Investigating L2 spoken English through the role play learner corpus[1]

Andrea Nava
University of Milan
andrea.nava@unimi.it

Luciana Pedrazzini
University of Milan
luciana.pedrazzini@unimi.it

Abstract

We describe an exploratory study carried out within the University of Milan, Department of English the aim of which was to analyse features of the spoken English of first-year Modern Languages undergraduates. We compiled a learner corpus, the "Role Play" corpus, which consisted of 69 role-play interactions in English carried out by first-year students at B1+-B2 levels according to the *Common European Framework of Reference*. The analysis focused on the students' use of two features of spoken English grammar, tails and the discourse markers 'yes' and 'yeah'. Instances of these features from the data were compared with examples of British native speaker, learner and Italian native speaker usage. Preliminary findings pointed to the role of the students' first language, L2 proficiency and specific task features in the range and frequency of these phenomena as well as in the functions they deployed in the spoken discourse of the informants.

Keywords: learner corpus research, second language spoken discourse, spoken grammar, tails, discourse markers

---

[1] An earlier version of this paper was presented at the International Conference "First and Second Languages: exploring the relationship in pedagogy-related contexts", held at the Department of Education, University of Oxford, 27-28 March 2009.

The study of learner language has developed apace in the last few decades. Instrumental to the great strides made by second language acquisition research has been both the availability of learner corpora and the greater emphasis placed on the investigation of spoken discourse and how learners display L2 pragmatic use and develop L2 pragmatic knowledge (Alcon-Soler & Martinez-Flor, 2008). Advances in the application of learner corpora to the study of language learning processes have been mostly the result of the widening of the types of corpora available. As pointed out by Myles (2005), the fact that early learner corpora traditionally consisted of written cross-sectional data, often unannotated, conspired against their exploitation in SLA research. The last few years have, however, seen a reversal of this trend, as the number and size of learner corpora of spoken language have increased, such as the well-known LINDSEI project (Gilquin, De Cock, & Granger, 2010)[2], whose size has grown considerably due to the addition to the database of new L1 subcorpora. The wider availability of spoken corpora has prompted the exploration and description of the features which characterise spoken learner language, with a particular focus on spoken grammar and pragmalinguistic aspects of spoken interactions, in keeping with similar research being conducted with native speaker corpora (e.g., McCarthy, 1998; O'Keeffe, McCarthy, & Carter, 2007).

The first issue that corpus-based research into spoken English grammar has had to grapple with has been the definition of the status of the grammar of conversation with respect to the grammar of written language. Does English have only one overarching grammatical framework or is the grammar of speech so fundamentally different from that of writing as to qualify as an autonomous system? Valid arguments have been put forward in support of both the 'one grammar' and the 'two grammars' hypotheses.

Proponents of the first hypothesis (D. Biber, S. Johansson, G. Leech, S. Conrad and E. Finegan, authors of the groundbreaking corpus-based *Longman Grammar of Spoken and Written English* – Biber et al., 1999; also Leech, 2000) stress the fact that the grammatical differences between the registers of English are more a matter of quantity than of quality. The same phenomena are present in each variety; only their numerical distribution is different. For example, it is pointed out that even such features as dysfluencies and repetitions, which are usually viewed as a trademark of conversation, do indeed recur, albeit to a far lesser extent, in written English.

---

[2] The *Louvain International Database of Spoken English Interlanguage* (LINDSEI) is a corpus of spoken learner language featuring interviews with higher intermediate to advanced EFL learners from 11 mother tongue language backgrounds and contains over one million words.

The opposite camp is chiefly represented by the researchers that Leech (2000) has called the 'Nottingham school' (Ronald Carter and Michael McCarthy), who have developed the CANCODE corpus[3], as well as by David Brazil, the author of the *Grammar of Speech* (1995). These scholars maintain that categories rooted in a written model of grammar are often ill-suited to describe spoken English grammar. Traditional sentence-based syntax, for example, does not account for such phenomena as indeterminate and incomplete structures and jointly produced grammatical units, which are endemic to spoken English (Carter & McCarthy, 2001). The proponents of the 'two grammars' hypothesis point out that, to appreciate the distinguishing features of spoken grammar, it is crucial to start off by analyzing the context in which communication takes place (McCarthy, 1998) and the speakers' communicative needs and purposes (Brazil, 1995), in that, as McCarthy (1998, p. 78) has very cogently put it, it is "discourse" that "drives grammar, not the reverse".

This discourse-driven view of spoken grammar appears to have been particularly influential in recent corpus-based investigations of spoken native and learner English (O'Keeffe et al., 2007). At the same time recent investigations of spoken learner language associated with the field of interlanguage pragmatics – the area of second language acquisition concerned with the "the study of non-native speakers' use and acquisition of L2 pragmatic knowledge" (Kasper & Rose, 1999, p. 81) – have shifted their focus from traditional sociopragmatic areas, such as the study of the cross-cultural realization of speech acts, to include the study of pragmalinguistic features, such as L2 learners' use of information highlighting options and discourse markers (e.g., Callies, 2009; Callies & Keller, 2008; Romero Trillo, 2008).

