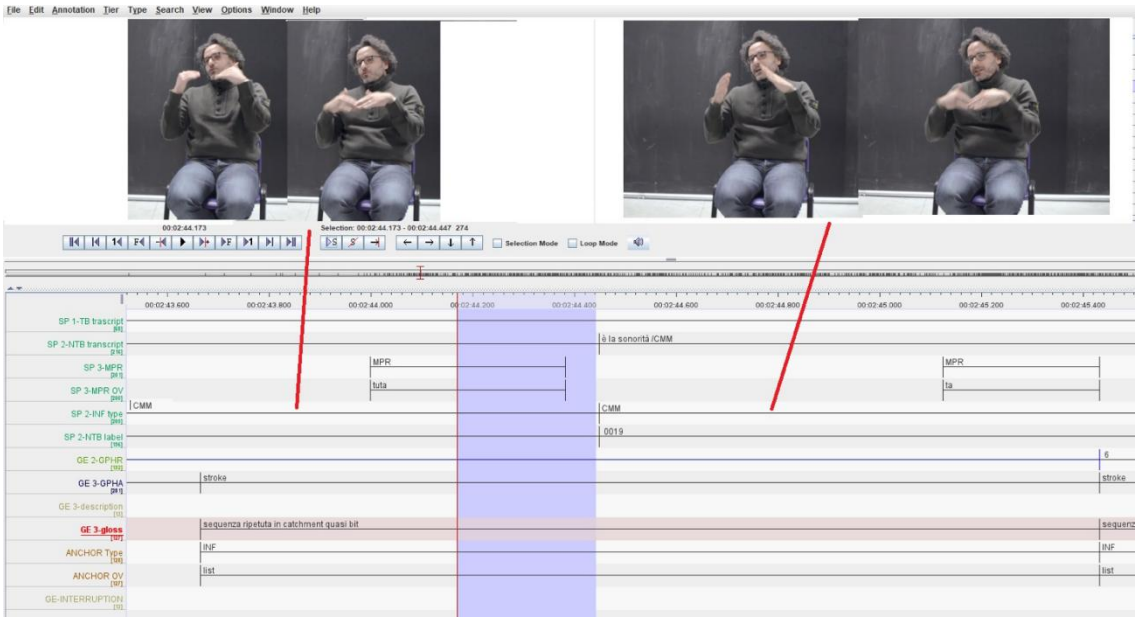

Figure 8a: Stroke series as Catchment affiliated with the Listing pattern in (8)

The functional concept of the list is expressed by repeating the same gesture as Catchment (McNeil et al., 2001), marking the cohesion of the Illocutionary pattern structure.[17] As Figure 8a shows, the gesture consists of *palm-down flat hands realizing a circular up-down movement towards the speaker's body.* Each phase is synchronous with each CMM and reaches its prosodic prominence. Therefore, the catchment strategy finds its specific value when affiliated with the listing activity.

[17] A *catchment* is recognized when one or more gesture features occur in at least two (not necessarily consecutive) gestures.

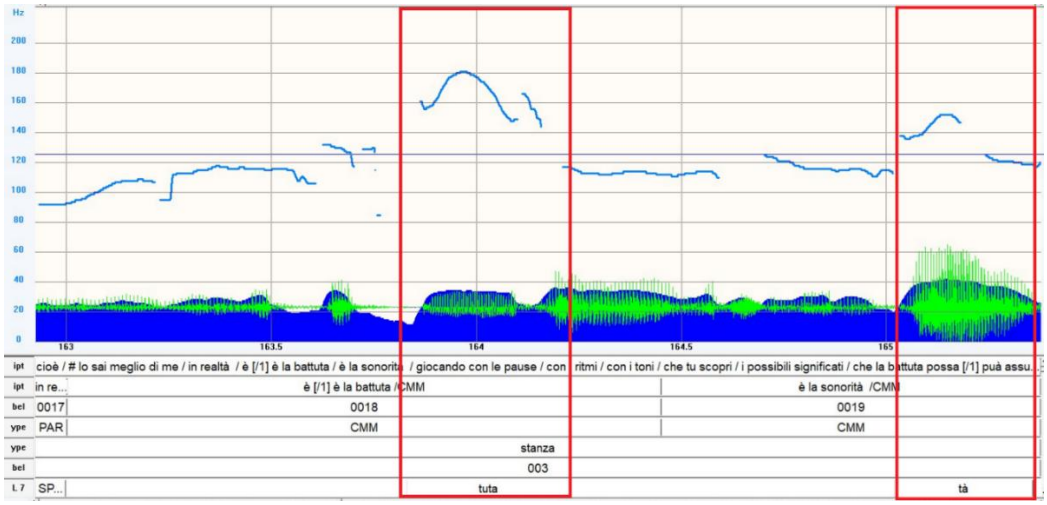

Figure 8b: Annotation of prominences on the f0 track of (8)

Notice that the gesture could be considered a sequence of beats. However, the synthetic nature of the gesture must be taken into account. These gestures also possess semantic qualities: movements have direction (top-down), so acting the term 'battuta' (line), and extend over time and space, like speech and sound. Therefore, the gesture is affiliated with the listing function as well as with lexical elements.

This possibility is even more evident in the following example. The illocutionary pattern in (9) expresses an alternative between two qualities (opposition) characterizing the outcome of an actor's performance when interpreting a line.

(9) è comprensibile /^{CMM} ma non esplicito /^{CMM}
'it is understandable /^{CMM} but not explicit /^{CMM}'

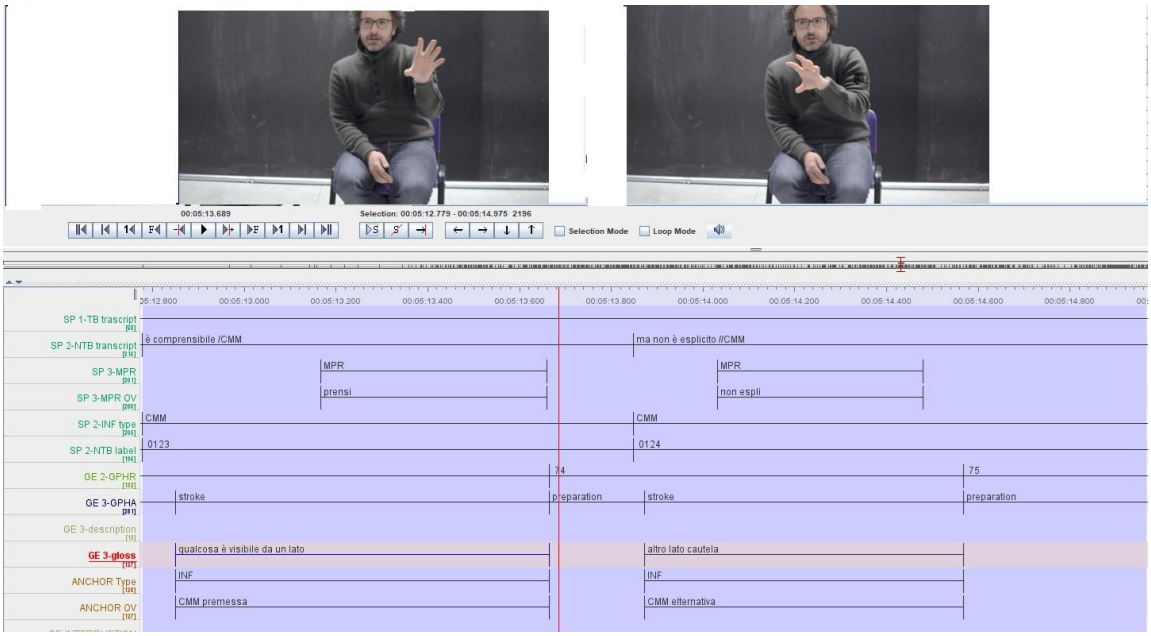

Figure 9a: Stroke series affiliated with the Alternation pattern in (9)

The information structure links two Comments within an illocutionary pattern, comparing two alternatives. The gesture highlights this link, marking the two sides of the comparison through *two hand movements, respectively, on the right and left sides of the torso, mid-hair*.

Figure 9b shows the prosodic pattern and the MPRs of the two CMMs. Both gestures are synchronous with these MPRs and are pragmatic gestures affiliated with the comparison pattern.

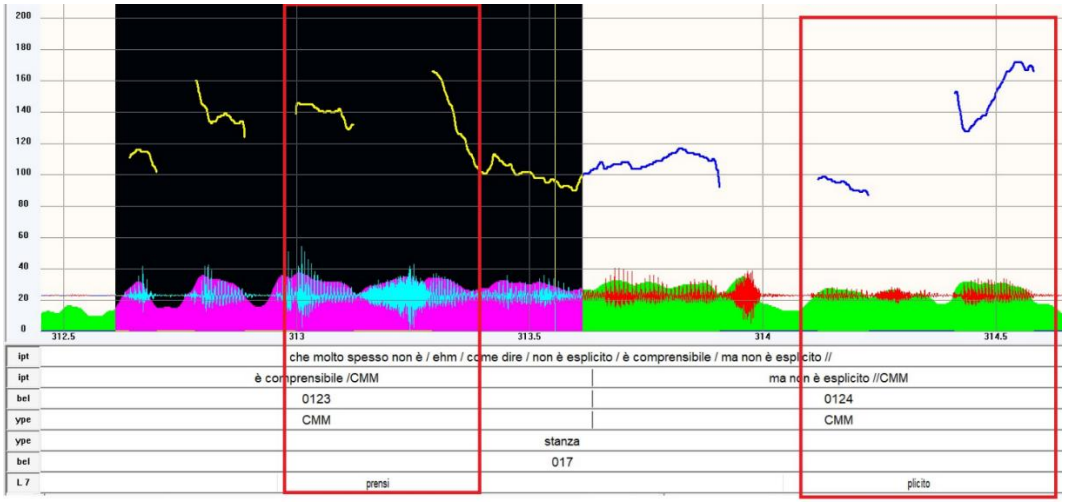

Figure 9b: Annotation of prominences on the f0 track of (9)

However, relying on visual cues, the gesture picks up more than one feature simultaneously in a synthetic manner. In particular, the form of the hand (*from bunch* to s*pread flat hand towards the recipient*), in connection with the word 'comprensibile' (understandable), leads to adding the feature 'shed light on' to the movement, which selects the first side of the comparison. In other words, the gesture is also affiliated with the meaning of one word.

On the other hand, in the second movement *from right to left, the open hand moves down toward the interlocutor* and rather expresses the value 'caution, it is my point of view', which is instead affiliated with the illocutionary value of *assertion taken for granted* of the second CMM. Therefore, the concept of Linguistic affiliation is a valid tool for disclosing gesture semantics, but it should consider the synthetic nature of visual information, allowing for multiple affiliations.

3.2.4 Parenthesis Function

The information function of Parenthesis is marked by prosody through a lowering of f0 range with respect to the other units of the information pattern (Dehé 2014; Saccone & Panunzi 2021; Saccone & Trombetta 2021); it is inserted in the utterance as a metalinguistic commentary, passing from an objective point of view to the subjective point of view of the speaker (Tucci 2010; Tucci 2011) and is not compositional within the utterance. The prosodic profile is flat but may be marked by some emphasis.[18]

(10) is an example of parenthetical insertion (white part of the f0 teach in Figure 10b), which adds an explicative commentary after a Bound Comment (black part of Figure 10b). The IU shows a lengthening of the syllable in the red box, marking with *MPR the word 'schematici' (schematic).

