**«Illuminazioni» (ISSN: 2037-609X), n. 41, luglio-settembre 2017**

**Raffaele Zago**

# CROSS-LINGUISTIC DIMENSIONS OF COMPARABILITY IN FILM DIALOGUE

ABSTRACT. This paper examines original film dialogue from a cross-linguistic perspective. More specifically, the paper will identify and compare the most frequent 3-grams – i.e. 3-word clusters – in a corpus of original English and original Italian films. This will be done with the specific aim of exploring the dimensions of comparability between the language of English films and the language of Italian films. It will be shown that the dialogues of English and Italian films exhibit a pronounced degree of similarity not only in terms of their decidedly clausal 'texture' and markedly interactional focus but also at the level of individual 3-grams – namely the English *I don't know* and the Italian *non lo so*, whose various functions will be described.

## 1. Introduction

A vast number of studies have focused on the translation of film dialogue from the points of view of dubbing and subtitling. Also, several descriptions have been provided of original, non-translated film dialogue in specific languages[1]. Instead, the cross-linguistic analysis of original films, i.e. the comparative study of original films in two or more languages, represents a somewhat neglected area in the literature on film dialogue.

The cross-linguistic analysis of original films constitutes an effective method capable of enhancing the description of film dialogue in two respects (cf. Biber 1995). First, it immediately highlights the linguistic similarities shared by films in different languages, and given that, if identified, the said similarities have emerged in spite of the typological diversity between the languages under consideration, they can confidently be viewed as important descriptors of the core linguistic features and communicative functions of film dialogue as a register. Second, the cross-linguistic approach might 'flag' possible idiosyncratic differences existing among films produced in two or more languages. In short, such approach makes it possible to throw light on the dimensions of comparability among films produced in different languages, a topic on which not much empirical work has been carried out so far.

The objective of this paper is precisely to study original film dialogue from a cross-linguistic perspective. The paper will identify and compare the most frequent phraseological clusters in a corpus of English and Italian films. This will be done in the interest of exploring the level of comparability between the language of English

[1] For a bibliography of linguistic research on films and TV series, see Bednarek and Zago (2017).

films and the language of Italian films. Section 2 will illustrate the corpus used and the methodology adopted in the present study.

## 2. Data and methodology

As stated in the previous section, this paper examines the language of original English films and the language of original Italian films with the aim of throwing light on the dimensions of comparability between them.

The corpus used in the present analysis is the *Pavia Corpus of Film Dialogue* (Freddi and Pavesi 2009; Pavesi 2014), henceforth PCFD, which comprises the complete dialogues of 24 British and American films, their dubbed Italian translations[2], as well as the complete dialogues of 24 original Italian films, for a total of approximately 695,000 words[3]. The specificity of the films included in the PCFD is that they are "conversational" (Freddi and Pavesi 2009: 98), i.e. they portray interactions taking place in contemporary, naturalistic settings and present dialogues which have been designed so as to reproduce everyday spontaneous conversation. This implies that the results discussed in section 3 should be considered as readily

[2] The dubbed films of the PCFD will not be considered in this study.

[3] In addition to Freddi and Pavesi (2009) and Pavesi (2014), an updated description of the rationale and design of the PCFD, together with a list of publications based on the corpus, can be found at the following link: http://studiumanistici.unipv.it/?pagina=p&titolo=pcfd.

generalisable to the population of conversational films set in realistic, contemporary settings, while they are not necessarily generalisable to other types of films such as westerns or costume films.