This paper aims to contribute to research on spoken learner discourse by showing that learner corpus research can provide useful tools for the investigation of spoken grammar features of Italian university students' English. We report on an exploratory study carried out at the University of Milan whose aim was to identify two features of spoken English grammar — tails and the discourse markers 'yes' and 'yeah'— in a learner corpus of role-play interactions. These features were chosen in that they are among those items which do not show up in traditional EFL syllabuses and are thus not explicitly taught in mainstream EFL courses. As a consequence, learners may not be fully aware of their formal and functional characteristics in spoken English.

---

[3] The CANCODE (Cambridge and Nottingham Corpus of Discourse in English) is a five-million-word corpus of spoken British English developed at Nottingham University in the late 1990s. Data collection was based on five contexts given by the type of relationship of the speakers involved: transactional, professional, pedagogical, socializing and intimate (McCarthy, 1998).

In the applied linguistics literature, tails and discourse markers have also been associated with those pragmalinguistic features which O'Keeffe et al. (2007, p. 159) loosely refer to as "relational language", i.e., linguistic devices that "create and maintain good relations between the speaker and hearer". As examples of relational language, O'Keeffe et al. single out conversational routines (such as those used for appreciating, complimenting or thanking, e.g., *thank god for that, thank you ever so much, thanks for your help*), small talk, discourse markers, hedging expressions (used to lessen the directness of what is said, e.g., *kind of, like, sort of*), vagueness and approximation markers (e.g., *and that sort of thing, and so on and so forth*).

Corpus-based research summarized in O'Keeffe et al. (2007) has shown that relational language is pervasive even in more overtly transactional communicative exchanges among native speakers (such as service encounters), the overarching function of such relational language episodes being pre-eminently affective. Hence Carter and McCarthy (2003, p. 119) maintain that the appropriate display of relational language is one of the criteria which mark out the "successful user of English", whose performance is "judged (…) by how well they communicate, including how well they fit what they say to the needs of their listener(s)".

The Study

This study is part of a research project (Rizzardi, Pedrazzini, & Nava, 2008) which was started in the Department of English Language and Literature of Milan University in 2002. We had introduced the use of the role play as a teaching and testing technique with first year students of English and wanted to find out how effective it was. In particular, two main issues were at stake, that is whether the role play encourages the use of "natural" (Sinclair, 1984) language from learners and what features of spoken learner English are produced.

The topic of how task design variables influence fluency, accuracy and complexity of performance has been prominent in SLA researchers' agenda for some time now (Bygate, Skehan, & Swain, 2001; Ellis, 2003). The role play seems to conform to the conditions of tasks which are likely to promote fluency, that is to say, it provides contextual support, it is about familiar or involving topics, it poses a single demand, it is closed and has a clear inherent structure of the outcome (Ellis, 2003, p. 123). These features would then appear to encourage students to communicate quite naturally, once they have become involved in the task. Indeed, even within the constraints of playing given roles, students are still free to make their own linguistic choices and to organize their discourse as they want and are able to do.

How does the role play work as a testing technique at the University of Milan? The students work in pairs. They are given one role-card each. After reading it, they carry out a short conversation without being interrupted by the teacher. The role-play is bound to last about 5-10 minutes. The kind of context created in the role play encourages the participants to use both trans-actional and relational language (Figure 1).

Role-play – Student A
You are spending some time in London and you want to attend an English language course for two weeks. You have gone to enrol at a local language school. Explain the situation to the school secre-tary and enquire about:
- cost of the course/payment
- lessons: levels/number of lessons/timetable/materials/activities/etc.
- teachers
- extra-school activities
- procedure to enrol
Thank the secretary and close the conversation in a suitable way.

Role-play – Student B
You work in an English language school in London. A student comes to the school to enrol for a course. Be prepared to answer questions about:
- cost of the course/payment (where/how, etc.)
- lessons: levels/number of lessons/timetable/materials/activities/etc.
- teachers
- extra-school activities
- procedure to enrol (forms to complete/photographs/proof of identification, etc.)

Figure 1 Example of role play cards

First-year Milan university students' role-play interactions have been re-corded and transcribed with the aim to compile a small-scale corpus of spoken learner language – the Role-Play learner corpus.[4] The Role Play learner corpus gathers data from a fairly homogeneous group of learners: they are Italian mother tongue speakers and their level of proficiency is roughly intermediate ranging from B1+ to B2 according to the *Common European Framework of Reference* (Council of Europe, 2001). To date 69 interactions of first-year un-dergraduate students of English have been recorded. The corpus currently stands at approximately 50,063 words, but data assembly is still going on.

In this paper we report on the preliminary findings of the analysis of the occurrences in the Role Play learner corpus of tails and the discourse markers 'yes' and 'yeah'. Instances of these features from the Role Play corpus are compared with examples of British native speaker, learner and Italian native speaker usage from existing corpora.