(10) apparentemente non logici /^{COB} che vuol dire non schematici /^{PAR}
'apparently illogical /^{COB} that means not schematic /^{PAR}'

[18] Parentheticals have been recently considered at the utterance level instead of at the IU function level (Saccone & Trombetta 2021). We did not find parenthetical utterances in this data set, so we skipped the point.

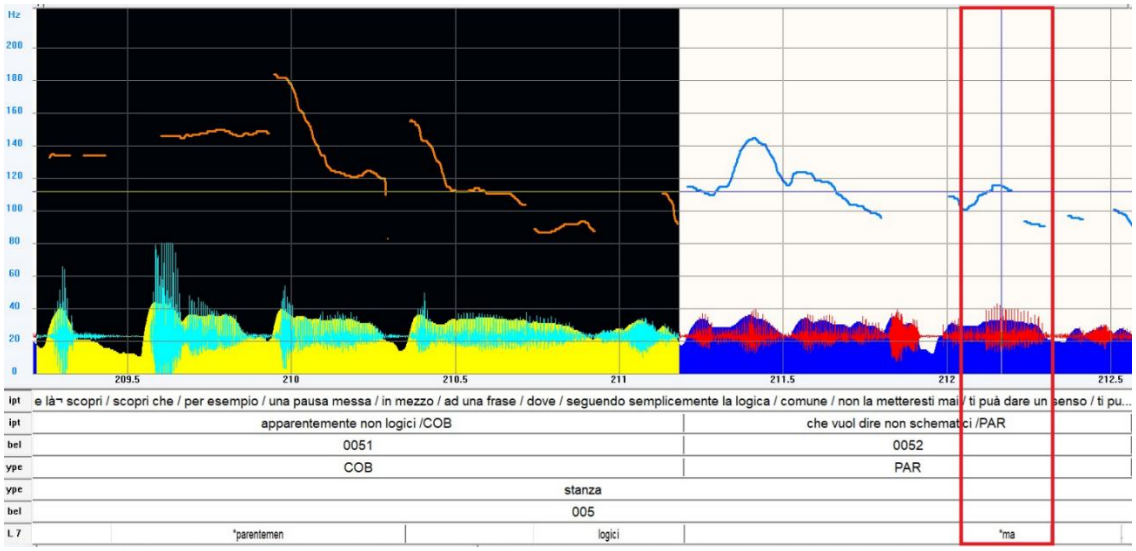

Figure 10b: Annotation of prominences on the f0 track of (10)

As Figure10a shows, while performing the Parenthesis, the subject *orients both hands toward his torso, palm-up open hand, fingers 2-5 bent, finger 1 stretched*. The gesture synchronizes to the unit and overwrites the *MPR.

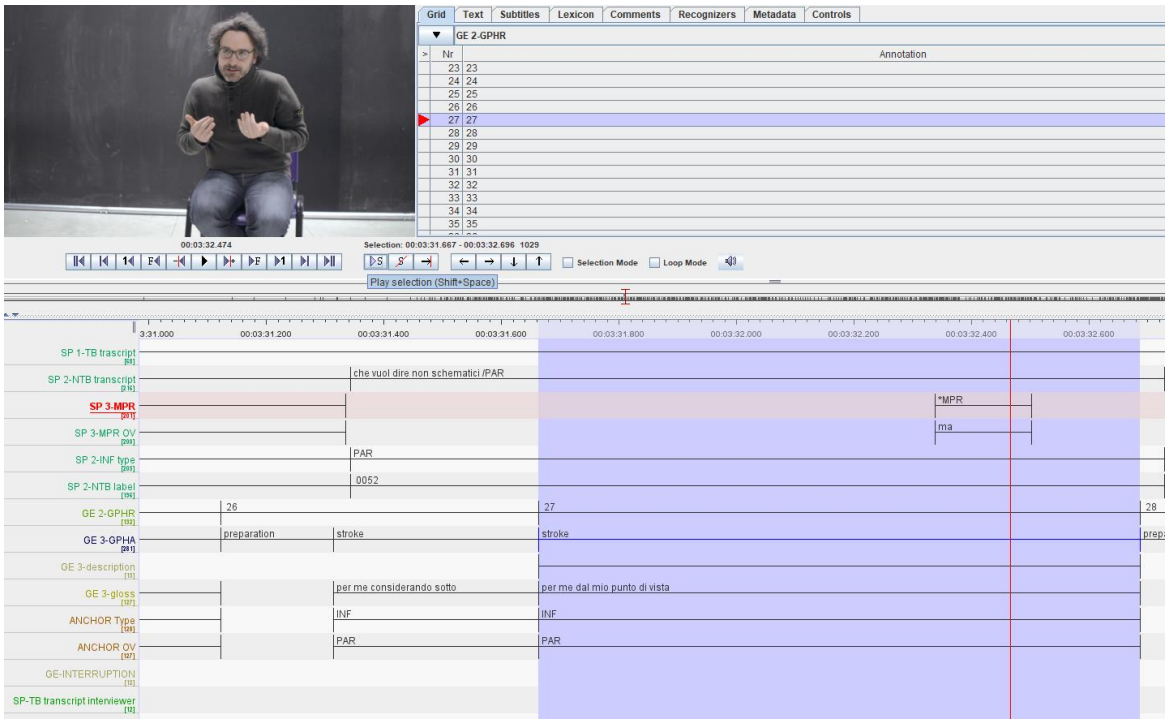

Figure 10a: Stroke affiliated with the Parenthesis function in (10)

Again, the gesture cannot be interpreted if referred to the words or the phrase 'che vuol dire non schematici' (that means not schematic). Still, it is immediately understandable if affiliated with the parenthetic function of the IU. Orienting the hands deictically toward himself, the subject means something like 'said from my point of view.' Thus, the deictic gesture is pragmatic, as it signals to the addressee the informational value of the IU expressed by prosody; i.e., the change in perspective of the IU content.

3.3 Information unit functions not anchoring pragmatic gestures (SCA, Interrupted units, Appendix, and Discourse Connectors)

As we might expect from the L-AcT perspective, we did not find any pragmatic gesture affiliated with SCA, whose function is played out at the locutive level and lacks pragmatic relevance. SCA units only guest gestures affiliated with lexical content or bearing a modal value (see below).

The lack of pragmatic function also regards interrupted units of Topic and Comment, which may have lexical relevance but do not bear the MPR, which is placed on the right side of the prosodic unit, after the interruption.

Both cases confirm that pragmatic gestures strictly correlate with a Perceptively Relevant Prosodic Contour bearing a functional value and must synchronize with units that play an information function. However, we did not find pragmatic gestures in synchrony with information units that serve as Discourse Connectors, or in the few Appendix units. In other words, no gesture in the data set was affiliated with the function of "connecting a unit to something previously said" nor to the function of "adding supplementary information", which, in principle, might be possible.[19]

It may be interesting, however, to notice that when the Discourse connection is coordinated with a gesture, the gesture goes in parallel with the semantics of the connector. For instance, the disjunctive value of 'ma' (but) in (11) is synchronous with *a rapid movement of both hands from the center to the left side*. The gesture can be interpreted as 'let us consider the other side', and is affiliated with the meaning of 'ma'(but).

(11) non schematico /^{COB} **ma** /^{DCT} questo non vuol dire /^{SCA} che non sia logico //^{COM}
'not schematic /^{COB} **but** /^{DCT} that doesn't mean /^{SCA} it's not logical //^{COM}'

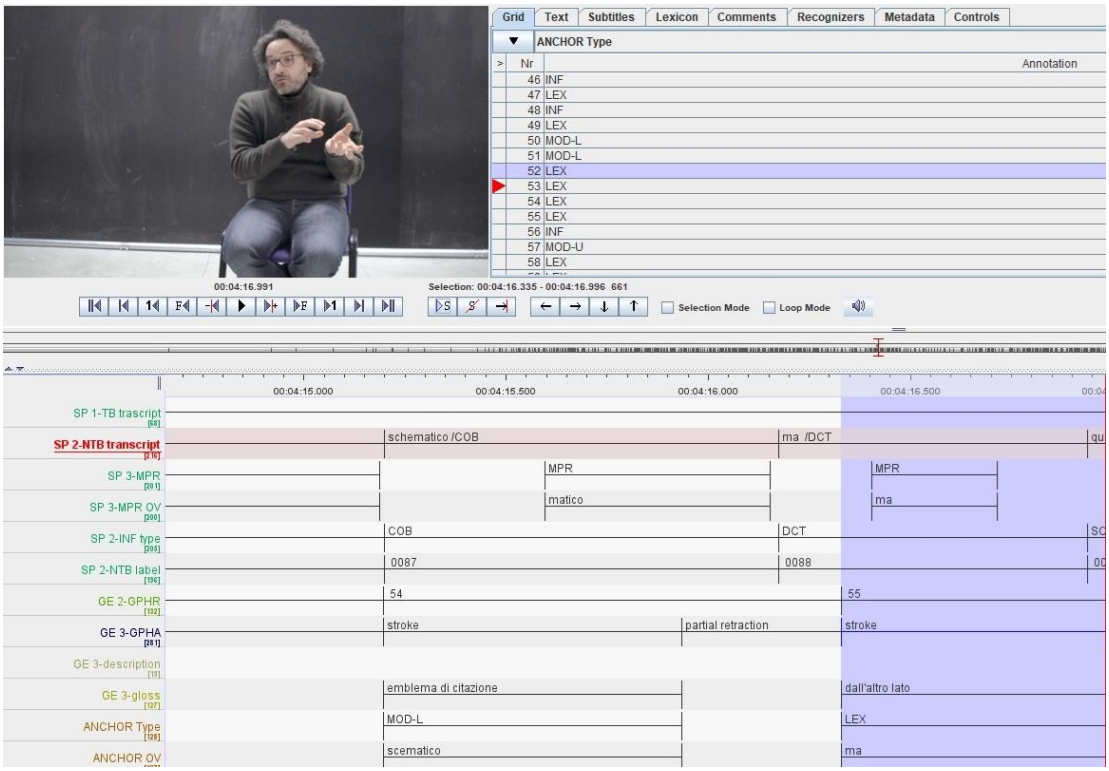

Figure 11a: Stroke affiliated with the lexical information of the DCT unit 'ma' in (11)

[19] The data set is too restricted to derive any conclusion from this lack. Indeed, the possibility of gestures affiliated with the Appendix function was hypnotized in a previous study (Moneglia 2024).

In example (12), the DCT function is played by the conjunction ‘perché’ (because), which has a different meaning, and the DCT is synchronous with a gesture affiliated with this meaning.