From a methodological point of view, the present analysis focuses on n-grams, i.e. “multi-word strings of two or more uninterrupted word forms” (Stubbs and Barth 2003: 62), and in particular on 3-grams, that is, 3-word clusters such as *I don’t know*, *what do you*, *you want to*, etc., in the English films of the PCFD, and *non lo so*, *non è che*, *ha detto che*, etc., in the Italian films of the PCFD[4]. N-grams – also known as “lexical bundles”, among several other labels – are characterised by the fact that they recur frequently in a given corpus. To put it in Biber *et al.*’s (1999: 989) words, they are “the sequences of words that most commonly co-occur in a register”. As such, they represent “important textual building blocks used in spoken and written discourse” (Tracy-Ventura *et al.* 2007: 217). Another characteristic of n-grams is that, while they may be complete constructions, as for example in the case of the interjection *oh my god* or the reassurance *non ti preoccupare* in the PCFD, they tend to be fragmentary, i.e. to “extend across structural units” (Biber *et al.* 1999: 991), as exemplified, among others, by *don’t know what* and *lo sai che* in the PCFD.

By computing and analysing n-grams, the researcher is in the position to capture a rich set of fundamental phraseological units which, due to their being recurrent, can reasonably be considered as marking salient communicative functions

[4] See Culpeper and Kytö (2010: 106), who offer methodological reasons justifying the choice of 3-grams over other types of n-grams.

associated with the register under study, not to mention that the patterns of register variation highlighted by n-gram frequency lists would probably go unnoticed, or would at least be less immediately noticeable, in investigations based on single-word frequency lists. In short, n-grams offer "a powerful window" (Granger 2014: 69) onto registers, and this is what has motivated their adoption in the present study.

The extraction of the n-grams from the PCFD has been carried out via the linguistic software package *WordSmith 6.0* (Scott 2011). More specifically, *WordSmith 6.0* has been used to generate two lists, one comprising the 100 most frequent 3-grams in the original English films of the PCFD, the other comprising the 100 most frequent 3-grams in the original Italian films of the corpus. Once extracted, the two lists have been compared with one another with the specific aim of identifying cross-linguistic similarities[5].

## 3. Cross-linguistic similarities in film dialogue

The following subsections will illustrate the similarities which have emerged in the comparison between the 3-grams extracted from the original English films and those extracted from the original Italian films of the PCFD. In particular, subsection

[5] For a discussion of the methodological challenges posed by the cross-linguistic analysis of n-grams, see Granger (2014). Useful observations on such issue can also be found in Cortes (2008) and Tracy-Ventura *et al*. (2007).

3.1 will deal with the decidedly clausal ‘texture’ characterising both groups of 3-grams, while subsection 3.2 will highlight their markedly interpersonal focus. Finally, subsection 3.3 will examine the degree of comparability between the 3-grams *I don’t know* and *non lo so* from both a quantitative and a qualitative point of view.

### *3.1 The clausal ‘texture’ of film dialogue*

A first cross-linguistic similarity which has emerged in this analysis is that both the 3-grams extracted from the original English films and those extracted from the original Italian films of the PCFD have a clausal ‘texture’. To put this in figures, 82 of the 100 most frequent English 3-grams (e.g*. I don’t know*, *are you doing*, *to talk to*, etc.) and 58 of the 100 most frequent Italian 3-grams (e.g. *non lo so*, *ha detto che*, *che è successo*, etc.) include verbs. It should be added that many of the Italian 3-grams which do not include a verb are interpersonal, negative-polarity fragments which are ‘completed’ by verbs. For example, a look at the concordances shows that *non ce la* has the verbs *faccio* and *facevo* as its most frequent right collocates, *non me ne* has the verbs *frega* and *importa* as its most frequent right collocates, *che non mi* has the verbs *piace* and *vuoi* as its most frequent right collocates, etc.

The clausal ‘texture’ exhibited by the English and the Italian films of the PCFD, and presumably by films produced in other languages, is clearly a consequence of the fact that film dialogue often simulates spontaneous conversation, a register which is well known for its heavy reliance on verbs, as opposed to the decidedly nominal 'texture' of written expository registers such as academic prose (Biber *et al*. 1999: 65-66; Biber *et al*. 2004; Tracy-Ventura *et al*. 2007; Cortes 2008). The strong presence of verbs in film dialogue, however, is not only the reflection of a conversational feature, but is also a diegetic necessity, verbs being indispensable for carrying out a number of important narrative actions, such as the expression of the characters' thoughts and intentions (examples 1, 2), the presentation of communication activities (examples 3, 4), the presentation of spatial movements (examples 5, 6), etc.:

1. CYNTHIA: **I don't want** to upset my daughter, do I? [*Secrets and Lies*]
2. CAGNETTI: **Io non voglio** che in futuro penseremo che quello che stiamo facendo oggi è una cazzata. [*Come te nessuno mai*]
3. JULIANNE: Michael, I have **to talk to** you. [*My Best Friend's Wedding*]
4. DOCTOR: Mi **ha detto che** si è trasferito, che ora è in pensione e vive a Roma. [*L'aria salata*]
5. HONEY: **Get out of** here! You'll get arrested! [*Saving Grace*]
6. LUIGI: ((angrily)) Senti, non cominciare con questi discorsi, sennò prendo e **me ne vado**! [*La terra*]

The abundance of verbal material is a characteristic which films have in common with plays. In particular, the reference is here to Culpeper and Kytö (2010), who analysed 3-grams in a corpus of present-day and Early Modern English plays. Among many other things, the two scholars found that the 3-grams were “dominated by full or parts of verb phrases” (Culpeper and Kytö 2010: 119). The explanation they offered to account for such result, namely that verbs are frequently used in plays to construct “a dynamic interaction for public entertainment” (Culpeper and Kytö 2010: 119), is equally applicable to film dialogue and is in keeping with what has been argued in this subsection.

### *3.2 The phraseology of interpersonality in film dialogue*

Another similarity which has emerged in this study is the centrality of the ‘you and I’ dimension in the phraseology of both the original English films and the original Italian films of the PCFD. For example, many 3-grams include first person pronouns (*I have to*, *I was just*, *look at me*, etc. in the English films of the PCFD; *io non sono*, *non me ne*, *perché non mi*, etc. in the Italian films of the PCFD), second person pronouns (e.g. *what do you*, *you know what*, *you all right*, etc. in the English films of the PCFD; *sei tu che*, *non ti preoccupare*, *te l’ho detto*, etc. in the Italian films of the PCFD), or a combination of first and second person pronouns, as in the following examples:

7. JULIANNE: Michael, **I love you**. I've loved you for nine years. I've just been too arrogant and scared to realize it. Well, now I'm just scared so I, I, I realise this comes at a very inopportune time but I really have this gigantic favour to ask of you. Choose me. Marry me. Let me make you happy. Oh, that sounds like three favours, doesn't it? But... ((kisses him)) [*My Best Friend's Wedding*][6]
8. CURCI: ((on the phone)) **Io e te**, lontani, all'estero, dove vuoi tu, amore mio. A Batticaloa, Sri Lanka, c'è quell'amico, Fabio, te ne ho parlato, no? [*La cena*]
9. WILLIAM: No! No, no, no wait. **I thought you** were, um, someone else. **I thought you** were Spike. I'm thrilled that you're not. ((they kiss)) Wow! [*Notting Hill*]
10. ANTONIA: ((angrily, in a high voice)) Mamma, sei tu che devi vergognarti! **Io non ti** sopporto più, veramente! Ma perché non te ne torni a casa tua? Io voglio stare da sola, la capisci questa parola? ((shouting)) Sola! [*Le fate ignoranti*]

Interpersonality is also marked by possessives, as in the case of the 3-gram *this is <u>my</u>*, which has a practically equivalent counterpart in the Italian 3-grams *è il <u>mio</u>*

[6] Notice also that this example includes a 3-gram which is entirely made up of personal pronouns, namely the repeat *I I I*. In films, dysfluencies such as *I I I* convey an impression of realism by evoking the unplanned, online nature of spontaneous conversations. Also, they are used to mark emotional involvement, as is especially evident in example 7 (cf. Zago 2016: 118).

and *è la mia*. Moreover, in the Italian list of 3-grams, the ‘you and I’ dimension is often evident in, and recoverable from, the very form of the verb, even in the absence of personal pronouns and possessives (e.g. *lo sai che*, *ho bisogno di*, *che hai fatto*, *non riesco a*, *non ho capito*, *ma lo sai*, etc.), Italian being a pro-drop language.

