A corpus-based study of pragmatic markers at CEFR level B1

Um estudo de marcadores pragmáticos no nível CEFR B1 baseado em linguística de corpus

Cristiane Ruzicki Corsetti
Cristina Lopes Perna
Pontifícia Universidade Católica do Rio Grande do Sul – Porto Alegre, RS, Brasil

Abstract: This article summarises the findings of a corpus-based study of pragmatic markers. The study comprised quantitative and qualitative analyses of a small specialised corpus of Brazilian learners' oral production in English at CEFR B1 and comparisons with benchmark corpora. It examined the most common discourse marking adverbs used to mediate segments of discourse in conversations, the most common explicit and implicit adverbial hedges used to mitigate representative speech acts and the most common minimal response tokens used to express good listenership. Subjects produced a limited range of discourse marking adverbs and response tokens. Investigations followed a form-to-function approach and indicated that learners underused the pragmatic and discourse functions of "*well*" and "*actually*" but demonstrated a consistent use of "*really*"; they overused "*maybe*" to convey epistemic stance and underused "*just*" as an implicit adverbial hedge; they produced "*yeah*" and "*uhuh*" as minimal response tokens, mostly to convey convergence.

Keywords: Pragmatic markers; Corpus-based investigations; CEFR B1; Brazilian learners; Conversational competence.

Resumo: Este artigo detalha os resultados de um estudo de marcadores pragmáticos baseado em linguística de corpus. O estudo envolveu análises quantitativas e qualitativas de um pequeno corpus especializado, com a produção oral de aprendizes brasileiros de inglês no nível CEFR B1, e comparações com corpora de referência. Examinou-se os advérbios mais comuns utilizados para mediar segmentos de discurso, os "hedges" adverbiais explícitos e implícitos utilizados para mitigar atos de fala representativos e as partículas de resposta mínimas utilizadas pelo interlocutor para expressar boa receptividade. Os sujeitos produziram número limitado de marcadores de discurso e de partículas de resposta. As investigações analisaram funções de marcadores específicos e indicaram subutilização das funções pragmáticas e discursivas de "*well*" e "*actually*", porém uso consistente de "*really*", sobreutilização de "*maybe*" como forma de posicionamento epistêmico, subutilização de "*just*" como mitigador. Os aprendizes utilizaram "*yeah*" e "*uhuh*" como partículas de resposta mínimas, principalmente para convergir.

Palavras-chave: Marcadores pragmáticos; Investigações baseadas em Linguística de Corpus; CEFR B1; Aprendizes brasileiros; Competência conversacional.

Introduction

This article summarises the results of a corpus pragmatic study, carried out as part of a doctoral thesis. Corsetti (2015) addressed the theme "conversational competence as a second language" due to its relevance to teachers preparing Brazilian learners of English for international examinations. For the empirical project, she focused on level B1 from the Common European Framework for Language Reference (COUNCIL OF EUROPE, 2001) so as to tackle the intermediate plateau phenomenon.

Corsetti's aims (2015) were three-fold: to revisit the construct "conversational competence" in L2 and place it within a pragmatic domain; to detail key discourse, pragmatic and interactional phenomena which comprise conversational competence in L2; to investigate some of the key discourse, pragmatic and interactional phenomena

comprising conversational competence in the oral production of twenty Brazilian learners of English at CEFR B1.

The study dealt with three research questions:

a) What are the most common discourse marking adverbs used to mediate segments of discourse in conversations by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?

b) What are the most common explicit and implicit adverbial hedges used to mitigate representative speech acts by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?

c) What are the most common minimal response tokens used to express good listenership by Brazilian learners at CEFR B1? How do their frequency of use and pragmatic functions compare to those of native speakers?

Investigations followed a form-to-function approach and indicated that subjects produced a limited range of discourse marking adverbs and of response tokens in general. Specific analyses and results will be detailed in section 3.

Theoretical background

In previous communicative competence frameworks (CANALE & SWAIN, 198O; CANALE, 1983; BACHMAN, 1990), conversational language use is addressed within competences that account for both oral and written discourse. However, these modes of discourse differ tremendously vis-à-vis the demands they make on language producers (BROWN & YULE, 1983). The prototypical characteristics of spoken conversation (CULPEPER & KYTO, 2010) corroborate the uniqueness of spoken discourse, especially when one considers its real-time processing and interactivity. Based on such differences, Corsetti (2015) argues that conversational competence deserves a detailed framework for its characterisation as a separate component.