(12) in realtà aveva senso //^{COM} **<u>perché</u>** /^{DCT} dava un ritmo alla battuta […]
‘it actually made sense //^{COM} **because** /^{DCT} it gave a rhythm to the line […]’

Figure12a depicts *a mid-air gesture in the upper center space, left palm down hand, fingers 1-5 crooked*. The subject designs a point in space, which in connection to the conjunction ‘perché’(why) indicates a concept that turns out not obvious, i.e., ‘the reason why.’ Therefore, by changing the semantics of the DCT, the gesture also changes and can be interpreted accordingly.[20]

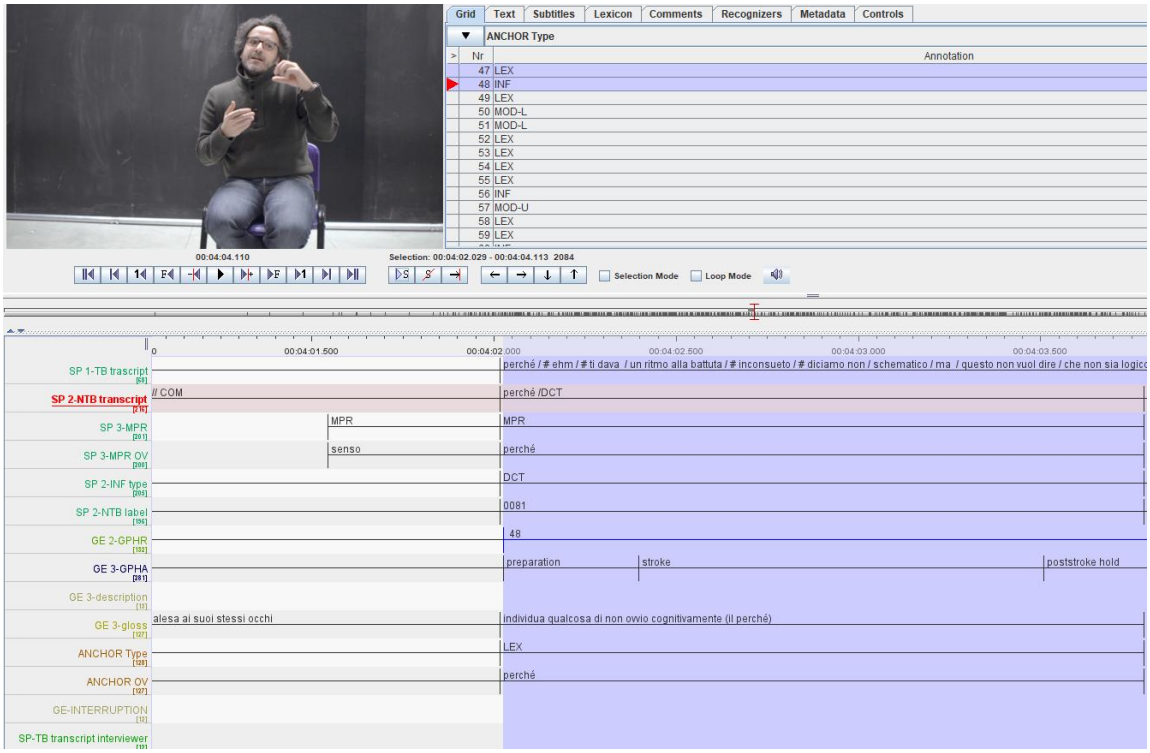

Figure 12a: Stroke affiliated with the lexical information of the DCT unit in (12)

3.4 Modal values

When searching for the linguistic affiliation, gestures bearing a modal value, e.g., the stance taken by the speaker (Biber & Finegan 1989) emerge clearly. (13) is a chain of two Comment units linked within an Illocutionary pattern of Reinforcement, and each unit of the pattern shows a gesture conveying an epistemic modality.

(13) **possono funzionare entr<u>ambi</u>** /^{CMM} **sono straordinari entr<u>ambi</u>** /^{CMM}
‘they can both work /^{CMM} **they are both amazing** //^{CMM},

[20] The affiliation of the gesture to the meaning of the lexical expression working as DCT seems to contradict the proposal made by Cresti and Moneglia (2021) that DCT units can be substituted one to the other and are, for this reason, semantically empty. However, the question has a further complexity since also the gestures synchronous to ‘ma’ (but) and perché (because) in (11) and (12) can, in principle, be substituted the one to the other.

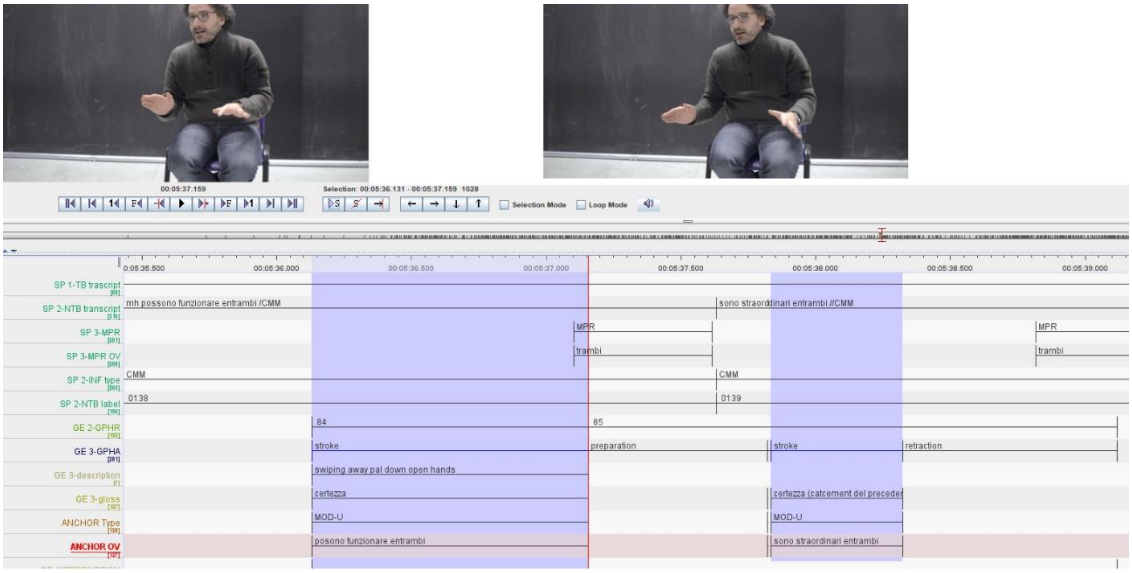

Figure 13a: Strokes with Modal-Epistemic value in (13)

The same gesture, *sweeping away palm down open hands,* is repeated in the two units, marking their integration in the same prosodic model. The gesture ends at MPR, and, as Figure 13b shows, the f0 presents a final rise, which characterizes the illocution of Assertion taken-for-granted (Cresti, 2020).

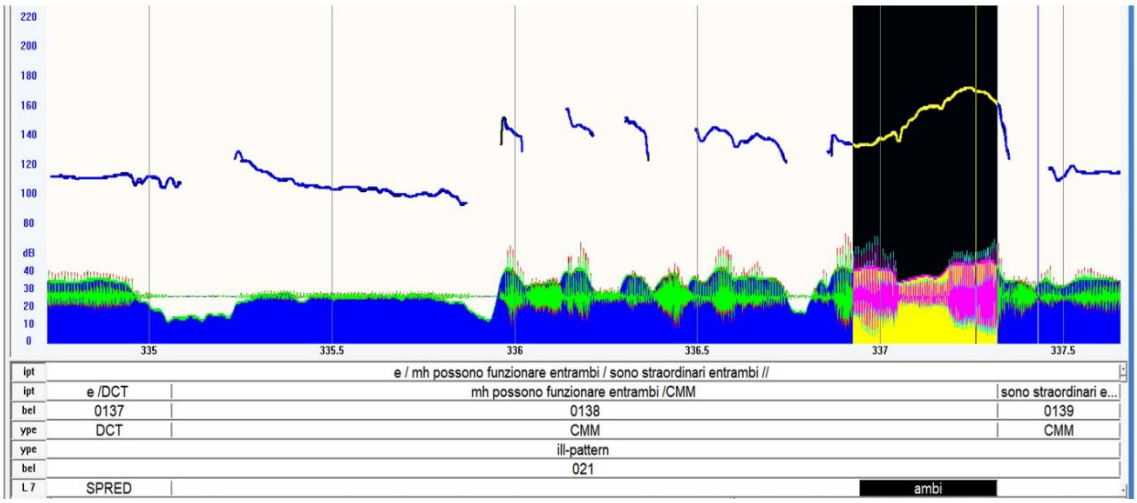

Figure 13b: Annotation of prominences on the f0 track of (13)

Looking for the affiliation, however, none of the lexical items ("work"; "amazing") allows for its interpretation. The gesture expresses an epistemic stance of 'certainty' on the propositional content of the assertion. The underlying metaphor is 'no alternative; all swept away.'[21] Therefore, the gesture works like the speaker's modal evaluation of 'certainty' added to the linguistic content of the IU.[22] No linguistic device expresses the modal evaluation; it is just added to the multimodal utterance through visual cues.

[21] The gesture is similar in meaning to that in (1) but is performed with both hands and does not depict a crux. Two different ways of expressing the same meaning (Frege 1892)

[22] It may be interesting to observe that the gesture is repeated in a *catchment* in the second CMM unit of the illocutionary pattern, strengthening the cohesion of the illocutionary pattern construction. In this case,

Gestures conveying a modal evaluation may concern a single word, instead of the overall value of the utterance, as just seen in (13). For instance, (14) is a Comment IU. The speaker concludes that the ways of segmenting a line can appear strange, but it is not illogical. The gesture in Figure 14a is the standard emblem for something said 'in the quote.' The gesture is synchronous with the word 'apparentemente' (apparently).

(14) [i momenti in cui tu segmenti la battuta sono] **apparentemente** non logici //^{COM}
[the moments in which you segment the line are] apparently not logical //^{COM}

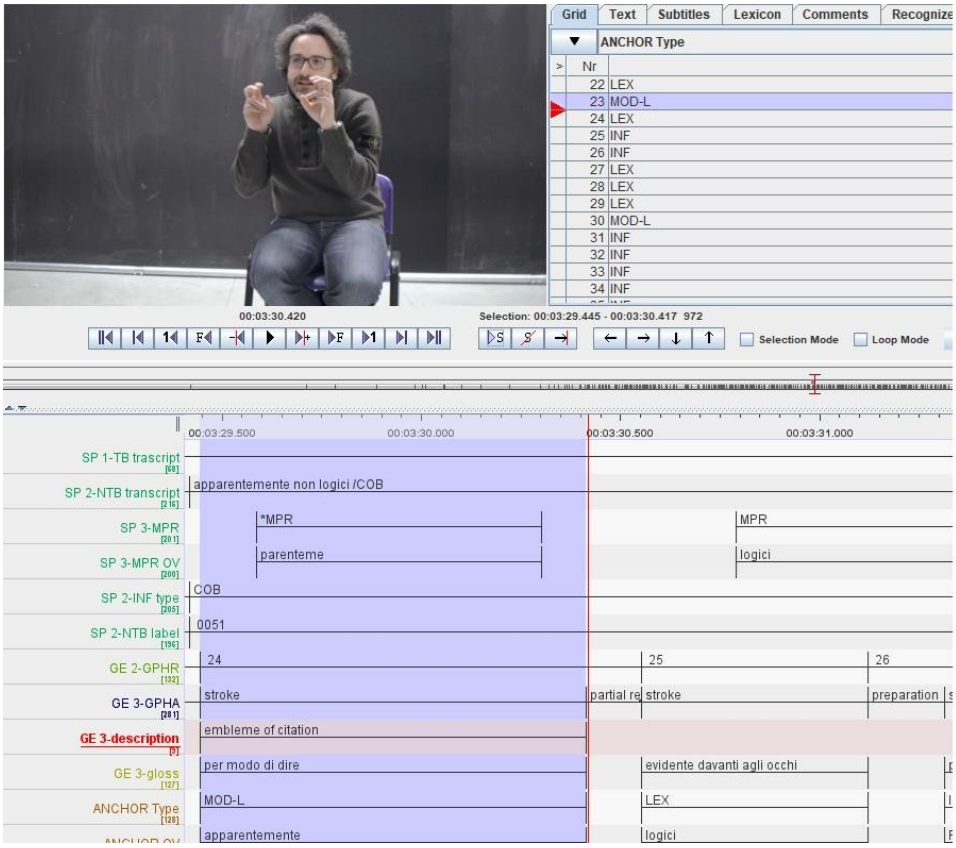

Figure 14a: Emblem of citation as a Modal gesture in (14)

The emblem of citation does not regard the full content of the Conclusion; it is precisely the word 'apparentemente'(apparently) that is said in a multimodal quote.