These results were predictable in that a film is ultimately an extended conversation and, as such, it reproduces the most fundamental communicative dimension of conversational registers, namely what O’Keeffe *et al*. (2007: 68) have referred to as “the speaker-listener world of *I* and *you*”. A film is an extended conversation also in the sense that its language is systematically dialogical and co-constructed even in those scenes or parts of scenes whose aim is more evidently and openly informational, that is, those scenes or parts of scenes which are conceived by screenwriters as especially important for the advancement of the narrative chain (Veirano Pinto 2014: 117; Zago 2016: 112). For instance, in films, phone calls are used to disclose information rather than as phatic activities; nonetheless, the phone call is one of those filmic situations where the disclosure of information takes place between an *I* – the caller – and a *you* – the receiver of the call – i.e. in a dialogical fashion, with informational language, hence, being ‘concealed’ and packaged conversationally.

The phraseology of interpersonality is another characteristic which films share with plays, where, as again illustrated by Culpeper and Kytö (2010: 132), a wide range of interpersonal 3-grams are instrumental in “the articulation of personal desires and negotiation of social relationships", not to mention that a pronounced interpersonal focus is also typical of the language of TV series, as pointed out, for example, by Bednarek (2011: 71).

### *3.3 'I don't know' and 'non lo so' in close*-up

A further cross-linguistic similarity between the original English component and the original Italian component of the PCFD is that the most frequent 3-gram in English films – i.e. *I don't know*, occurring 294 times in all the 24 English films – is practically the same as the most frequent 3-gram in Italian films – i.e. *non lo so*, occurring 191 times in 23 Italian films. The correspondence between these 3-grams is not only at the level of the lexical verbs involved – i.e. the English mental verb *to know* and its Italian synonym *sapere* – but also at the level of polarity, which is negative in both cases.

The high frequency of *I don't know* in the PCFD aligns with what happens in other English corpora of spontaneous conversation and fictional dialogue, according to a trend that the previous subsections have already highlighted. For example, in the conversation section of the *Longman Spoken and Written English Corpus* (LSWE Corpus), *I don't know* was found to be the most frequent 3-gram by Biber *et al*. (1999: 994). More in general, Biber *et al*. (1999) point out that many occurrences of the most frequent type of n-gram in conversation – i.e. personal pronoun + lexical verb phrase + complement-clause fragment – "report negative personal states in the first person", as in the case of "*I don't know*, *I don't think*, *I don't want*, *I don't like*" (Biber *et al*. 1999: 1004). *I don't know* was found to be the most frequent 3-gram also in the *Cambridge and Nottingham Corpus of Discourse in English* (CANCODE) and in the North American spoken segment of the *Cambridge International Corpus* (CIC) analysed by O'Keeffe *et al*. (2007: 66-67), as well as in the corpus of TV series investigated by Bednarek (2011: 65) and, finally, in the corpus of present-day plays investigated by Culpeper and Kytö (2010: 116-117).

Similarly, the high frequency of *non lo so* in the PCFD is in line with the results offered by the *Perugia Corpus* (PEC), a reference corpus of spoken and written Italian comprising more than 26 million words (Spina 2014). A search in the PEC shows that *non lo so* is almost twice as frequent in Italian film dialogue as in general spoken Italian (see table 1). When the distribution across the subsections of the PEC is considered, one finds that while *non lo so* is most frequent in spontaneous conversation (on the telephone and face-to-face, respectively), it is both appreciably frequent and more widely dispersed in fictional TV programmes and film dialogue (see table 2). Such correlation between high frequency and wide dispersion for *non lo so* in the PEC, which closely mirrors what happens in the PCFD, qualifies this 3-gram as a salient marker of fictional, telecinematic discourse.