---

[4] The role play interactions have been transcribed according to the guidelines for the LINDSEI.

The Use of Tails

"Tails" (Carter and McCarthy, 1997; McCarthy, 1998), also called "amplificatory tags" (Quirk, Leech, Greenbaum, & Svartvik, 1985), "tags" (Biber, Johansson, Leech, Conrad, & Finegan, 1999) and "right dislocation" (Huddleston & Pullum, 2002), are elements appended to the end of sentences[5] which display strong interpersonal meanings. At least three possible types of tails have been identified in the literature: noun phrase, declarative and question tags[6]:

The following extract, drawn from the CANCODE corpus, embodies a sequence of two types of tails:

*They do I suppose take up a lot of time **don't they kids***?

(CANCODE, Carter & McCarthy, 1997, p. 43*)*

The first tail (*don't they*) is the familiar 'question tag', with the inverted sequence of subject and auxiliary, whereas the final tail (*kids*) is a 'noun phrase tag', which elaborates the antecedent subject pronoun. Such clusterings of relational features are very common in native spoken English and have the effect of enhancing the affective impact of utterances.

An example of a 'declarative tag' is shown in the following CANCODE extract:

*She was a character **she was** really*

(CANCODE, Carter & McCarthy, 1997, p.66)

In this case, a tail is created by reiterating the subject and the main verb. In informal spoken English, tails often accompany evaluative statements such as the one embodied in the previous extract (Carter & McCarthy, 1997; McCarthy & Carter, 1997). Prosodically, tails are characterized by rising intonation (Aijmer, 1989; Quirk et al., 1985), with the exception of question tags, which admit either a rising or a falling intonation.

From the point of view of informativity, tails tend to be discourse-old while in their relational/interactional function they are mainly used as 'listener-sensitive' devices. By using a tail the speaker clarifies or emphasizes aspects of the message and the context (e.g., by restating a participant), eases processability (by shifting a 'heavy' constituent to the end of the utterance), and more generally establishes an affective bond with the listener.

Although relational language appears to be a pervasive feature of native spoken English, corpus-based studies have shown that the frequency of occurrence of tails is not particularly high. For example, the conversation section of

---

[5] Researchers disagree as to whether tails should be viewed as part of clause structure (Nava, 2005).

[6] We will use "tail" as the superordinate term and refer to each of the three subtypes with the label "tag".

the *Longman Spoken and Written English Corpus* (Biber et al., 1999) contains approximately 200 noun phrase tags per million words compared to 5500 occurrences per million words of the discourse marker *well*.

Tail-like phenomena also occur – albeit rather infrequently – in native spoken Italian. However, of the three types of tails identified above, only noun phrase tags appear to be attested in Italian, as in the following example:

<u>Le</u> *hai comprate.* **le acciughe**!

(Bazzanella, 1994, p. 125)

Like other features of spoken grammar, tails are routinely neglected by traditional EFL syllabuses and coursebooks. Brief mention is made in Swan (2005), a grammar and usage book for advanced learners and teachers and in two grammar practice books which draw on CANCODE materials (Carter, Hughes, & McCarthy, 2000; Nettle & Hopkins, 2003), while we found no evidence of this phenomenon being dealt with in major pedagogical grammar books for teachers (e.g., Celce-Murcia & Larsen Freeman, 1999).

Analysis of the Role Play corpus (Nava, 2005) shows that tails occur in Italian EFL learners' interlanguage sporadically – only 14 occurrences of tails were found in the whole 50,000-word corpus. Most instances were noun phrase tags used in an English native-like way in over 80% of cases. Only two occurrences of question tags were attested, while no instances of declarative tags were found. In this extract the speakers have been involved in a role play in which they discuss possible holiday destinations:

<B2> *to the mountains. oh no. I don't want to go absolutely to the mountain I prefer to go to the sea. it's more relaxing* <\B2>

<B1> <u>*it's*</u> *boring* **the sea** <\B1>

(Role Play Corpus)

B1 expresses an evaluation of one alternative (going to the seaside) in an utterance with a final noun phrase tag (*the sea*). The noun phrase tag in this extract accords with standard native English norms, as it is made up of a noun phrase which names the subject anaphorically.

A small number of instances of phenomena which look like noun phrase tags from a formal point of view, but are anomalous from a discoursal point of view were also produced, as in the following extract:

<B2> *I think he considers himself very handsome and so <laughs> I think it would be better for him* **something like cosmetics em items** *or I know in his house there's lot of mirrors and he likes cosmetics* <\B2>

<B1> *yes I know but I think it's it would be better* **a city map** <\B1>

(Role Play Corpus)

Although formally resembling standard English tails, the units in bold introduce new information, thus contravening one of the discoursal requirements

of standard English tails, i.e., that they should be discourse-old (Quirk et al., 1985). What appears to be happening here is that as the informants do not fully master the pragmalinguistic options that are available to English speakers to reconcile the requirements of information structure with the English grammatical word order, they rely on transfer from Italian of forms which appear to serve a similar function in Italian as English tails.