In parallel, the prosodic performance of the word synchronous to the gesture clearly emphasizes the lexical choice with *MPR. As Figure 14b shows, the word is lengthened, and its f0 jumps to a high pitch (in red box), while the MPR conveying the pragmatic force of the conclusion is the final falling movement (in red circle).

Therefore, we also get complete synchrony among the Stroke, the *MPR marking the lexical choice, and the lexical affiliate to which the modal evaluation is added through visual cues.

Considering both types of modal gestures (scope on the IU and scope on one word), we recorded a variety of modal meanings expressed through gestures, covering the main stance types: Epistemic stance (Possibility, Surprise, Certainty, Uncertainty); Relevance stance (Preciseness, Approximation); Intersubjective stance (Caution, Compliance).

the synchrony with the prosodic prominence is missed. This seems a typical consequence of catchment phenomena (see below).

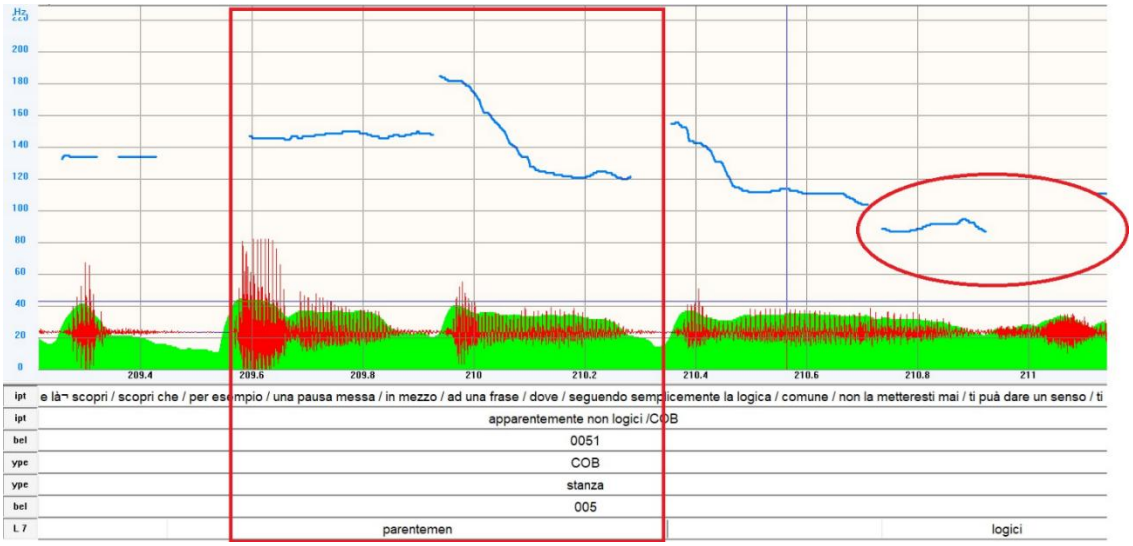

Figure 14b: Annotation of prominences on the f0 track of (14)

3.5 Gestures in silent pauses

We showed that finding the affiliation of a gesture in speech is a key path to its interpretation. However, gestures may also occur and be affiliated to a time-taking unit, which, in principle, from the point of view of speech, are not units of information. Considering the time-taking units (TMT) not filled by linguistic material in the dataset, three of them co-occur with a gesture. For instance, during the 600 ms pause in the speech flow, the subject takes time to restart his discourse, and this silence is filled by raising both index fingers at the upper center toward his *head*.

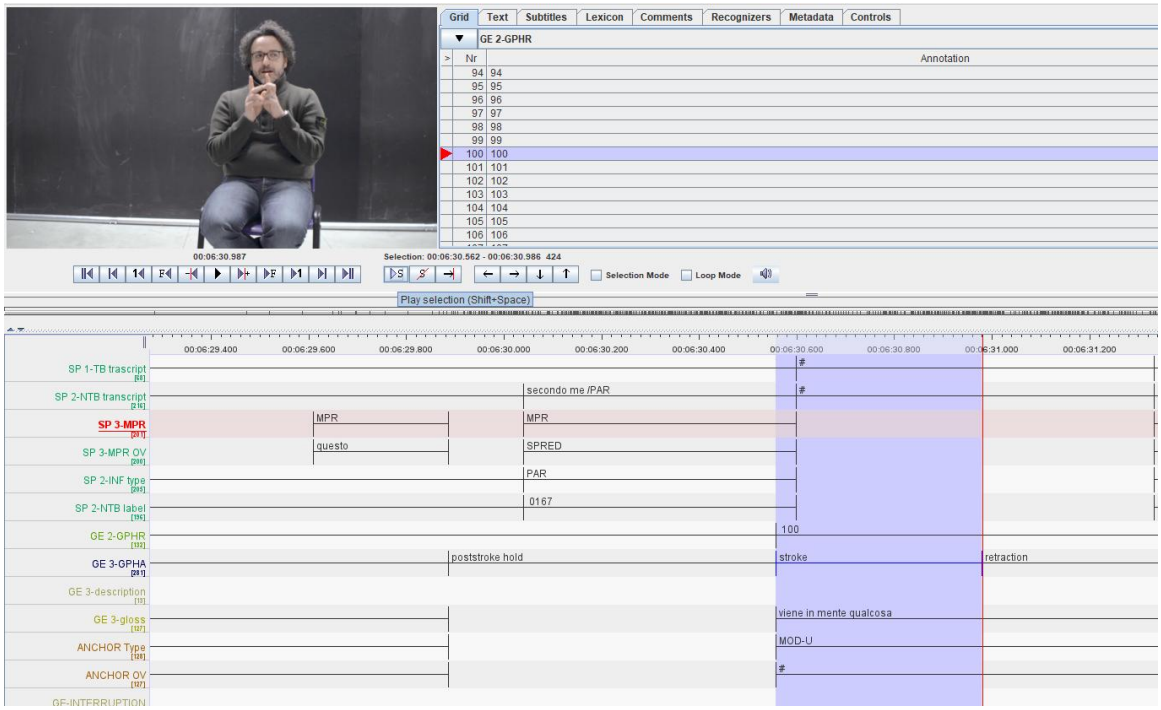

Figure 15a: Stroke affiliated with the meaning of a pause

The gesture *per se* cannot be considered an emblem with conventional meaning, but it can be interpreted in connection with the fact that it occurs during the pause of a monologue. If we consider the Stroke affiliated with the reason for the pause, the gesture gets a meaning: ‘an idea

may come to my mind' / 'there is something I may say'. Therefore, due to the inclusion of visual information, the silent unit becomes a multimodal unit of information.

3.6 Quantitative and qualitative results

3.6.1 Types of affiliations and their relation to IUs

The first result of the research concerns the type of linguistic affiliations that we found in this monologue, described by the histograms in Figure 16

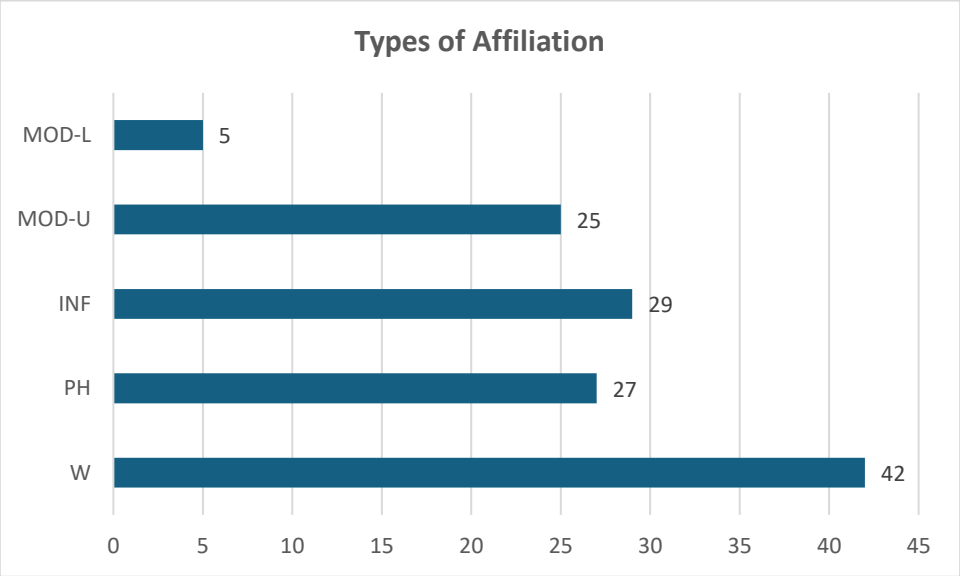

Figure 16: Affiliation of gestures in the dataset by type

Considering 128 GPHRS, 69 were found to be affiliated with locutive information (word or phrase), but a large number of gestures (59) concern non-lexical information. In this set, half are pragmatic gestures affiliated with information functions, and half are modal gestures, which add evaluation in the visual modality. Therefore, the strategy of affiliating gestures with lexical information prevails; however gestures that find scope in the information underlying the locutive phrasing (modality and information function) constitute a substantial set, and all types are consistently represented even in this small dataset.

Figure 17 details the relation between types of affiliation and the IU types foreseen in the L-AcT approach. Comment, Topic, and Parenthesis IUs guest gestures that find affiliation of all kinds (Locutive, Functional, and Modal). However, this is not the case with Discourse Connectors and Appendices, which do not affiliate with modal or pragmatic gestures.

It is crucial to notice again that the IUs with no functional values and no MPR (Scanning units and Interrupted units) never affiliate a pragmatic gesture. This confirms one of the assumptions of the L-AcT framework. Scanning IUs refer to the same locution of the unit they parse and do not constitute an independent information unit with a specific information function.