**Table 1 – The four sections of the PEC in which *non lo so* is most frequent**

| **Sections** | **Size** | **Hits** | **Dispersion** | **Frequency per million words** |
|---|---|---|---|---|
| Film dialogue | 626,289 | 260 | 59 out of 66 | 415.14 |
| General spoken Italian[7] | 2,158,555 | 457 | 201 out of 1041 | 211.72 |
| Italian spoken on TV[8] | 1,147,255 | 226 | 71 out of 127 | 196.99 |
| Novels | 3,545,430 | 379 | 55 out of 60 | 106.90 |

**Table 2 – The four subsections of the PEC in which *non lo so* is most frequent**

| **Subsections** | **Size** | **Hits** | **Dispersion** | **Frequency per million words** |
|---|---|---|---|---|
| Telephone conversations between peers | 283,652 | 210 | 105 out of 440 | 740.34 |
| Face-to-face conversations between peers | 187,454 | 127 | 49 out of 97 | 677.50 |
| Fictional TV | 127,026 | 58 | 16 out of 17 | 456.60 |

[7] 'General spoken Italian' is a label used here to refer to a varied spoken section of the PEC including: face-to-face and telephone conversations between peers; dialogic language spoken in institutional settings; various types of monologic spoken language, namely conferences, lessons, trials, political discourse, religious discourse, monologic language spoken in other institutional settings, and songs (Spina 2014).

[8] As explained in Spina (2014), this section of the PEC comprises news programmes (also including programmes such as *Report* and *Ballarò*) and various types of entertainment programmes (namely talk shows, fictional programmes, sports programmes, and other TV shows).

| | | | | |
|---|---|---|---|---|
| programmes | | | | |
| Film dialogue | 626,289 | 260 | 59 out of 66 | 415.14 |

To recapitulate, *I don't know* and *non lo so* are the most frequent 3-grams in their respective components of the PCFD. This finding partly derives from the fact that these 3-grams are already prominent in unscripted conversation both in English and in Italian, as documented above by reference to corpora including spontaneous spoken language (i.e. the LSWE Corpus, the CANCODE, the CIC and the PEC). Further, it has been pointed out that *I don't know* and *non lo so* are also very frequent in other corpora of fictional dialogue (i.e. the corpus of TV series analysed by Bednarek 2011; the corpus of present-day plays analysed by Culpeper and Kytö 2010; the telecinematic sections of the PEC).

The issue which has now to be addressed is that of the reasons for the prominence of *I don't know* and *non lo so* in the PCFD and, more in general, in film dialogue. A first likely reason is that *I don't know* and *non lo so* are useful in film dialogue in that they conceal scriptedness by simulating unplannedness. This can be seen, for instance, in example 11, where *I don't know* gives viewers the impression that Gerry does not have a clear idea as to how to describe his interlocutor's behaviour, an impression which is also conveyed by the hesitator *ah* and by the repeat *a bit ... a bit*.

11. GERRY: You’re sure? You’ve been a bit, ah, **I don’t know**, a bit distant since I’ve got back. [*Sliding Doors*]

Similar cases can be found in Italian films as well, as exemplified in 12, where the repetition of *non lo so* provides viewers with the impression that Piero is thinking of examples (e.g. *la sicurezza dolce*; *invece che i muri, gli incontri*; etc.) as he speaks. Notice also that, in example 12, the two occurrences of *non lo so* cooperate with other markers of unplannedness, namely the repeats *che-che* and *la- ... la*.

12. PIERO: Sì, ma che almeno siano balle di sinistra! **Non lo so**, la sicurezza dolce. Invece che i muri, gli incontri. Invece che-che i divieti, la- **non lo so**, la gente per strada! [*Diverso da chi?*]

In short, inserting *I don’t know* and *non lo so* in a turn seems a strategy whereby English and Italian screenwriters, or the very actors, can give an air of unplannedness to lines which, in reality, have been carefully pre-planned, thus making film dialogue sound less ‘artificial’ and more realistically conversational.