The functions expressed by the native-like examples of tails produced by Italian EFL students appear to be in keeping with their relational nature – with the affective bond-creating function being prominent in most of the occurrences, as clearly illustrated in the following example, in which the noun phrase tag (*hair*) is associated with an evaluative stance:

<B1> *also in* <name of place> *there is a male and a female hairdresser* <\B1>
<B2> *oh yeah but it's not important for you hair for me I have a long hair*
(laughs) <\B2>
(Role Play Corpus)

To sum up, it would appear that intermediate-level Italian speakers of English are moving towards native-like use of spoken grammar when a main strategy such end-focusing is considered, but to implement this strategy, they sometimes resort to features which deviate from native English speaking norms and are likely to be the result of first language influence. To account for these findings, we should point out that our informants had been learning English mainly in an Italian context, with little exposure to 'positive evidence', that is informal native spoken English, and had not been taught spoken grammar explicitly.

If we subscribe to the view that the interface between syntactic knowledge and discourse/pragmatic knowledge in interlanguage development occurs by means of mapping rules (Bos, Hollebrands, & Sleeman, 2004), it may be argued that Italian learners' path in the acquisition of the form, meaning and use (Larsen Freeman, 2003) of English tails involves a process of resetting of the relative strength of a focusing rule. In other words, Italian learners need to learn that focusing through the use of tails in English is subject to syntactic/pragmatic constraints which do not hold in Italian. As the resetting of the strengths of optional one-to more mapping rules is arguably (Sleeman, 2004) harder to achieve that the acquisition of a new, obligatory one-to-one mapping rule, the occurrence of non-native like pragmalinguistic features in Italian learners' role-play interactions is hardly surprising. Lack of exposure to informal spoken English or explicit awareness of the principles of information organization in English discourse may also have prevented the informants from 'blocking' the transfer of tail-like structures.

The Use of 'Yes' and 'Yeah'

In this second part of the paper we focus on two features of spoken grammar which seemed particularly relevant for the analysis of spoken learner discourse in the Role Play corpus: 'yes' and 'yeah' used as discourse markers[7]. The issue which will be considered here is to what extent a particular use of these markers is characteristic of our learners compared to other L2 learners and native speakers.

A great deal of research on discourse markers has been carried out in the past two decades and the term 'discourse marker' (e.g., Frazer, 1999; Schiffrin, 1987; Schourup, 1985) has been competing with other terms, such as 'discourse particles' (e.g., Aijmer, 2002), 'connectives' (e.g., Bazzanella, 1990), 'pragmatic markers' (e.g., Brinton, 1996), to mention just a few. 'Discourse marker' is probably the most frequently used term. Aijmer and Simon-Vandenbergen (2011, p. 226) argue that "'pragmatic marker' is, however, most commonly used as a general or umbrella term covering forms with a wide variety of functions both on the interpersonal and textual levels". This varied terminology employed in research reflects different approaches to the analysis of these linguistic items or expressions allowing a multitude of frameworks (Aijmer & Simon-Vandenbergen, 2011; Schiffrin, 2001,).

Schiffrin (1987, p. 31) operationally defines discourse markers as "sequentially dependant elements which bracket units of talk". According to Aijmer (2002, p. 2), they are "a class of words with unique formal, functional and pragmatic properties" and "dispensable elements functioning as signposts in the communication facilitating the hearer's interpretation of the utterance on the basis of various contextual clues". The most common discourse markers in everyday informal spoken language are single words such as *anyway, cos, fine, good, great, like, now, oh, okay, right, so, well*, and phrasal and clausal items such as *you know, I mean, as I say*, etc. (Carter & McCarthy, 2006, p. 208). They are drawn from different grammatical and lexical inventories so their classification in terms of the conventional word classes is problematic since "they stand outside of phrase and clause structure" and for this reason, "they are best considered as a class in their own right" (Carter & McCarthy, 2006, p. 209). The main functions which a discourse marker may perform are quite varied and "their literal meanings are 'overridden' by pragmatic functions in-

---

[7] Besides being used as a response form to questions, 'yeah' and its variants (yes, uh huh, mhm), also frequently serve as a response to a statement. According to Biber et al. (1999, p. 1091), "functionally, these inserts can be classified as backchannels because of their signalling feedback to the speaker that the message is being understood and accepted".

volving the speaker's relationship to the hearer, to the utterance or to the whole text" (Aijmer, 2002, p. 2). This constitutes one major difficulty for learners who have to deal with the different meanings of these markers according to their specific functions in discourse.

Drawing on previous studies, Fung and Carter (2007, pp. 412-414) define discourse markers by five main criteria: position (utterance initial for most of them but also flexible for some discourse markers); prosody (discourse markers should be independent from the utterances they introduce); multigrammaticality (due to their categorical heterogeneity); indexicality (they signal "the relation of an utterance to the preceding context"); optionality (which makes discourse markers "semantically and grammatically optional so that their existence does not affect the truth condition of the propositions"). Any of these criteria is a necessary but not sufficient condition for determining whether a linguistic item carries the status of 'discourse marker'. It follows that a combination of them together with sociolinguistic variables need to be considered.