Corsetti (2015) places conversational competence within a pragmatic domain grounded on two assumptions: conversation is prototypical of language usage (LEVINSON, 1983); language in use involves the negotiation of meaning (THOMAS, 1995). Second language learners perform language functions, convey and negotiate meanings through spoken discourse, relying on the complex interplay between discourse, pragmatic and interactional features.

Corsetti (2015, p. 104-108) then proposes a framework for the characterisation of conversational competence which consists of three interrelated facets: the management of discourse, the negotiation of illocutionary meaning and the deployment of conversational practices.

The management of discourse includes a learner's ability: to produce and sustain (longer) stretches of discourse, including the use of cohesive devices, co-reference, substitution, ellipsis and lexical relationships; to produce coherent discourse by adhering to thematic and information structure, which may be strengthened by cohesive relations; to employ discourse markers to launch and conclude topics, to signal relationships of sequence, to monitor and manage the ongoing discourse, to indicate that they have not selected the most appropriate way of expressing things and that they are adding to or refining what they say, to signal that they are sensitive to listeners needs and that they are monitoring the state of shared knowledge and to mark their stance or attitude towards the message.

The negotiation of illocutionary meaning involves a learner's ability: to produce illocutionary acts according to the intended illocutionary force; to employ positive and negative politeness strategies in order to minimise the degree of imposition of face-threatening acts and to produce face-saving acts; to use pragmatic force modifiers so as to soften or strengthen the force of their messages and to make concepts fuzzier or less fuzzy; to interpret the illocutionary force of utterances, based on pragmalinguistic and sociopragmatic knowledge, illocutionary force indicating devices and contextual clues.

The deployment of conversational practices refers to a learner's ability: to carry out small talk as a means to initiate a conversation, including conversational routines; to orient themselves to the rules of turn-taking; to construct sequences of utterances by reference to the practices of adjacency pair organisation; to show an orientation to preference structure; to employ backchannels in order to indicate that they are following what is being said and also to inform how the message is being received.

In order to investigate key discourse, pragmatic and interactional phenomena of her proposed framework in subjects oral production, Corsetti (2015) follows co-joint methods. According to O'Keeffe and Walsh (2012), Conversation Analysis and Corpus Linguistics have different origins and focus on different research objects. However, both fields share methodological similarities and can be used as complementary approaches: they use empirical, naturally occurring data; they refer to baseline comparisons with other types of interactions (sequential order and reference corpora, respectively); they investigate language in its social context; analyses are carried out at turn level for a better understanding of context.

Furthermore, some discourse, pragmatic and interactional phenomena are triggered by individual

lexical items (e.g. pragmatic markers), lexical bundles, clause patterns and formulae, and thus, may be retrieved by means of corpus searches. In a broad sense, Corpus Pragmatics refers to the "studies of language use that employ large, computer-readable collections of language" (JUCKER, 2013, p. 1). Such studies are quite recent and differ from philosophical approaches to pragmatics as they are placed within empirical data-based pragmatics.

"Form-to-function mapping" refers to investigations of particular linguistic entities and their functions, for instance, the different uses of a specific discourse marker. Form-to-function research questions lend themselves more easily to corpus linguistics investigations since it is relatively easy to translate them into a search string. "Function-to-form mapping", on the other hand, comprises investigations of specific functions and the range of linguistic devices employed to perform them, for instance, the realisations of a specific speech act.

Corsetti (2015) analyses the functions of pragmatic markers that lend themselves to corpus investigation and seem to be prototypical of the three proposed facets of conversational competence (despite some possible overlap), namely discourse marking adverbs, adverbial hedges and backchannels.