Another plausible reason explaining the high frequency of *I don’t know* and *non lo so* in film dialogue is that these 3-grams turn out as instrumental for screenwriters in managing the amount of information presented to viewers as well as the timing with which the information is presented. In particular, *I don’t know* and *non lo so* seem to be used by screenwriters to delay the presentation of key pieces of information, which are effectively disclosed at some later point in the film. An example is offered in 13, where Sister Helen is pressing the death row prisoner Matthew Poncelet with her questions in an attempt to make him confess to the murder of a young couple, a crime which, according to Poncelet, has been committed by Carl Vitello. Sister Helen's questions are met with evasiveness by Poncelet, who says that he does not know exactly how he got involved in the murder perpetrated by Vitello. In actual fact, as is clear from the rest of the film, in this scene *I don't know* is a narrative tactic allowing the screenwriter to delay Poncelet's admission of guilt until the very end of the film, when Poncelet is about to be executed by lethal injection.

13. SISTER HELEN: […] What possessed you to be in the woods that night?

MATTHEW PONCELET: ((shouting)) I told you, I was stoned out of my head!

SISTER HELEN: Now don't blame the drugs. You were harassing couples for weeks before this happened. Months! What was it?

MATTHEW PONCELET: What do you mean?

SISTER HELEN: Did you look up to Vitello? Did you think he was cool? Did you wanna impress him?

MATTHEW PONCELET: **I don't know**.

SISTER HELEN: ((shouting)) You could've just walked away.

MATTHEW PONCELET: Hey, he went psycho on me.

SISTER HELEN: Don't blame him. You blame him, you blame the government, you blame drugs, you blame blacks. ((shouting)) You blame the Percys, you blame the kids for being there. What about Matthew Poncelet? Where's he in this story? What, is he just an innocent? A victim? [*Dead Man Walking*]

An analogous function is performed by *non lo so* in example 14, where Elsa asks her husband Michele why Roberto did not attend her birthday party. Michele replies by saying that he does not know the reason why Roberto missed the party. Again, as was the case for *I don't know* in the previous example, *non lo so* is used here to withhold two central pieces of information which are introduced in the following scene, namely that Michele has lost his job and that Roberto played an important role in Michele's dismissal.

14. ELSA: […] Grazie, amore mio! È stata una festa bellissima! C'erano tutti i miei amici, tutti i nostri amici! Mi sono divertita tantissimo! C'erano proprio tutti tutti, eh! Roberto non c'era!

MICHELE: Non ce l'ha fatta.

ELSA: Come mai?

MICHELE: **Non lo so**, Elsa. Non ho capito.

ELSA: Eh, ti sei offeso! Ma dai, avrà avuto da fare, su! Lo perdoniamo, eh? ((singing)) Ti stringerò. Giuro che ti farò male. [*Giorni e nuvole*]

*I don't know* and *non lo so*, hence, seem to be deployed by screenwriters to 'ration' narrative information. More specifically, they are used to temporarily withhold vital pieces of information which are subsequently disclosed at some crucial point in the film. Such withholding has the effect of insinuating doubts and arousing questions among viewers concerning the missing pieces of information which the characters have presented as unknown; this, in turn, keeps viewers engaged, thus preparing the ground for, and enhancing the impact of, the subsequent disclosure (cf. Kozloff 2000: 37-43).

One further reason which might be suggested to account for the frequent occurrence of *I don't know* and *non lo so* in English and Italian films is that these 3-grams seem to lend themselves to the expression of emotional content, both positive and negative. For instance, in example 15 *I don't know* launches a confrontational utterance with which Nola expresses frustration over her clandestine relationship with Chris (i.e. *I don't know what I'm doing with you*)[9]. Another emotionally-loaded occurrence of *I don't know*, this time as part of a closeness-marking utterance, can be found in Chris's second turn (i.e. *I don't know what I'd do if I couldn't see you*).