Most research on discourse markers has focused on native English discourse and just a limited number of studies have been undertaken on the usage of discourse markers by non-native speakers, in particular second or foreign language learners (e.g., Aijmer, 2004; Fuller, 2003; Gilquin, 2008; Hasselgren, 2002; Müller, 2005; Pulcini & Damascelli, 2005; Pulcini & Furiassi, 2004; Romero Trillo, 2002). Compared with native speaker use, these studies suggest that learners use discourse markers according to different patterns of frequency and preference, and there is a strong correlation between usage and overall proficiency although even the most proficient learners do not use them to the degree that native speakers would do (e.g., Hellerman & Vergun, 2007 for a review of studies).

Discourse markers also seem to be an area where there has generally been a lack of focus by teachers, who normally tend to rely on "the notion of 'filler' or (non)explanations such as 'it doesn't really mean anything,' or 'you just have to learn it'" (Wichmann & Chanet, 2009, p. 24). Even dictionaries and some popular resource and methodology books for teachers often provide insufficient references to descriptions of the functions of these markers in different types of discourse (Callies, 2009; Hellermann & Vergun, 2007). Further, while there are EFL/ESL coursebooks which aim at the explicit teaching communication strategies in the spoken language, from a thorough analysis it emerges that "these communicative texts do not focus on students' discourse markers as objects for instruction in and for themselves" (Hellermann & Vergun, 2007, p. 162).

More recently, a study by Fung and Carter (2007) has shed new light on this area of research through the analysis and comparison of data from a pedagogic sub-corpus of CANCODE (460,055 words in size) and a corpus of inter-

active classroom discourse of secondary students in Hong Kong with the purpose of investigating the extent to which intermediate-advanced Chinese ESL learners are able to use discourse markers in their interactions[8]. In their adoption of a corpus-driven approach, the researchers emphasize "the descriptive value for classroom discourse of recurrent patterns and of frequency distribution" (Fung & Carter, 2007, p. 414), – an opinion likewise shared by the authors of this paper – and view discourse markers as "both pragmatically significant and socially sensitive". Following both Schiffrin's (1987) notion of a multi-dimensional model of coherence and Aijmer's (2002) interpersonal perspective, they propose a theoretical framework which categorizes discourse markers under four functional types: interpersonal, referential, structural and cognitive given that any instance may perform more than one of these functions.

Among the categories identified in Fung and Carter's framework, discourse markers of the interpersonal category seem to be of the utmost relevance for our analysis of relational language displayed in the Role Play corpus. Discourse markers with an interpersonal function are used to mark shared knowledge (*you know, you see, see, listen*), to indicate attitudes (*well, really, obviously, I think, absolutely, etc*.) and responses (*OK/ okay, oh, right/alright, yeah, yes, I see, great, ok great, sure*)[9] (Fung & Carter, 2007, p. 418). For example, in this extract from the pedagogic sub-corpus in CANCODE the speakers respond to each other at various points using *yeah* showing that they are "expressing a general acknowledgement of the preceding interactive unit":

[…]
<4> *Beans Means Heinz so why would you buy any other beans*. <G?>
<5> *You're kind of buying*
<2> **Yeah**
<5> *half a bean*.
<4> **Yeah.**
<5> <G?> *bean*.
<1> **Yeah.** <E> laughs <\E>
(CANCODE, Fung & Carter, 2007, p. 432)

Our analysis of the Role Play corpus data has focused on the use of the discourse markers *yes* and *yeah* investigating aspects of frequency and usage by our

---

[8] The Hong Kong learner corpus contains 14,157 words from group discussions of 49 intermediate-advanced learners of English in a secondary school in Hong Kong. The subjects were involved in a task in which they had to act as the staff of a toy company and needed to make a proposal about the type of toy they intended to manufacture (Fung & Carter 2007, p. 416).
[9] This third type of discourse markers are also defined "listener response tokens", underlying the active and responsive role of the listeners in conversation, that is to say their "listenership" (O'Keeffe et al., 2007, p. 141).

group of learners. Findings so far highlight more frequent usage of *yes* used in the similar prototypical interpersonal function as *yeah*. *Yes* was found to be the tenth most frequent word in the Role Play corpus compared to *yeah,* which occurs less frequently being the thirty-fourth most frequent word. The following extract illustrates an example of use of *yes* appearing in isolation in turn initial position with the main function of responding and possibly marking continuation:

    <B1> *I've seen an advertisement on a newspaper* <\B1>
    <B2> ***yes****.* <\B2>
    <B1> *it's about a house and* […] <\B1>
    <B2> *Mmh.* <\B2>
    <B1> a*nd it's very good because it's in an excellent condition* <\B1>
    <B2> ***yes*** <\B2>
    <B1> *and. it has two garage … they have two cars* <\B1>
    <B2> *ah* <\B2>
    <B1> *mmh. a big garden in front of the house* <\B1>
    <B2> ***yes*** <\B2>
    (Role Play corpus)