A corpus-pragmatic study

The study encompassed both quantitative and qualitative analyses of empirical data retrieved from native speaker and learner corpora. It counted on the participation of twenty Brazilian adult learners of English at CEFR B1, who attended a general English course at a language institute in the South of Brazil between the second semester of 2012 and the second semester of 2013. In order to collect learner data, a public version of the FCE speaking test was selected, based on two main reasons. As a formal assessment instrument, it is designed and validated to give learners of English the opportunity to produce enough samples of language to have their speaking ability assessed in four analytical criteria. Secondly, it is delivered in a paired format, giving them the opportunity to display their conversation management ability.

The recordings of the ten FCE interviews amounted 2 hours 40 minutes and 14 seconds. The interviews were manually transcribed following the conventions for transcribing audio files suggested by O'Keeffe, McCarthy and Carter (2007, p. 6), which included extra-linguistic information, filled pauses and backchannels. Learners' oral production was reproduced as originally uttered by them. The final outcome was a small specialised corpus of 23,803 word tokens, displaying Brazilian learners' oral production at CEFR B1.

Spoken sub-corpora derived from "The BNC Sampler" and "The Diachronic Corpus of Present-Day Spoken English" were selected as native speaker benchmark corpora. As subjects had been exposed to textbooks and classroom materials mostly based on the British variety, it felt methodologically adequate to focus on British English native speaker corpora available on CQPweb for comparisons. The spoken sub-corpus of the BNC Sampler contains 1,140,055 words distributed in 98 texts. The Diacronic Corpus of Present-day Spoken English metadata information was used as a filter in the compilation of a specialised sub-corpus re-named as "face-to-face conversations from the 1990s", containing 271,360 words distributed in 110 texts.

Comparisons of relative frequencies derived from the three corpora were normalised to 10,000 due to the differences in corpus size. The benchmark corpora investigations relied on the CQPweb server while the Brazilian learner corpus was analysed using the free software AntConc 3.4.1 (ANTHONY, 2014). Other native speaker corpora and learner sub-corpora (LINDSEI) were referred to for specific purposes.

Discourse marking adverbs: data analysis

Corsetti's subjects (2015) produced the following most frequent adverbs that fit the characterisaton of discourse markers: "*maybe*", "*really*", "*just*", "*well*", "*probably*" and "*actually*". The adverbs "*really*", "*well*" and "*actually*" were chosen for the investigation of discourse markers since they are commonly used to mediate segments of discourse with regards to the different parts of discourse and also to the relations between speakers and parts of discourse. The adverbs "*maybe*", "*just*" and "*probably*" are traditionally associated with mitigation and were investigated as hedges.

In the analysis of the discourse marking adverb "*really*" (CORSETTI, 2015, p. 178-190), the comparison of relative frequencies (per 10,000 words) between the native speaker benchmark corpora and the Brazilian learner corpus indicated a consistent use by subjects: BNC Sampler spoken sub-corpus (15.3501); Face-to-face conversations from the 1990s (29.0389); Brazilian learner corpus (19.7454).

Learners were able to employ "*really*" to perform all the pragmatic categories proposed by Paradis (2003). "*Really*" was more frequently used as a pragmatic marker modifying speaker meaning, considering the degree reinforce (42.55%), emphasiser (42.55%) and de-emphasiser functions (6.38%). While the degree reinfoncer "*really*" reinforces scalar properties denoted by adjectives, the emphasiser "*really*" has the effect of adding subjective emphasis of situations and accompanies

attitudinal verbs. As for discourse properties, truth attester "*really*" has an epistemic function, since it ensures the truth of the assertion that it takes in its scope and provides implicit evidence based in reality. Truth attester "*really*" was observed in utterance initial and final positions mediating segments of discourse, but with a lower percentage (8.52%).

Further comparisons between native speaker data (PARADIS, 2003) and the Brazilian learner data showed that both used "*really*" mostly as a pragmatic marker modifying speaker meaning within utterances, either to emphasise scalar properties of adjectives or to reinforce the subjectivity of situations when accompanying attitudinal verbs. Native speakers produced truth attester "*really*" more frequently than Brazilian learners did, which suggests that subjects were less familiar with this discourse function of "*really*".

As for "*well*" (CORSETTI, p. 169-178), the comparison of relative frequencies confirmed considerable underuse of the discourse marker by subjects: BNC Sampler spoken sub-corpus (54.3833); Face-to-face conversations from the 1990s (63.1264); Brazilian learner corpus (3.7810).