[9] On n-grams as utterance launchers see Biber *et al*. (1999: 1003), Biber *et al*. (2004: 399), and Culpeper and Kytö (2010: 140).

15. NOLA: ((in a loud voice)) **I don't know** what I'm doing with you. You're never gonna leave Chloe.

    CHRIS: Maybe I will.

    NOLA: Don't say that unless you mean it.

    CHRIS: ((sighs)) Chloe is just so desperate to get pregnant. I mean, it's mechanical. ((sighs)) **I don't know** what I'd do if I couldn't see you. Really. I mean it. [*Match Point*]

The role of *non lo so* as a marker of emotional content in Italian films is documented in example 16, extracted from a scene featuring Paolo who bursts into Arianna's house. When Arianna, Paolo's ex-girlfriend, asks him the reason for his unexpected visit, he first says *non lo so* twice, but then confesses the real reason: he wants to tell her that he will always wait for them to get back together. *Non lo so*, here, clearly marks Paolo's lack of control over his actions due to his emotional turmoil. The emotionally-loaded tone of the two occurrences of *non lo so* – and of the rest of Paolo's turn – is confirmed, among other things, by Paolo's aggressive delivery – see the indication "((aggressively))" in the transcription – as well as by Arianna's reply *Ma tu sei impazzito!*. Notice also that example 16 includes two further cases in which the phraseology of 'not knowing' is associated with the expression of emotional content or, in other words, is exploited in its emotional implications, namely *non so* in *Io senza di te non so starci!* and *non ci so* in *Io non ci so stare!*.

16. ARIANNA: Ma che cosa sei venuto a fare?

    PAOLO: ((aggressively)) **Non lo so**, **non lo so**! Sono solo venuto a dirti che passeranno i mesi, passeranno gli anni...forse mi odierai per tutta la vita, forse un giorno invece riuscirai a ripensarci! Allora io sarò qui ad aspettarti capito! Io sarò sempre qui ad aspettarti!

    ARIANNA: Ma tu sei impazzito!

    PAOLO: ((holding her face nervously)) Senti io senza di te non so starci! ((overlap)) Io non ci so stare!

    BOY 2: ((shouting)) ((overlap)) Lasciala!

    PAOLO: ((shouting)) Io ti amo! Io ti amo. Ti sei mai sentita come mi sento io adesso eh? Ci devi passare per capire. Ma che cazzo fai in mutande eh?

    ARIANNA: Paolo vattene! [*L'ultimo bacio*]

In sum, *I don't know* and *non lo so* seem to be of use in film dialogue not only to simulate unplannedness and to distribute information strategically, but also to express the characters' emotions, with this last function being ultimately intended to establish, maintain or increase the viewers' emotional involvement (cf. Kozloff 2000: 49-51; Quaglio 2009: 87-105; Bednarek 2012).

## 4. Conclusions

This paper has adopted a cross-linguistic approach to the analysis of film dialogue. The software package *WordSmith 6.0* has been used to identify the most frequent 3-grams in the original English component and the original Italian component of the PCFD. Once extracted, the English and the Italian 3-grams have been compared with one another with the specific aim of exploring their degree of comparability.

The results of the study indicate that the original English films and the original Italian films of the PCFD are similar in at least three respects. First of all, both the English and the Italian films of the PCFD have been shown to exhibit a markedly clausal 'texture'. Second, both the English and the Italian films of the PCFD have been found to be dominated by what this paper has referred to as the phraseology of interpersonality. Finally, it has been illustrated that the most frequent 3-gram in the English films, i.e. *I don't know*, is practically the same as the most frequent 3-gram in the Italian films, i.e. *non lo so*, with these 3-grams being comparable also in functional terms, particularly as markers of unplannedness, as tools to distribute narrative information strategically and as carriers of emotional content. Taken together, these results suggest that a high degree of comparability exists between the language of English films and the language of Italian films, with film dialogue, hence, appearing as a register having a rather rigid and distinct lexico-grammatical profile cross-linguistically (cf. Biber 1995; Taylor 2008; Veirano Pinto 2014; Zago 2016).