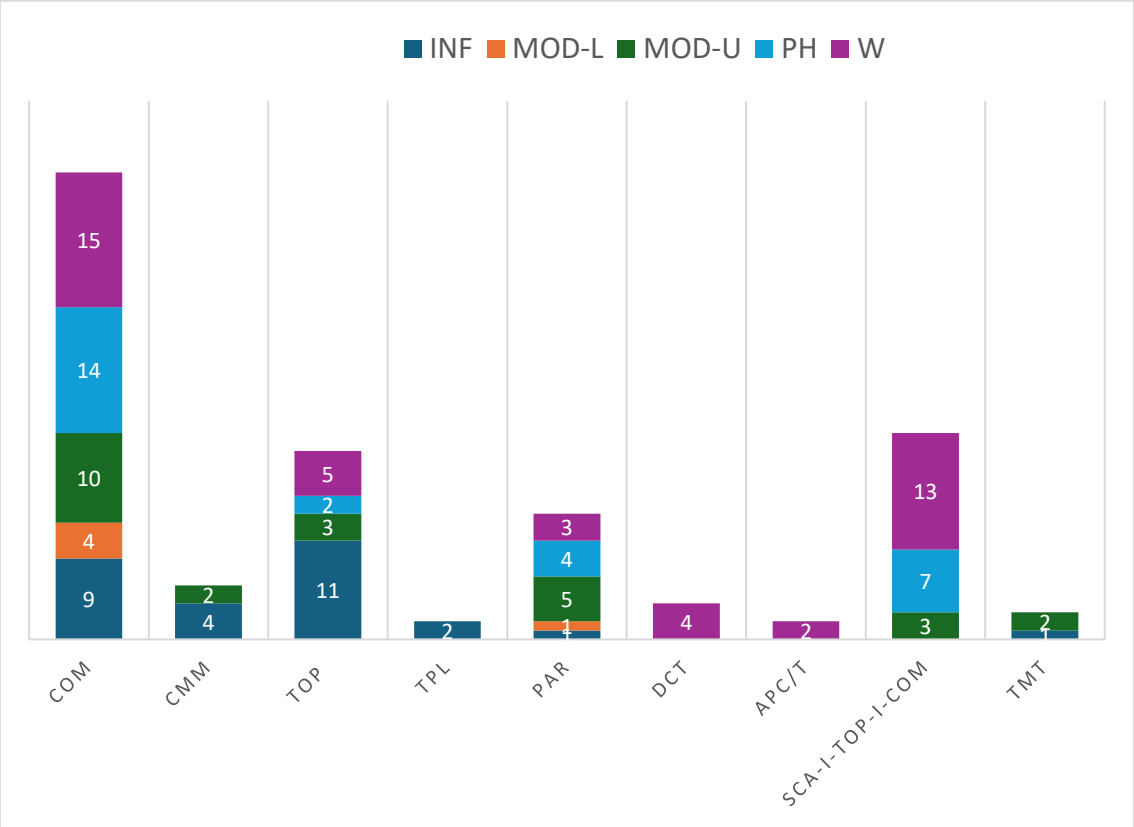

Figure 17: Affiliation of gestures by type for each IU type

The above figures indirectly highlight the strict relationship between IUs and gestures. As far as we consider prosodically marked IUs as the linguistic counterpart of the flow of thought (Chafe 1994), it is not surprising that gestures are found systematically affiliated with the linguistic content of the IU to which they are synchronous or, in other words, get their scope in that IU, referring to one or more characters of the underlying *idea unit* from which gesture and speech origin.

From the analysis of our dataset, it seems that gestures are never affiliated backward or forward with respect to the IU they synchronize with. The prediction, however, may appear too compromising since the gesture/IU relation is not necessarily one-to-one. In principle, the following two cases might conflict with this assumption and deserve close attention.

First case. Sometimes, one IU can synchronize with more than one Stroke. We found five cases in the dataset where two Strokes occurred in the same IU. As far as the Strokes are affiliated with the linguistic information of the same IU, this does not conflict with the stated principle. However, the observation of these contexts has theoretical relevance. Under the L-AcT assumption that the MPR expresses the functional value of an IU and that only one MPR can occur in the same IU, if it is true that pragmatic gesture must synchronize with an MPR and is also true that a Stroke must find its affiliation in the IU in synchrony, then, in no circumstance does one IU guest two pragmatic gestures. This sounds reasonable, and although a larger dataset is necessary to confirm this prediction, our observations corroborate this restriction on gesture distribution.

Second case. In the gesture/IU relation, we found various Strokes extending over two IUs. This, in theory, conflicts with the above principle, which should be downgraded to a tendency. However, observing these contexts, we see that the exception proves the rule. Strokes extend over

two IUs only in two cases, and in both cases, the two prosodic units share the same functional value:

a) when the units belong to the same underlying idea, as in SCA/COM; SCA/TOP; i-TOP/TOP;
b) when the gesture is repeated in a catchment, marking the foci of an illocutionary pattern or a list (as in Figure 8), it can be considered a pragmatic gesture marking the cohesion of the illocutionary pattern.

Although larger data sets are needed to verify the extent to which the affiliation relation Stroke /IU holds, the positive and negative correlations observed confirm the claim that the affiliation of the Stroke in the Information unit in synchrony is more than a quantitative tendency.

3.6.2 Asynchronies

The synchronization relationship between Gesture, Affiliate, and Prosodic prominence is substantial. Considering the 128 Strokes, only a few cases (9) do not satisfy the complete synchrony among the three variables.

Despite the small number of asynchronies, our fine-grained analysis revealed that they pertain only to gestures affiliated with lexical entries and can be framed within informational limits and specific gesture strategies. The following possible asynchronies can, in principle, occur:

a) Semantics first: gesture synchrony to the affiliate not marked by MPR(*)
b) Prosody first: gesture synchrony to MPR(*), which does not select any affiliation
c) Complete asynchrony: gesture not synchronized to its affiliate nor to one *MPR

(15) is a context of Semantics first, relevant for its theoretical implications on the relation between prosodic structure and gesture. As anticipated in 3.2.1, L-AcT foresees that the MPR characterizing the information function of Topic is placed on the right side, on the last syllables of the IU. Under this condition, to comply with the principle that a prosodic prominence signals the affiliation of a gesture, to affiliate a gesture with a word preceding the MPR, this word should be emphasized. This is not the case in (15), where the Topic is filled by an NP whose semantic head is modified (on the right) by a PP guesting the MPR (red box in 18b).

(15) se tu hai un **gruppo** di lavoro /^{TOP}
‘if you have a **group** of work /^{TOP},

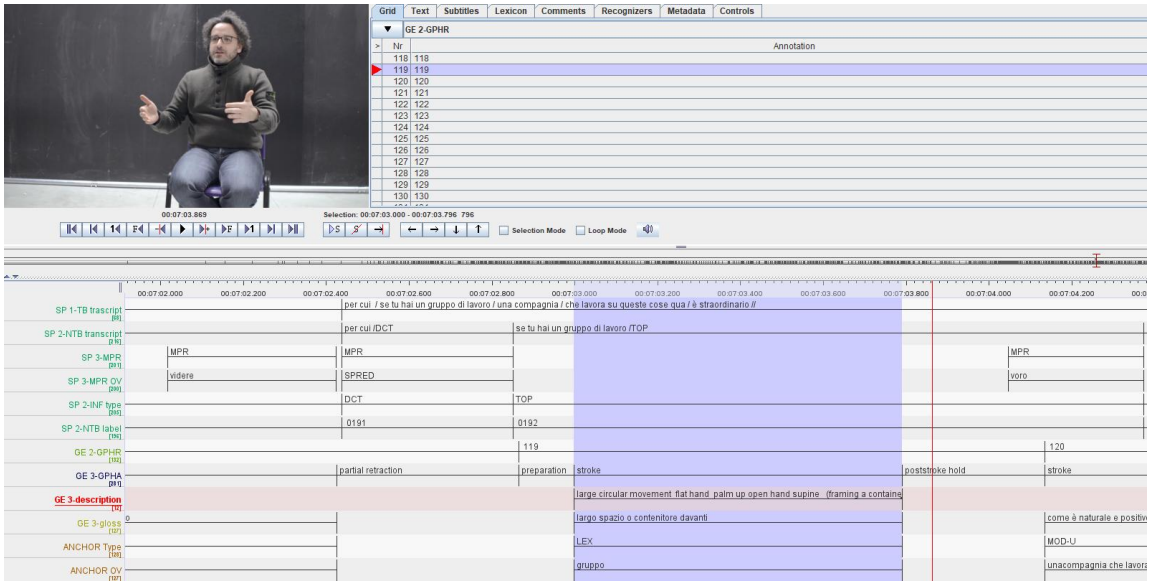

Figure 18a: Stroke affiliated with one word preceding the MPR in the Topic unit of (15)

The gesture depicted in Figure 18a is a *large circular movement, flat hand palm up, open hand supine*. The gesture is synchronous to the word 'gruppo' (group), constituting the semantic head of the phrase in Topic. In affiliation with that word, the gesture can be interpreted as 'framing a container.' So, the affiliate is a word on the left side of the unit, but, as Fig (18b) shows, the MPR is on the last word (in red box), while the word 'gruppo' (in a circle) is not emphasized.

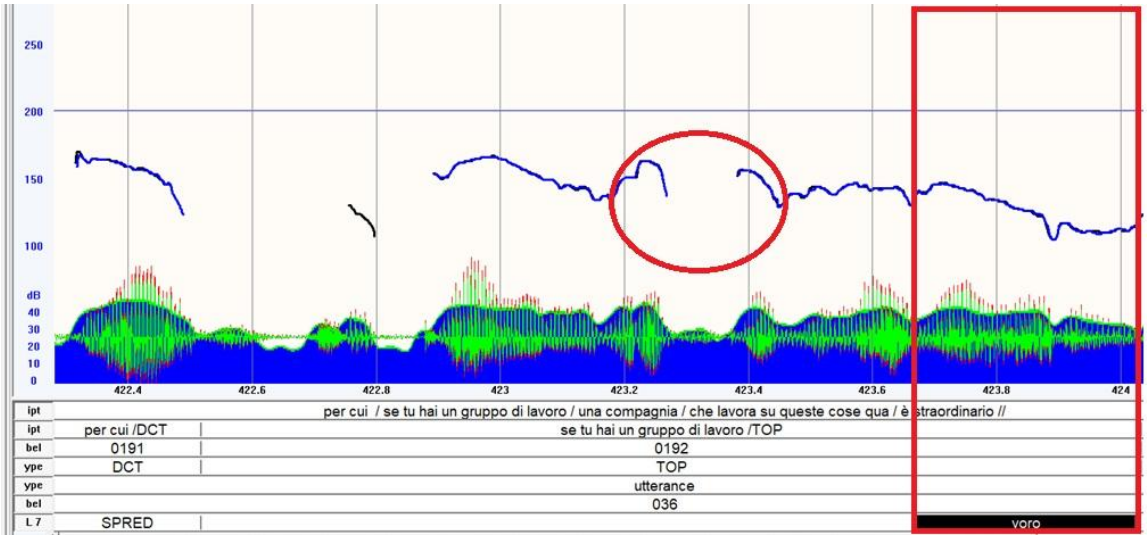

Figure 18b: Annotation of prominences on the f0 track of (15)

Therefore, the prosodic properties of the Topic IU can predict the possibility of a Semantic first phenomenon: the Stroke can synchronize with a lexical affiliate preceding the MPR, which may not be emphasized.[23]

One of the gestures synchronous with (16) replicates the same phenomenon

(16) ma se tu sei un **attore** che viene **lanciato** /^{TOP} su un **set** /^{APT}
'but if you are an **actor** who is **launched** /^{TOP} on a **set** /^{APT}'

The prosodic contour marking the Topic function of the phrase, which is headed by the word 'attore' (actor) modified by a relative clause, is on the word 'lanciato' (launched), i.e., the modifier (on the right), as in the first box in Figure 19b.