The higher number of instances of *yes* compared to *yeah* in the Role Play corpus seems to confirm the results discussed in similar studies on the use of spoken English by foreign language learners with different L1 backgrounds. For example, a comparison with the findings reported by Fung and Carter (2007, p. 431) reveals that "there is an over reliance of *yes* rather than *yeah* among the Hong Kong subjects". *Yes* was found to be the fourth most frequent word in the student corpus while *yeah* is the third least represented and the eighth most frequent discourse marker. On the other hand, *yeah* occurred as the third most frequent word in the pedagogic sub-corpus of CANCODE of native English. In the same vein, the higher frequency of *yes* over *yeah* is also attested in the analysis carried out by Pulcini and Furiassi (2004) on the Italian component of the LINDSEI, a learner corpus of spoken English amounting to a total of 79,264 words made up of 50 interviews to Italian speaking university students. We may conclude with Fung and Carter (2007, p. 431) that learners do not seem to exploit "the range of possibilities available with *yeah* that native speakers do as a way to exhibit understanding or acknowledgement (interpersonal category) or as a continuer of the progress of the primary speaker's turn (structural category)" but prefer its formal alternative *yes*.

The preference for the discourse marker *yes* may be accounted for by the learners' familiarity with this lexical item exclusively in its use as affirmative response form. It is also evident that learners are transferring the pragmatic meaning of the equivalent L1 discourse marker "sì" normally used in Italian with the same interpersonal function, as shown in these examples:

A. *di solito noi diamo più attenzione /,/*

B. ***sì***

A. *ad altri fenomeni secondari*!

(Bazzanella, 1994, p.157).

With regard to a direct influence from the learners' L1, O'Keeffe et al. (2007, p. 157) argue that "(…) while a form may have an equivalent in another language, it does not mean that it is directly transferable in all instances". In some cases this would lead to a pragmatic error, such as in the reduplication of *yes* in the following extract from the Role Play corpus in which one of the participants (B1), while agreeing with B2, may also unintentionally convey impatience:

<B2> ah *let's talk about Trani I've never been there and I wanna go how is* … <\B2>

<B1> *Oh **yes yes** it's a beautiful town* […] <\B1>

<B2> *Just like the. little town I've been* <\B2>

<B1> ***yes*** <\B1>

<B2> In Tuscany <\B2>

<B1> *Eh **yes yes yes yes** eh but I like Trani because* […]

(Role Play corpus)

In a study of the use of discourse-markers by native and non-native speakers of English, Romero Trillo (2002, p. 770) provides further explanation of this phenomenon hypothesizing a "binary track" for learners of a foreign or second language. The formal track is related to the acquisition of grammatical and semantic rules whereas the pragmatic track relates to the social use of language. Non-native learners in a non-target language environment would develop these two tracks mainly in instructional contexts in which the learning of certain forms will be contextualized and put into use at different subsequent stages. This would explain what he defines "pragmatic fossilization", that is the inappropriate use of certain forms at the pragmatic level of communication.

Conclusions

Although provisional, the findings from the analysis of the use of selected spoken grammar features by Italian EFL learners provide an interesting picture of the complex interplay of different phenomena which are shown to have a bearing on L2 spoken discourse. In particular, the study has attested the participants' difficulty using relational language in order to sound 'natural' English language users. This raises a fundamental issue tackled in previous studies (e.g., Carter & McCarthy, 1995), that is a need for explicit teaching of spoken grammar in L2 curricula in order to "strengthen learners' pragmatic competence in spoken language" providing them with opportunities of acquiring

explicit information about how to use the L2 in "culturally, socially, and situationally appropriate ways" (Fung & Carter, 2007, p. 433).

In the context of the undergraduate English courses at the University of Milan data from our spoken learner corpus have been used to encourage data-driven learning (Granger & Tribble, 1998; Mukherjee & Rohrbach, 2006; Nesselhauf, 2004), that is a kind of language learning based on discovery activities. In using learner corpora for pedagogic purposes, Aston (2008, p. 352) urges teachers to take "a more forgiving approach": "(…) we should not just mark up errors, which may be due to particular contextual constraints, but also communicative successes". The teacher might start off by presenting some instances taken from the learner corpus to draw students' attention to aspects of use of particular linguistic features. Then these instances can be compared with examples taken from native speaker corpora (both L2 and L1 spoken corpora). Discourse completion tasks, role play and real play can also provide opportunities for further communicative practice.

The pedagogic use of learner corpora highlights new opportunities for teacher development as it would give teachers the opportunity to find out what "they have always wanted to know" (Tsui, 2004) about language and their students' language learning. Discovery learning with both native and learner corpus data is "not only empowering for learners, but for teachers as well", and not only for "non-native speaker teachers" (Bernardini, 2004, p. 28) but, we would argue, for native speaker teachers too.

Although this small case study does not allow us to draw any strong conclusions, three factors appear to have emerged as having some bearing on how Italian learners of English use selected spoken grammar features and would thus merit further investigation: L1 influence, L2 proficiency and specific task features.