All occurrences of "*well*" portrayed its use as a discourse marker, mediating segments of discourse: choice related (for pausing and planning what to say next 44.5%), change related (for repair 22.2%) and opinion (for conveying stance 33.3%). "*Well*" was mostly used for speech management functions and often preceded by filled and unfilled pauses in initial position or in medial position, being used as a floor holding device, similarly to Aijmer's results (2011).

Comparisons between native speaker data (AIJMER 2011) and the Brazilian learner corpus demonstrated that both used "*well*" more frequently for speech management functions. However, native speakers used "*well*" in the change-related function more frequently than Brazilian learners did while the latter used "*well*" in the choice-related function with a higher percentage among functions than native speakers did. Percentages for attitudinal "*well*" were also higher in native speaker data.

Lastly, the discourse marking adverb "*actually*" (CORSETTI, p. 190-200) was also underused by subjects as indicated in the comparison of relative frequencies: BNC Sampler spoken sub-corpus (10.324); Face-to-face conversations from the 1990s (18.9784); Brazilian learner corpus (1.6804).

Subjects used "*actually*" mostly as a discourse marker. "*Actually*" was placed in initial position in 3 occurrences to introduce one's opinion, conveying different attitude to the preceding discourse. The single occurrence of "*actually*" in medial position displayed its use as a mitigating device. All 4 occurrences were followed by "*I*" and verbs expressing preferences.

Aijmer's native speaker findings (2002, p. 260) show that the most frequent use of "*actually*" is in medial position, being employed as an adverb. Conversely, the most typical use of the discourse particle is in final position, being three times more frequent than in initial position. "*Actually*" in final position was not observed in the Brazilian data.

Explicit and implicit adverbial hedges: data analysis

According to Nikula (1996, p. 50-55), pragmatic force modifiers refer to modifying devices that speakers use either to soften or to strengthen the force of their messages. The author draws a distinction between explicit and implicit pragmatic force modifiers based on the semantic-pragmatic divide. Explicit modifiers include attitudinal adverbs, parenthetical constructions or adverbs of degree and indicate quite explicitly the degree to which speakers are committed to the truth or preciseness of their messages. Conversely, implicit modifiers have a tendency to remain ambiguous or fuzzy even in the context, often making various interpretations possible.

Subjects (CORSETTI, 2015, p. 212-254) produced a range of word forms traditionally associated with mitigation such as attitudinal verbs, modal verbs, nouns and epistemic and restrictive adverbs. However, not all forms were necessarily used for mitigating purposes as demonstrated in the analyses of concordance lines.

The attitudinal verb "*think*" was the form associated with mitigation most frequently produced subjects. 215 occurences (90.3247 relative frequency per 10,000 words) out of 245 included the collocation "*I*" + "*think*". This formula is widely used by Brazilian learners in general as a means to introduce their opinion rather than as a hedging device. Another typical form used to introduce learners' opinion is the fixed expression "*In my opinion*", also present in the Brazilian learner corpus, but with a lower frequency.

The second most frequent form was the modal verb "*can*", found mostly in contexts of suggestions, expressing possibility and ability. The distal forms "*could*" and "*should*" were used with a much lower frequency. "*You can*" was the most common collocation with 71 occurrences (29.8281 relative frequency per 10,000 words) expressing mostly possibility. Only 14 occurrences of "*I can*" were found, out of which 11 expressed ability (mostly followed by sensory verbs as a marker of sensory evidence) and 3 ocurrences conveyed possibility.

The third most frequent form was the lemma "*things/thing*". All occurrences of "*things*" displayed

its use for expressing vagueness within categories. The most common collocating items with "*things*" were as follows: "***the*** *things*" (11 occurrences), "***a lot of*** *things*" (4 occurrences) and "***some*** *things*" (3 occurrences). Other forms for expressing vagueness included "*kind of*", "*a bit*", "and "*types/type*", but with lower frequencies. As for the singular form "*thing*", it also expressed vagueness.

The adverbial forms "*maybe*", "*probably*" and "*just*" were found mostly within contexts of representative speech acts such as asserting, making deductions and concluding and were selected for the empirical investigations of hedges.