Figure 19a shows that in synchrony with the word 'attore' (actor), the subject performs a gesture identifying an object in space, while in synchrony with this MPR on the word 'lanciato' (thrown), he performs an iconic gesture *acting the launching* (Müller 2014).

[23] We can notice, however, that the gesture is kept in post-Stroke hold until the MPR. Therefore, it still follows the tendency to synchronize with a PRC in its expressive phase. This possibility is common. In five TOP units in the dataset, the gesture extends over the entire unit and reaches the MPR in its post-stroke hold phase, maintaining synchrony within the gesture's expressive phase (Kita et al. 1998).

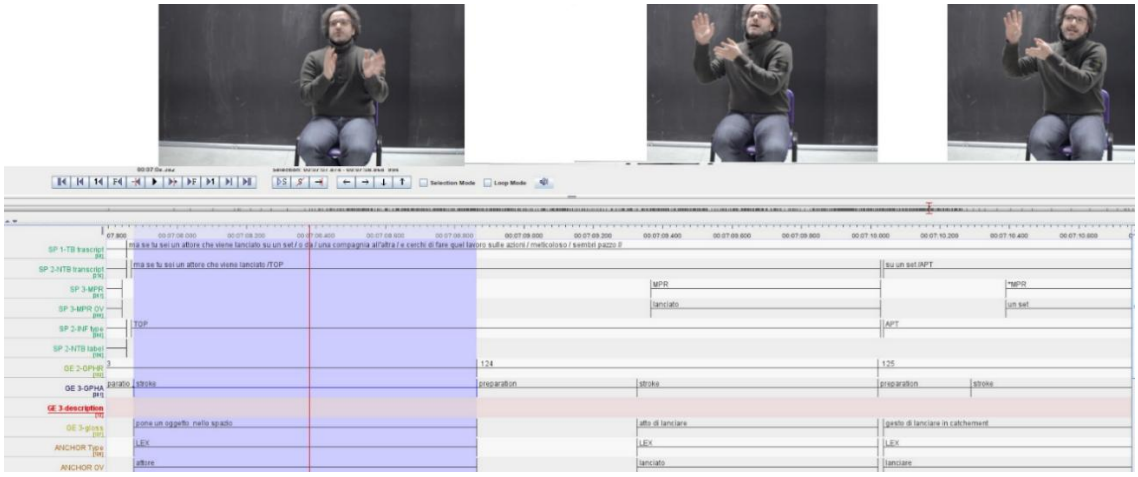

Figure 19a: Strokes affiliated with lexical content in Topic and Appendix units of (16)

So, both words are the affiliation of a gesture, but the gesture affiliated with the word 'attore'(actor) is not synchronous with any prosodic emphasis on that word (f0 in a red circle in Figure 19b).

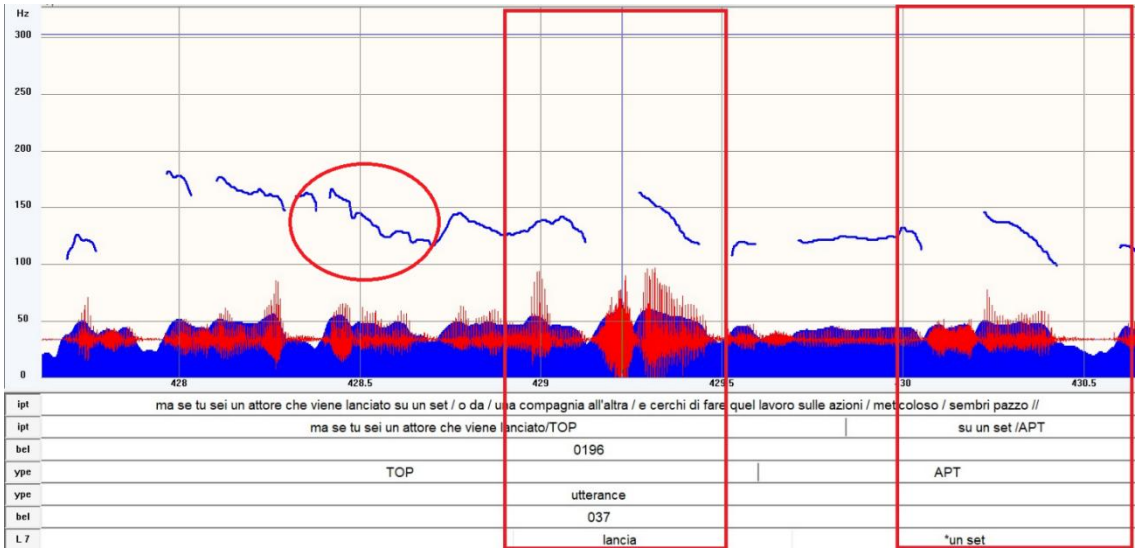

Figure 19b: Annotation of prominences on the f0 track of (16)

Therefore, the synchrony with a prosodic prominence of gestures affiliated with a word is not strictly necessary if the gesture is synchronous with that word (Semantic first).
(16), however, also presents the opposite type of asynchrony (Prosody first), which is not connected to the prosodic properties of the information structure, such as the fixed position of the MPR in Topic units but can be framed within the gesture strategy of Catchment (McNeill et al. 2001).

The gesture of throwing something in Figure 19a fully synchronizes with the Topic MPR (in the first box of Figure 19b) placed on the lexical affiliate 'lanciato'(thrown). Still, the same gesture is repeated in a catchment in the following Appendix unit, attracted by the prosodic emphasis on the word 'set' (second red box in Figure 19b). However, the word 'set' does not constitute a possible semantic affiliation for the gesture. The catchment strategy can cause asynchrony between the Prosodic prominence synchronized with the gesture and its affiliation.

4 Conclusions

Identifying the linguistic affiliation of gestures through gesture interpretation in the context in which they appear is productive. Considering all the various levels of information encoded in speech, all gestures in the data set find an affiliation. By selecting the range of linguistic information that allows for the disclosure of the metaphor underlying the gesture, we identify the type of information conveyed by gestures. Beyond lexical affiliation, pragmatic and modal gestures are so well identified.

From a quantitative perspective, we observe a tendency to affiliate gestures with the locutive content, complementing the meaning expressed by words or phrases; however, pragmatic and modal gestures are not occasional behaviors and occur with nearly the same frequency. Given that pragmatic/modal gestures are frequently Recurrent gestures, the methodology can be significant in developing a bottom-up approach for the induction of their abstract meaning.

The L-AcT approach assumes that IUs necessarily correspond to prosodic units and that prosodic prominence is connected to intentional prosodic variations conveying the information function of the unit. This frame enables the capture of relevant generalizations concerning the relationship between gesture/prosody synchrony and the possible scope of semantic affiliation. Once the affiliation is identified, we have strong confirmation that gestures are semantically connected to the information units (Unit of thought in Chafe) to which they are synchronized, neither backward nor forward. This correlation confirms that gesture and speech units originate from the same rich underlying unconscious entity (Idea unit in Kendon, Grow point in McNeill).

Gestures may extend over two IUs only when both IUs refer to the same locutive space, as in SCA units or Illocutionary Patterns, confirming the overall synchronization principle.

Within this general synchrony rule, we have evidence of a strong tendency to synchronize among the Stroke, a Perceptively Relevant Prosodic Contour, and a semantic affiliation, all belonging to the same IU. The expectation that the Stroke starts before the prominence and finds its apex when reaching it, as stated in previous research, is also confirmed, even marking prominences within the L-AcT frame.

Asynchrony may rarely occur in both directions (Semantic first and Prosody first) and specifically regard Gestures affiliated with the lexical choice. We could observe two cases that have theoretical relevance. The Semantic first case occurs when the prosodic contour has a fixed position, as in the Topic IU, and the lexical affiliate does not coincide with this position; i.e, emphasis is not compulsory for attracting gestures. The Prosody first case can be foreseen when gestures are repeated in a catchment, marking the cohesion of a construction, as typically happens in illocutionary patterns.

The general question posed by the article has a clear but not univocal answer. The synchrony of gestures with prosodic prominence, both functional and emphasis, is substantial and foresees a semantic affiliation in the IU in which it appears. However, it does not specify the type of affiliation (locutive, pragmatic, or modal). In other words, we cannot foresee a gesture's semantic relation with linguistic information only by looking at gesture/prosody synchrony. Disclosing the image underlying the gesture in connection with the various semantic levels encoded in one IU is necessary.

Nonetheless, this research highlighted distributive restrictions, correlating with the L-AcT's principles, that constrain gesture affiliation. Gestures with a lexical affiliate and modal gestures can synchronize with both lexical emphasis and functional contours. Scanning units, which do not play any information function, can contain only prosodic emphasis and never attract a gesture with a functional value. Gestures with a functional affiliation tend to extend over the

prosodic unit, never synchronize to an emphasis, and necessarily synchronize with a prosodic contour bearing functional value. If one IU with functional value is synchronous to two gestures, only one of them can be a pragmatic gesture, while if one gesture extends over two PUs, they must share the functional value.

Gestures affiliated with the Topic function, the Parenthesis function, and the Illocutionary pattern structures convey the corresponding pragmatic value. If affiliated with the illocution, they are synchronized with Comment IUs, supplementing the illocutionary information. Modal gestures specify the speaker's stance about the content of the IU or just about one word within the IU, supplementing the information expressed linguistically. These results are interesting for Information Structure theory, as they constitute an independently motivated correlation to the L-AcT assumption, which predicts that one specific prosodic unit type will perform one specific information function type.

We did not find pragmatic gestures affiliated with IUs working as Discourse Connectors or Appendices, which, on the contrary, can be affiliated with a gesture in their locutive content. This lack of gesture/function correlation may be due to the reduced dimension of the dataset, but may also signal the low pragmatic relevance of these information units.

REFERENCES

1. Amir N, Silber-Varod V, Izre'el S. Characteristics of Intonation Unit Boundaries in Spontaneous Spoken Hebrew: Perception and Acoustic Correlates. In: Proceedings of Speech Prosody 2004. Nara: ISCA; 2004. p. 677-680. http://dx.doi.org/10.21437/SpeechProsody.2004-156

2. Austin J. L. How to Do Things with Words. Oxford: Oxford University Press; 1962.

3. Barbosa P, Raso T. Spontaneous Speech Segmentation: Functional and Prosodic Aspects with Applications for Automatic Segmentation / A segmentação da fala espontânea: aspectos prosódicos, funcionais e aplicações para a tecnologia. Rev. Estud. Ling. 2018; 26/4: 1361-1396. http://dx.doi.org/10.17851/2237-2083.26.4.1361-1396

4. Biber D, & Finegan E. Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect. Text. 1989; 9: 93-124. https://doi.org/10.1515/text.1.1989.9.1.93

5. Biber D, Johansson S, Leech G, Conrad S, Finegan E. The Longman Grammar of Spoken and Written English. London: Longman; 1999.