As shown in the previous extracts, the role of the learners' L1 emerges quite clearly. Given that the learners involved in the study had hardly experienced any focused instruction on the features of spoken English grammar or any sustained exposure to natural L2 spoken interactions, they tended to rely on L1 transfer as a default strategy when sophisticated pragmatic ability needed to be displayed. This type of process is also attested in studies on the development of L2 pragmatics and grammar in naturalistic contexts showing that in the very early stages adult learners "build on their available pragmatic knowledge, making do with whatever L2 grammar they have and at the same time acquiring the grammar needed to accomplish actions in L2" (Kasper & Rose, 2002, pp. 187-188).

The study also seems to provide evidence of the fact that a learner's level of L2 proficiency has some bearing on the likelihood that this transfer process will lead to accurate and appropriate L2 output. Indeed, in their role play interactions, the participants failed to block the transfer of forms which,

despite fulfilling the same function as tails and discourse markers in their L1, are not acceptable in standard native English.

A third factor which cannot be overlooked in a study of this kind is the role of the task itself. It has been pointed out that role plays may under-represent learners' pragmatic ability (Kasper & Rose, 2002, p. 89). Indeed, the fictive world of the role play may affect learners differently from native speakers. Whereas in authentic interaction, "participants' planning and execution of communicative action is supported by rich social context", participants in a role play may feel under pressure having to imagine their own and their co-actors' roles. It is thus possible that this will "reduce their capacity for online input processing and utterance planning".

References

Aijmer, K. (1989). Themes and tails: the discourse functions of dislocated. elements. *Nordic Journal of Linguistics, 12*, 137-154.

Aijmer, K. (2002). *English discourse particles.* Amsterdam: John Benjamins.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3,* 173-190.

Aijmer, K., & Simon-Vandenbergen, A. (2011). Pragmatic markers. In J. Zienkowski, J., Östman, & J. Verschueren (Eds.), *Discursive pragmatics* (pp. 223-247). Amsterdam: John Benjamins.

Alcon Soler, E., & Martinez Flor, A. (Eds.). (2008). *Investigating pragmatics in foreign language learning, teaching and testing*. Bristol: Multilingual Matters.

Aston, G. (2008). It's only human…. In A. Martelli, & V. Pulcini (Eds.), *Investigating English with corpora. Studies in honour of Maria Teresa Prat* (pp. 343-354). Monza: Polimetrica Publisher.

Bardovi-Harlig, K., & Griffin, R. (2005). L2 Pragmatic awareness: Evidence from the ESL classroom. *System, 33*, 401-415.

Bazzanella, C. (1990). Phatic connectives as interactional cues in contemporary spoken Italian. *Journal of Pragmatics*, *14*, 629-647.

Bazzanella, C. (1994). *Le facce del parlare. Un approccio pragmatico all'italiano parlato*. Firenze: La Nuova Italia.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S., &. Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bos, P., Hollebrandse, B., & Sleeman, P. (2004). The pragmatics-syntax and the semantics-syntax interface in acquisition. *IRAL, 42*, 101-110.

Brazil, D. 1995. *A grammar of speech.* Oxford: Oxford University Press.

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks: Second language learning, teaching and testing*. London: Longman.

Callies, M., & Keller, W. (2008). The teaching and acquisition of focus constructions: An integrated approach to language awareness across the curriculum. *Language Awareness, 17*, 249-266.

Callies, M. (2009). *Information highlighting in advanced learner English. The syntax-pragmatics interface in second language acquisition*. Amsterdam: John Benjamins.

Carter, R., & McCarthy, M. J. (1995). Grammar and the spoken language. *Applied Linguistics, 16*, 141-158.

Carter, R., & McCarthy, M. J. (1997). *Exploring spoken English*. Cambridge: Cambridge University Press.

Carter, R., & McCarthy, M. J. (2001). Size isn't everything: spoken English, corpus and the classroom. *TESOL Quarterly 35*, 337-340.

Carter, R., & McCarthy, M. J. (2003). If you ever hear a native speaker, please let us know. In A. Pulverness (Ed.), *Brighton conference selections* (pp. 116-123). Canterbury, UK: IATEFL.

Carter, R., & McCarthy, M. J. (2006). *Cambridge grammar of English: A comprehensive guide to spoken and written grammar and usage*. Cambridge: Cambridge University Press.

Carter, R., Hughes, R., & McCarthy, M. J. (2000). *Exploring grammar in context.* Cambridge: Cambridge University Press.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston: Thomson Heinle.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Eslami-Rasekh, A. (2005). Raising the pragmatic awareness of language learners. *ELT Journal, 59*, 199-208.

Eslami, Z. R., & Eslami-Rasekh, A. (2008). Enhancing the pragmatic competence of non-native English-speaking teacher candidates (NNESTCs) in an EFL context. In E. Alcon Soler, & A. Martinez Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 178-197). Bristol: Multilingual Matters.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*, 931-952.