The Brazilian learner corpus displayed 71 occurrences of "*maybe*" and 13 occurrences of "*probably*" (CORSETTI, 2015, p. 221-235). Relative frequencies (per 10,000 words) were as follows: BNC Sampler spoken sub-corpus "*probably*" (4.5962) and "*maybe*" (2.7103); Face-to-face conversations from the 1990s "*probably*" (8.0336) and "*maybe*" (3.6114); Brazilian learner corpus "*probably*" (5.4614) and "*maybe*" (29.8281).

The comparison of relative frequencies between the native speaker benchmark corpora and the Brazilian learner corpus indicated that "*probably*" occurred more frequently than "*maybe*" in both British English native speaker corpora. Subjects, on the other hand, used "*maybe*" more frequently than "*probably*", differing from British native speaker linguistic behaviour but aligned with American English patterns (BIBER et al., 1999).

Different reasons may account for subjects' favouritism towards the epistemic adverb "*maybe*". Brazilian learners in general have more exposure to the American variety: Brazil is geographically closer to the USA, which may facilitate more frequent visits to the USA; most Brazilian schools (both regular schools and language institutions) adopt American English textbooks; a regular and more frequent access to American movies and sitcoms.

Moreover, the high frequency of "*maybe*" indicates overuse. Subjects did not produce a single occurrence of "*may*" or "*might*" for expressing degrees of certainty. One can conclude that learners resorted to the epistemic adverb "*maybe*" (and possibly "*probably*" with a much lower frequency) to perform this function. This high frequency may be due to transfer as in Brazilian Portuguese, epistemic stance is often conveyed through the adverb "*talvez*", the direct translation of "*maybe*".

Qualitative analyses indicated that subjects employed "*probably*" and "*maybe*" to perform the following functions:

1. "*Probably*": mitigating the degree of certainty of propositions (77.00%); marking uncertainty when discussing pros and cons (23.00%).

2. "*Maybe*": mitigating the degree of certainty of propositions (47.89%); marking uncertainty when discussing pros and cons (26.77%); softening preferences (16.91%); mitigating suggestions (2.81%); planning ahead (2.81%); politeness disagreement (2.81%).

Both epistemic adverbs were employed as explicit hedges to mitigate representative speech acts. All occurrences of "*probably*" displayed its use as an explicit adverbial hedge used to mitigate assertions. 94.38% of the occurrences of "*maybe*" also showed its use to mitigate assertions, mitigating the degree of certainty of propositions, marking uncertainty when discussing pros and cons of suggested ideas, softening individual preferences and as a politeness hedge when expressing disagreement with proposed ideas. 2.81% of the occurrences of "*maybe*" demonstrated its use mitigating another speech act category, namely directives (suggestions).

As for the third adverbial hedge under investigation (CORSETTI, 2015, p. 235-247), subjects produced 17 occurrences of "*just*". Comparisons with native speaker data attested an extremely high frequency of "*just*" in the benchmark corpora and indicated low figures in the Brazilian learner corpus, which signals underuse. Both native speakers and learners produced "*just*" mostly in medial position. Relative frequencies were as follows: BNC Sampler spoken sub-corpus (34.0685); Face-to-face conversations from the 1990s (43.0424); Brazilian learner corpus (7.1419).

Subjects employed "*just*" in the following functions: emphatic, including downtoning (52. 94%); restrictive adverb (41.18%); within a conditional clause (5.88%). "*Just*" was more frequently used to emphasise elements within utterances as a means to modify their pragmatic force (mitigating). 8 out of 9 occurrences diplayed its use as an implicit hedge mitigating assertions. There was only 1 occurrence of "*just*" strengthening assertions. The effect of "*just*" on the pragmatic force of utterances was sometimes fuzzy, even within its contexts. The second most common function was its use as a restrictive adverb, taking scope over the truth condition of particular elements within the utterance. The empathic function of "*just*" seems to be the unmarked use by subjects, similarly to Aijmer's findings (2002) in native speaker corpora.