6. Bressem J. A linguistic perspective on the notation of form features in gestures. In: C. Müller C, Cienki A, Fricke E, Ladewig S.H, McNeill D, Teßendorf S. Body – Language – Communication: An International Handbook on Multimodality in Human Interaction, Vol. 1. Berlin: De Gruyter-Mouton; 2013. p. 1079–1098. http://dx.doi.org/10.1515/9783110261318.1079

7. Bressem J, Ladewig SH, Müller C. Linguistic Annotation System for Gestures (LASG). In: Müller C, Cienki A, Fricke E. Ladewig SH., McNeill D, Teßendorf S., editors. Body – Language – Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38) Vol. 1, Berlin: De Gruyter Mouton; 2013. p. 1098–1125.

8. Bressem J, Müller C. A repertoire of German recurrent gestures with pragmatic functions. In: Müller C, Cienki A, Fricke H, Ladewig S, McNeill D, Bressem J. Body – Language – Communication: An International Handbook on Multimodality in Human Interaction, Vol 2.Berlin: De Gruyter Mouton; 2014. p. 1575-1591. http://dx.doi.org/10.1515/9783110261318.1098

9. Buhmann J, Caspers J, van Heuven V, Hoekstra H, Martens J-P, Swerts M. Annotation of prominent words, prosodic boundaries, and segmental lengthening by non-expert transcribers in the spoken Dutch Corpus. In: Gonzales Rodriguez M, Suarez Araujo C., editors. Proceedings of the International Conference LREC2002. Paris: ELDA; 2002. p. 779–785.

10. Beckman M, Hirschberg J, Shattuck-Hufnagel S. The original ToBI system and the Evolution of the ToBI Framework. In: Sun-Ah J., editor. Prosodic Typology. The Phonology of Intonation and Phrasing. Oxford: Oxford University Press; 2005. p. 9–54. DOI: 10.1093/acprof:oso/9780199249633.003.0002

11. Cantalini G. Corpus multimodale annotato per lo studio della gestualità co-verbale nel parlato-parlato e nel parlato-recitato. In; Cresti E, Moneglia M., editors. Corpora e Studi Linguistici. Milano: Officinaventuno; 2022. p. 135-149. https://doi.org/10.48448/3k4n-hg49

12. Cantalini G, Moneglia M. The annotation of Gesture and Gesture / Prosody synchronization in Multimodal Speech Corpora. Journal of Speech Science, 9, 1-24, 2020. http://dx.doi.org/10.20396/joss.v9i00.14956

13. Cantalini G, Gagliardi G, Moneglia M, Proietti M. La correlazione gesto/prosodia e la sua variabilità: il parlato spontaneo di contro alla performance attorale. In: De Meo A, Dovetto F., editors. La Comunicazione Parlata, Atti del convegno: GSCP (Napoli, 12-14 Dicembre 2018), Roma: Aracne; 2020. p. 63-90. DOI:10.4399/97888255339034

14. Cavalcante F.A. The information unit of Topic: a crosslinguistic, statistical study based on spontaneous speech corpora. PhD diss., Belo Horizonte: UFMG; 2020. http://hdl.handle.net/1843/33673

15. Chafe W. Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing. Chicago: The University of Chicago Press; 1994.

16. Cienki A. The study of gesture in cognitive linguistics: How it could inform and inspire other research in cognitive science. WIREs Cognitive Science. 2022; 13(6): e1623. https://doi.org/10.1002/wcs.1623

17. Cienki A. Speakers' Gestures and Semantic Analysis. Cognitive Semantics. 2023; 9: 167–191. http://dx.doi.org/10.1163/23526416-bja10051

18. Cienki A, Müller C., editors. Metaphor and gesture. Amsterdam: Benjamins; 2008. https://doi.org/10.1075/gs.3?locatt=mode:legacy

19. Collier R, Rolelof D, de Pijper J, Sanderman A. Perceived prosodic boundaries and their phonetic correlate. In: Human Language Technology: Proceedings of a Workshop (Plainsboro, NJ, USA, March 21-24). Morgan Kaufmann Publishers, Inc; 1993. p. 341-345. https://aclanthology.org/H93-1068.pdf

20. Couper-Kuhlen E, Selting M. Interactional Linguistics. Studying Language in Social Interaction, Chapter F, Cambridge: Cambridge University Press; 2018.

21. Cresti E. Corpus di italiano parlato. Firenze: Accademia della Crusca; 2000.

22. Cresti E. La Stanza: un'unità di costruzione testuale del parlato. In: Ferrari A., editor. Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione e giustapposizione. Atti del X° Convegno SILFI. Firenze, Cesati; 2010. p. 713-732.

23. Cresti E. L'unité de suffixe: identification et interprétation des unités de la langue parlé. In: Caddéo S, Roubaud M-N, Rouquier M, Sabio F., editors. Penser les langues avec Claire Blanche-Benveniste. Aix-en-Provence: Presses Universitaires de Provence; 2012. p. 201–213.

24. Cresti E. The pragmatic analysis of speech and its illocutionary classification according to the Language into Act Theory. In: Izre'el S, Mello H, Panunzi A, Raso T., editors. In Search of Basic Units of Spoken Language: A Corpus-Driven Approach. Amsterdam: Benjamins; 2020. p. 181-220. http://dx.doi.org/10.1075/scl.94.06cre

25, Cresti E. The Appendix of Comment according to Language into Act Theory: A corpus-based research on Italian spoken data. CHIMERA: Romance Corpora and Linguistic Studies. 2021; 8:45–69. https://doi.org/10.15366/chimera2021.8.003

26. Cresti E, Moneglia M. The definition of the Topic within Language into Act Theory and its identification in spontaneous speech corpora, in Borreguero, M., Atayan, V. & Gross, S., editors. Special issue on Models of Discourse Units, Revue Romane. 2018; 53(1): 30-62. https://doi.org/10.1075/rro.00005.cre?locatt=mode:legacy

27. Cresti E, Moneglia M. The illocutionary basis of information structure. Language into Act Theory (L-AcT). In: Adamou E, Haude K, Vanhove M., editors. Information structure in lesser-described languages: Studies in prosody and syntax. Amsterdam: Benjamins; 2018. p. 359-401. https://doi.org/10.1075/slcs.199.13cre?locatt=mode:legacy

28. Cresti E, Moneglia M. The definition of Connectors, Connectives, and Conjunctions according to Language into Act Theory, in Ferrari A, Pecorari F., editors. Special Issue, Nuove prospettive di analisi dei connettivi, SILTA. 2021; 1: 31-51.

29. Cresti E, Moneglia M. The role of prosody for the expression of illocutionary types. The prosodic system of questions in spoken Italian and French according to Language into Act Theory. Front. Commun. 2023; 8:1124513. https://doi.org/10.3389/fcomm.2023.1124513

30. Dehé N. Parentheticals in Spoken English: The Syntax-Prosody Relation. Cambridge: Cambridge University Press; 2014. https://doi.org/10.1017/CBO9781139032391

31. Dankel P, Satti I. Multimodale Listen. Form und Funktion körperlicher Ressourcen in Aufzählungen in französischen, spanischen und italienischen Interaktionen. Romanistisches Jahrbuch. 2019; 70(1): 58–104. http://dx.doi.org/10.1515/roja-2019-0003

32. Dressel D, Dankel P, Kalkhoff A.M.T. What can collaboratively produced lists tell us about constructions? A multimodal analysis of co-constructed enumeration practices in spoken Spanish, in Constructions in Spanish; 2023. p. 340–374. Available at: 10.1075/cal.34.13dre.

33. Firenzuoli V. Le forme intonative di valore illocutivo dell'italiano parlato: analisi sperimentale di un corpus di parlato spontaneo (LABLITA) [PhD Thesis]. Firenze: Università di Firenze; 2003.

34. Firenzuoli V., Signorini S. L'unità informativa di Topic. Correlati intonativi. In: Marotta G, Nocchi N. editors. La Coarticolazione. Atti delle XIII Giornate del Gruppo di Fonetica Sperimentale, Pisa: ETS; 2003. p. 177-184.

35. Frege G. Über Sinn und Bedeutung. Zeitschrift für Philosophie Und Philosophische Kritik. 1892; 100/1: 25-50.

36. 't Hart J. Psychoacoustic backgrounds of pitch contour stylization. IPO Ann. Prog. Rep. 1976; 11: 11–19.

37. 't Hart J, Collier R, Cohen S. A Perceptual Study of Intonation. An Experimental-Phonetic Approach to Speech Melody. Cambridge: CUP; 1990. https://doi.org/10.1017/CBO9780511627743

38. Izre'el S, Mettouchi A. Representation of Speech in CorpAfroAs. Transcriptional Strategies and Prosodic Units. In: Mettouchi A, Vanhove M, Caubet D., editors. Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages. Amsterdam: Benjamins; 2015. p. 13–41. https://doi.org/10.1075/scl.68.01izr?locatt=mode:legacy

39. Izre'el S, Mello H, Panunzi A, Raso T. (2020). In Search of Basic Units of Spoken Language: A Corpus-Driven Approach. Amsterdam: Benjamins; 2020. https://doi.org/10.1075/scl.94?locatt=mode:legacy

40. Kanane S, Pietrandrea P, Gerdes K. The annotation of list structures, In Lacheret A, Kahane S, Pietrandrea P., editors. Rhapsodie: a Prosodic and Syntactic Treebank for Spoken French. Amsterdam: Benjamins; 2019. p. 69-95. http://dx.doi.org/10.1075/scl.89.06kah

41. Kendon A. Gesture: Visible Action as Utterance. Cambridge: CUP; 2004. http://dx.doi.org/10.1017/CBO9780511807572

42. Kendon A. Pragmatic functions of gestures: some observations on the history of their study and their nature. Gesture. 2017; 16: 157-175. https://doi.org/10.1075/gest.16.2.01ken?locatt=mode:legacy

43. Kirchhof C. So What's Your Affiliation With Gesture? In Kirchhof CZ. Malisz Z, Wagner P. editors. GeSpIn 2011, vol 2. Bielefeld. 2011. p. 1-7.