## BIBLIOGRAFIA

Bednarek M. (2011), "The language of fictional television: A case study of the 'dramedy' *Gilmore Girls*", *English Text Construction* 4(1), pp. 54-83.

Bednarek M. (2012), "Get us the hell out of here: Key words and trigrams in fictional television series", *International Journal of Corpus Linguistics* 17(1), pp. 35-63.

Bednarek M., Zago R. (2017), "Bibliography of linguistic research on fictional (narrative, scripted) television series and films/movies", version 1, available at http://unipv.academia.edu/RaffaeleZago.

Biber D. (1995), *Dimensions of Register Variation. A Cross-Linguistic Comparison*, Cambridge, Cambridge University Press.

Biber D., Johansson S., Leech G., Conrad S., Finegan E. (1999), *Longman Grammar of Spoken and Written English*, London, Longman.

Biber D., Conrad S., Cortes V. (2004), "If you look at…: Lexical bundles in university teaching and textbooks", *Applied Linguistics* 25(3), pp. 371-405.

Cortes V. (2008), "A comparative analysis of lexical bundles in academic history writing in English and Spanish", *Corpora* 3(1), pp. 43-57.

Culpeper J., Kytö M. (2010), *Early Modern English Dialogues*: *Spoken Interaction as Writing*, Cambridge, Cambridge University Press.

Freddi M., Pavesi M. (2009), "The Pavia Corpus of Film Dialogue: Research rationale and methodology", in *Analysing Audiovisual Dialogue. Linguistic and Translational Insights*, Freddi M., Pavesi M. (eds), pp. 95-100, Bologna, Clueb.

Granger S. (2014), "A lexical bundle approach to comparing languages: Stems in English and French", *Languages in Contrast* 14(1), pp. 58-72.

Kozloff S. (2000), *Overhearing Film Dialogue*, Berkeley/Los Angeles/London, University of California Press.

O'Keeffe A., McCarthy M., Carter R. (2007), *From Corpus to Classroom*: *Language Use and Language Teaching*, Cambridge, Cambridge University Press.

Pavesi M. (2014), "The Pavia Corpus of Film Dialogue: A means to several ends", in *The Languages of Dubbing. Mainstream Audiovisual Translation in Italy*, Pavesi M., Formentelli M., Ghia E. (eds), pp. 29-55, Bern, Peter Lang.

Quaglio P. (2009), *Television Dialogue*: *The Sitcom Friends vs. Natural Conversation*, Amsterdam/Philadelphia, John Benjamins.

Scott M. (2011), *WordSmith Tools version 6*, Liverpool, Lexical Analysis Software.

Spina S. (2014), "Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione", in *Proceedings of the First Italian*

*Conference on Computational Linguistics CLiC-it 2014* (vol. 1), Basili R., Lenci A., Magnini B. (eds), pp. 354-359, Pisa, Pisa University Press.

Stubbs M., Barth I. (2003), "Using recurrent phrases as text-type discriminators: A quantitative method and some findings", *Functions of Language* 10(1), pp. 61-104.

Taylor C. (2008), "Predictability in film language: Corpus-assisted research", in *Corpora for University Language Teachers*, Taylor Torsello C., Ackerley K., Castello E. (eds), pp. 167-181, Bern, Peter Lang.

Tracy-Ventura N., Cortes V., Biber D. (2007), "Lexical bundles in speech and writing", in *Working with Spanish Corpora*, Parodi G. (ed.), pp. 217-231, London, Continuum.

Veirano Pinto M. (2014), "Dimensions of variation in North American movies", in *Multi-Dimensional Analysis, 25 Years on. A Tribute to Douglas Biber*, Berber Sardinha T., Veirano Pinto M. (eds), pp. 109-147, Amsterdam, John Benjamins.

Zago R. (2016), *From Originals to Remakes. Colloquiality in English Film Dialogue over Time*, Acireale/Roma, Bonanno Editore