Minimal response tokens: data analysis

Backchannels are "verbal and non-verbal devices used to provide feedback and other supportive responses, normally as a way to encourage the speaker to continue" (CARTER & MCCARTHY, 2006, p. 892). "Minimal

response tokens" include short utterances and non-word vocalisations (O'KEEFFE & ADOLPHS, 2008). They are minimal forms as they have no homonyms in other word classes. "Non-minimal response tokens" include adverbs or adjectives functioning as pragmatic markers and short phrases or minimal clauses (O'KEEFFE & ADOLPHS, 2008).

O'Keeffe and Adolphs (2008, p. 84) adopt the following framework for classifying response tokens in casual conversation: continuer tokens (they maintain the flow of the discourse); convergence tokens (they mark agreement and convergence); engagement tokens (they mark high engagement where addressees respond on an affective level to the content of the message); information receipt tokens (they mark points in the discourse where adequate information has been received).

As adverbs had already been addressed as discourse markers and hedges, Corsetti (2015, p. 259-294) focused on minimal forms to avoid possible overlap. AntConc's word list tool was used to view the rank positions and the raw frequencies of all the words in the learner corpus. This produced a manual list of all the items that fit the form categorisation of minimal response tokens as described by O'Keeffe and Adolphs (2008). Adjectives and adverbs were disregarded. The list did not include "*yes*" and "*no*" tokens as they are traditionally found within answering moves to "*yes*" or "*no*" questions. Tokens with fewer than 4 occurrences were not included either.

The Brazilian learner corpus displayed the following most frequent tokens fitting the description of minimal forms: "*uh*" (351.2162), "*yeah*" (51.6741), "*hum*" (16.8955), "*ah*" (11.7632), "*uhuh*" (7.5620), "*oh*" (3.781) and "*aih*" (1.6804). Relative frequencies (per 10,000 words) relate to the total number of occurrences of each token rather than their function as response tokens.

A closer look at the contexts of usage of individual forms revealed that "*yeah*" was the only short utterance produced by subjects of as a response token, with 37 occurrences as one-word inserts. "*Uhuh*" was the non-word vocalisation most frequently used as a response token: 17 out of 18 occurrences of "*uhuh*" were placed in initial position, 12 of which were one-word inserts. As a result, "*yeah*" and "*uhuh*" were chosen for the empirical investigations.

Intonation played a part in the interpretation of minimal response forms. Although the Brazilian learner corpus is not prosodically annotated, the manual transcription of interviews required extracts to be played several times. The investigation of the functions of "*yeah*" and "*uhuh*" also included listening to specific extract parts again.

Firstly, the study (CORSETTI, 2015, p. 269-281) analysed 110 occurrences of "*yeah*", out of which 92

were in utterance initial position and 18 occurred immediately after some non-content turn-preface items or immediately after a lexical repetition in utterance initial position. Native speaker data attested a far higher frequency of "*yeah*" in utterance initial position than in medial and final positions. Subjects produced "*yeah*" less frequently than native speakers did but they displayed a consistent use of "*yeah*" as a listener response token in utterance initial position. Relative frequencies as shown below:

a) BNC Sampler spoken sub-corpus: utterance initial position (57.7691), other positions (19.4201);
b) Face-to-face conversations from the 1990s: utterance initial position (60.8785), other positions (14.3352);
c) Brazilian learner corpus: utterance initial position (46.2126), other positions (5.4614).

The functions of "*yeah*" as a listener response token in utterance initial position in the Brazilian data can be summarised as follows: convergence token (88.19%); continuer token (1.82%); information receipt token (0.90%). Other uses of "*yeah*" in utterance initial position included the following functions: similar to "*yes*" within answering moves (8.19%); as a compensation strategy (0.90%).

Subjects produced "*yeah*" in utterance initial position largely as a response token (90.91%). In addition, 88.19% of the occurrences in initial position displayed its use as a convergence token, signalling a tendency for agreement within preference structure. Some occurrences of "*yeah*" as a convergence token indicated its use as an informal politeness formula employed to reinforce common ground despite disagreement.

Secondly, the investigation (CORSETTI, 2015, p. 281-287) focused on 17 occurrences of "*uhuh*" in utterance initial position in the Brazilian learner corpus: as one-word insert sequences followed by full stops (14 occurrences); followed by commas (2 occurrences); followed by paralinguistic elements (1 occurrence).