44. Kita S, van Gijn I, van der Hulst H. Movement phases in signs and co-speech gestures, and their transcription by human coders. In: Wachsmuth I, Fröhlich M., editors. Gesture and Sign Language in Human-Computer Interaction Berlin: Springer; 1998. p. 23–35. https://doi.org/10.1017/CBO9780511620850.011

45. Krifka M. Basic notions of information structure. Acta Linguistica Hungarica. 2008; 55(3–4): 243-276. https://www.jstor.org/stable/26190905

46. Krifka M, Musan R., editors. The Expression of Information Structure. Berlin/Boston: De Gruyter Mouton; 2012. https://doi.org/10.1515/9783110261608

47. Ladd DR. Intonational Phonology (2nd ed.). Cambridge: Cambridge University Press; 2008. http://dx.doi.org/10.1075/fol.17.2.08ogr

48. Ladewig SH, Bressem J. A linguistic perspective on the notation of gesture phases. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Teßendorf S., editors. Body – Language – Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and

Communication Science 38) Vol. 1, Berlin: De Gruyter Mouton; 2013. p. 1060–1079. http://dx.doi.org/10.1515/9783110261318.1060

49. Loehr D. Gesture and Intonation [Ph.D Thesis]. Washington, DC: Georgetown University; 2004.

50. Loehr D. Temporal, structural, and pragmatic synchrony between intonation and gesture. Laboratory Phonology. 2012; 3(1): 71-89. https://doi.org/10.1515/lp-2012-0006

51. Loehr D. Gesture and prosody. In: Müller C, Cienki A, Fricke E, Ladewig SH, McNeill D, Teßendorf S., editors. Body – Language – Communication: An International Handbook on Multimodality in Human Interaction (38.2). Berlin, De Gruyter Mouton; 2014. p. 1381-1391. https://doi.org/10.1515/9783110302028.1381

52. Lopez-Ozieblo R. Proposing a revised functional classification of pragmatic gestures. Lingua. 2020; 247, 102870. http://dx.doi.org/10.1016/j.lingua.2020.102870

53. Martin PH. The Structure of Spoken Language Intonation in Romance, Cap. 11 Winpitch. Cambridge: Cambridge University Press; 2015. p. 259–271. https://doi.org/10.1017/CBO9781139566391.013

54. McNeill D. Hand and Mind: What Gestures Reveal about Thought. Chicago: University of Chicago Press; 1992. http://dx.doi.org/10.2307/1576015

55. McNeill D. Gesture and thought. University of Chicago Press; 2005. http://dx.doi.org/10.7208/chicago/9780226514642.001.0001

56. McNeill D. Why We Gesture. Cambridge: Cambridge University Press; 2015. https://doi.org/10.1017/CBO9781316480526

57. McNeill D, Duncan SD. Growth point in thinking-for-speaking. In: McNeill D., editor. Language and Gesture. Cambridge: Cambridge University Press; 2000. p. 141-161.

58. McNeill D, Quek F, McCullough KE, Duncan SD, Furuyama N, Bryll R, Furuyama N, Ansari R. Catchments, prosody and discourse. Gesture. 2001; 1(1): 9–33. http://dx.doi.org/10.1075/gest.1.1.03mcn

59. Mittmann-Malvessi M. O C-ORAL-BRASIL e o estudo da fala informal: um novo olhar sobre o Tópico no Português Brasileiro [PhD Thesis] - Universidade Federal de Minas Gerais, Belo Horizonte.

60. Moneglia M. Spoken corpora and pragmatics. Rev. Brasileira Ling. Apl. 2011; 11: 479–519. http://dx.doi.org/10.1590/S1984-63982011000200009

61. Moneglia M. Gesture/Speech synchronization and gesture's scope across Information Unit types, in Korotaev N, Sumbatova N., editors. Corpus scientiae: Papers in honor of Vera I. Podlesskaya. Moscow: Buki Vedi; 2024. p. 339-365.

62. Moneglia M, Raso T, Malvessi-Mittmann M, Mello H. Challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: C-ORAL-BRASIL / C-ORAL-ROM. In: Speech Prosody 2010, W1.09, Satellite workshop on Prosodic Prominence: Perceptual, Automatic Identification. Chicago; 2010. Retrieved from: https://www.isca-speech.org/archive/sp2010/sp10_2010.html, .

63. Moneglia M, Raso T. Notes on Language into Act Theory (L-AcT). In: Raso T, Mello H, editors. Spoken Corpora and Linguistic Studies, Amsterdam: Benjamins; 2014. p. 468–495. https://doi.org/10.1075/scl.61.15mon?locatt=mode:legacy

64. Moneglia M, Fabbri M, Quazza S, Panizza A, Danieli M, Garrido JM, Swert M. Evaluation of consensus on the annotation of terminal and non-terminal prosodic breaks in the C-ORAL-ROM corpus. In: Cresti E., Moneglia M., editors. C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages. Amsterdam: Benjamins; 2005. p. 257-276. https://doi.org/10.1075/scl.15.09mon?locatt=mode:legacy

65. Müller C. Gestural Modes of Representation as Techniques of Depiction. In: Müller C. Cienki A, Fricke E, Ladewig SH, McNeill D, Bressem J., editors. Body – Language – Communication: An International Handbook on Multimodality in Human Interaction. Berlin: De Gruyter Mouton; 2014. p. 1687–1702. http://dx.doi.org/10.1515/9783110302028.1687

66. Müller C. Gesture and Sign: Cataclysmic Break or Dynamic Relations? Front. Psychol. 2018; 9: 1-20. https://doi.org/10.3389/fpsyg.2018.01651

67. Müller C, Bressem J, Ladewig SH. Towards a grammar of gestures: A form-based view in Body Language Communication. In: Müller C, Cienki A, Fricke E, Ladewig SH., McNeill D, Teßendorf S., editors. Body – Language – Communication: An International Handbook on Multimodality in Human

Interaction, Vol. 1. Berlin: De Gruyter-Mouton; 2013. p. 707-733. http://dx.doi.org/10.1515/9783110261318.707

68. Pierrehumbert J, Hirshberg J. The meaning of Intonational Contour in the Interpretation of Discourse. In: Cohen PH., Morgan JL, Pollack ME., editors. Intensions in Communication. Cambridge MA: MIT Press; 1990. p. 271-311. http://dx.doi.org/10.7551/mitpress/3839.003.0016

69. Raso T. Prosodic Constraints for Discourse Markers. In: Raso T, Mello H., editors. Spoken Corpora and Linguistics Studies. Amsterdam: Benjamins; 2014. p. 411-467. https://doi.org/10.1075/scl.61.14ras?locatt=mode:legacy

70. Raso T, Vieira MA. description of Dialogic Units/Discourse Markers in spontaneous speech corpora based on phonetic parameters. Chimera: Romance Corpora and Linguistic Studies. 2016; 3: 221-249. http://dx.doi.org/10.15366/chimera2016.3.2.008

71. Raso T, Cavalcante F, Mittmann M. Prosodic forms of the Topic information unit in a cross-linguistic perspective: a first survey. In: de Meo A, Dovetto F., editors. La comunicazione parlata. Proceedings of the GSCP International Conference, 13-15 June 2016. Roma: Aracne; 2017. p. 473-498. DOI: 10.4399/978882552064427

72. Rocha B. Uma Metodologia Empírica Para a Identificação e Descrição de Ilocuções e a Sua Aplicação Para o Estudo da Ordem em, PB, e Italiano [Ph.D. thesis]. Belo Horizonte: UFMG; 2016.

73. Rohrer PA. temporal and pragmatic analysis of gesture-speech association: A corpus-based approach using the novel MultiModal MultiDimensional (M3D) labeling system [Ph.D. thesis]. Barcelona: University Pompeu Fabra; 2022. https://theses.hal.science/tel-03994053v1

74. Rohrer PA, Delais-Roussarie E, Prieto P. Visualizing prosodic structure: Manual gestures as highlighters of prosodic heads and edges in English academic discourses, Lingua. 2023. http://dx.doi.org/10.1016/j.lingua.2023.103583

75. Rossi M. Le seuil de glissando ou seuil de perception des variations tonales pour la parole. Phonetica. 1971; 23: 1–33. https://doi.org/10.1159/000259328

76. Rossi M. La perception des glissando descendants dans les con-tours prosodiques. Phonetica 35, 11–40. https://doi.org/10.1159/000259920

77. Panunzi A, Gregori L, Rocha B. Comparing annotations for the prosodic segmentation of spontaneous speech. In: S. Izre'el S., Mello H., Panunzi A., Raso T., editors. In Search of Basic Units of Spoken Language: A Corpus-Driven Approach. Amsterdam: Benjamins; 2020. p. 403–432. https://doi.org/10.1075/scl.94.17pan?locatt=mode:legacy

78. Panunzi A., Saccone V. Complex illocutive units in L-AcT: an analysis of non-terminal prosodic breaks of bound and multiple Comments. Rev. Estud. Ling. 2018; 26: 1647–1674. https://doi.org/10.17851/2237-2083.26.4.1647-1674

79. Saccone V. La stanza nella Teoria della Lingua in Atto. CHIMERA. Romance Corpora and Linguistic Studies. 2020; 7: 55-68. https://doi.org/10.15366/chimera2020.7.003

80. Saccone V, Panunzi A. Le unità di Comment multiplo. Analisi secondo la Teoria della Lingua in Atto. In: De Meo A., Dovetto F., editors. La comunicazione parlata. Atti del Congresso GSCP – Università degli Studi di Napoli L'Orientale (Napoli, 12-14 dicembre 2018). Roma: Aracne; 2020. p. 263-286.

81. Saccone V., Trombetta C., Parenthetical Units and Structures in Italian and German spoken language: prosodic and textual analysis. CHIMERA. Romance Corpora and Linguistic Studies. 2021; 8: 1-23. http://dx.doi.org/10.15366/chimera2021.8.001

82. Saccone V, Vieira M, Panunzi A. Illocutive Unit in Language into Act Theory: an analysis of non-terminal prosodic breaks of Bound Comments and Lists. Journal of Speech Sciences. 2018; 7(2): 51–64. http://dx.doi.org/10.20396/joss.v7i2.14998

83. Saccone V, Panunzi A. Le strutture parentetiche nel parlato italiano: classificazione funzionale e identificazione prosodica. In: Castagneto M, Ravetto M., editors. La comunicazione parlata, Atti del Congresso GSCP – 2021 Vercelli, Roma: Aracne; 2023. p. 461-486. DOI: 10.53136/979122180945917

84. Schriffin D. Discourse markers. Cambridge: Cambridge University Press; 1987. https://doi.org/10.1017/CBO9780511611841

85. Signorini S. Topic e soggetto in corpora di italiano Parlato [PhD. Thesis]. University of Florence; 2005.

86. Swerts M. Prosodic features at discourse boundaries of different strengths. J. Acoustic. Soc. Am. 1997; 101: 514–521. http://dx.doi.org/10.1121/1.418114

87. Tucci I. Obiter dictum: la funzione informativa delle unità parentetiche. In: Pettorino M, Giannini A, Dovetto F., editors. La Comunicazione parlata 3. Atti del Congresso GSCP (Napoli, 23-25 febbraio 2009). Napoli: Università l'Orientale Press; 2010. p. 635-654.

88. Tucci I. Illocution and Modality in Spoken Italian: performing a Speech Act through words and Judging their Semantic Content. A Corpus-Based Analysis. In: Mello H, Panunzi A, Raso T., editors. Pragmatics and Prosody Illocution, Modality, Attitude, Information Patterning, and Speech Annotation. Firenze: FUP; 2012. p. 83-100. 10.36253/978-88-6655-084-6

Online resources

89. WINPITCH https://www.winpitch.com/

90. PRAAT https://www.fon.hum.uva.nl/praat/

91. ELAN https://archive.mpi.nl/tla/elan