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: native and learner use in pedagogic settings. *Applied Linguistics, 28*, 410-429.

Fuller, J. M. (2003). Discourse marker use across speech contexts: a comparison of native and non-native speaker performance. *Multilingua 22*, 185-208.

Gilquin, G.(2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference? In J. Romero-Trillo (Ed.), *Corpus and pragmatics*. *A mutualistic entente* (pp. 119-149). Berlin and New York: Mouton de Gruyter.

Gilquin, G., De Cock, S., & Granger, S. (2010). *Louvain international database of spoken English interlanguage.* Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora,*

second language acquisition and foreign language teaching (pp. 3-33). Amsterdam: John Benjamins.

Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on computer* (pp. 119-131). Addison Wesley Longman: London and New York.

Hasselgren, A. (2002). Learner corpora and language testing: small-words as markers of learner fluency. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam: John Benjamins.

Hellermann J., & Vergun A. (2007). Language which is not taught: the discourse marker use of beginning adult learners of English. *Journal of Pragmatics, 39,* 157-179*.*

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Larsen Freeman, D. (2003). *Teaching language. From grammar to grammaring*. Boston: Heinle.

Leech, G. (2000). Grammars of spoken English: New outcomes of corpus-oriented research. *Language Learning, 50,* 675-724.

Kasper, G. (2007). Pragmatics in second language learning: An update. Remark addressed in the 2007 American Association for Applied Linguistics (AAAL) annual conference. Costa Mesa, California.

Kasper, G., & Rose, K. R. (1999). Pragmatics and SLA. *Annual Review of Applied Linguistics, 19*, 81-104.

Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.

Kondo, S. (2008). Effects on pragmatic development through awareness-raising instruction: Refusals by Japanese EFL learners. In E. Alcon Soler, & A. Martinez Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 153-177). Bristol: Multilingual Matters.

McCarthy, M. J. (1998). *Spoken language and applied linguistics.* Cambridge: Cambridge University Press.

McCarthy, M. J., & Carter, R. (1997)*.* Grammar, tails and affect: constructing expressive choices in discourse. *Text, 17*, 231-252.

Mukherjee, J., & Rohrbach, J. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: new departures in learner corpus research. In B. Kettemann, & G. Marko (Eds.), *Planning, gluing and painting corpora: Inside the applied corpus linguist's workshop* (pp. 205-232). Frankfurt am Main: Peter Lang.

Müller, S. (2005). *Discourse markers in native and non-native English discourse.* Amsterdam: John Benjamins.

Myles, F. (2005). Interlanguage corpora and SLA research. *Second Language Research*, *21*(4), 373-391.

Nava, A. (2005). Comparing tails. An exploratory study of tails in native spoken English and in Italian EFL learners' interlanguage. *Mots Palabras Words, 6*, 71-92.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125-152). Amsterdam: John Benjamins.

Nettle, M., & Hopkins, D. (2003). *Developing grammar in context.* Cambridge: Cambridge University Press.

O'Keeffe, A., McCarthy, M. J., & Carter, R. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.

Pulcini, V., & Damascelli, A. (2003). A corpus-based study of the discourse marker okay. In A. Bertacca (Ed.), *Historical linguistic studies of spoken discourse* (pp. 231-243). Pisa: Edizioni Plus.

Pulcini, V., & Furiassi, C. (2004). Spoken interaction and discourse markers in a corpus of learner English. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and discourse* (pp. 107-123)*.* Bern: Peter Lang.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Rizzardi, M.C., Pedrazzini, L., & Nava, A. (2008). The role-play learner corpus: A resource for investigating learner language. In A. Martelli, & V. Pulcini (Eds.), *Investigating English with corpora. Studies in honour of Maria Teresa Prat* (pp. 379-398). Monza: Polimetrica Publisher.

Romero Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English*. Journal of Pragmatics, 34*, 769-784.

Romero Trillo, J. (Ed.). (2008). *Pragmatics and corpus linguistics*. Berlin/New York: Mouton De Gruyter.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D. (2001). Discourse markers: Language, meaning and context. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 54-74). Oxford; Malden, MA: Blackwell Publishers.

Schourup, L. C. (1985). *Common discourse particles in English conversation.* New York: Garlands Publishing.

Sleeman, P. (2004). Guided learners of French and the acquisition of emphatic constructions. *International Review of Applied Linguistics, 42*, 129-151.

Sinclair, J. (1984). Naturalness in language. In J. Aarts, & W. Meijs (Eds.), *Corpus linguistics: recent developments in the use of computer corpora in English language research* (pp. 203-210). Amsterdam: Rodopi.

Swan, M. (2005). *Practical English usage* (3rd ed.). Oxford: Oxford University Press.

Tsui, A. (2004). What teachers have always wanted to know — and how corpora can help. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 39-61). Amsterdam: John Benjamins.

Wichmann, A., & Chanet, C. (2009). Discourse markers: a challenge for linguists and teachers. *Nouveaux cahiers de linguistique française, 29,* 23-40.