The distribution of "*uhuh*" among functions in utterance initial position in the Brazilian data was as follows: convergence token (47.05%); continuer token (35.30%); information receipt token (17.65%).

All occurrences of "*uhuh*" in utterance initial position showed its use as a listener response token. Subjects employed "*uhuh*" more frequently as a convergence token. Nevertheless, its use as a continuer token was also evident. There were occasional occurrences of "*uhuh*" as an information receipt token.

Due to insufficient native speaker corpus evidence of "*uhuh*", its relative frequency was compared to that of more advanced learners (Spanish component of LINDSEI). Subjects produced "*uhuh*" less frequently

than Spanish learners did. However, they were able to employ it as a listener response token performing the convergence, continuer and information receipt functions at relevant places in conversations. Relative frequencies were as follows: Brazilian learner corpus (7.5620); Spanish component of LINDSEI (8.2358).

In O'Keeffe and Adolphs's corpus-based study (2008, p. 90), young British women data indicated that convergence tokens occurred twice as frequently as engagement tokens, ten times as frequently as continuer tokens and eleven times as frequently as information receipt tokens. Adding the occurrences of both "*yeah*" and "*uhuh*", the Brazilian data displayed even larger differences: convergence tokens occurred thirteen times as frequently as continuer tokens and twenty six times as frequently as information receipt tokens.

Discussion

Corsetti's (2015) empirical investigations were intended to illustrate some of the key phenomena comprising conversational competence and to highlight some areas which seem to have posed more difficulties to Brazilian learners at CEFR B1. Overall, subjects produced a limited range of discourse marking adverbs and listener response forms. Although they displayed a consistent use of "*really*", they underused the pragmatic and discourse functions of "*well*" and "*actually*".

Moreover, the overuse of "*maybe*" and the absence of "*may*" and "*might*" in the Brazilian data indicated that learners resorted to the adverb to convey degrees of certainty, possibly due to transfer. They also underused "*just*" as an implicit hedge. Within EFL contexts, the modal verbs "*must*", "*may*", "*might*", "*could*" and "*can't*" are usually grouped as "modal verbs for deductions" and introduced to learners in general as from CEFR A2. The pragmatic functions of "*just*", on the other hand, are usually presented to students at more advanced levels, usually at CEFR B2 or C1.

From a pedagogical perspective, Thornbury and Slade (2006, p. 296) suggest that an approach to the teaching of conversation should include three elements, namely exposure, instruction and practice, not necessarily in this order. The authors point out that the sequence of events is less important than the fact that all the three elements are included in an instructional programme.

According to Thornbury and Slade (2006), in an exposure-instruction- practice cycle, learners may listen to conversation extracts, study specific features highlighted in transcripts and then attempt to use them in their own conversations. In an instruction-exposure-practice cycle, learners are given prior explicit instruction in a feature of conversation. In an exposure-practice-instruction cycle,

students receive feedback (instruction) after they have listened to conversation extracts and tried to implement specific features. A practice-exposure-instruction cycle encompasses attemping a conversational task, observing more proficient speakers performing the same task and receiving instruction, before repeating the task.

Taking Corsetti's findings (2015) into account, one may infer that Brazilian learners at CEFR B1 in general would benefit from pedagogical assistance for the acquisition of discourse markers, for different forms to convey epistemic stance and for more varied tokens to express good listenership. Corsetti (2015) describes a few classroom activities which encompass the exposure, instruction and practice elements and may be implemented in the future to cater for these pedagogical objectives.

References

AIJMER. Karin. *English Discourse Particles*: Evidence from a Corpus. Amsterdam: John Benjamins Publishing Company, 2002.

AIJMER. Karin. Well I'm Not Sure I Think…: The Use of Well by Non-native Speakers. *International Journal of Corpus Linguistics*, v. 16, n. 2, p. 231-254, 2011.

ANTHONY, Laurence. *AntConc* (Version 3.4.2) [Computer Software]. Tokyo, Japan: Waseda University, 2014. Disponível em: <http://www.antlab.sci.waseda.ac.jp/>. Acesso em: 11 abr. 2014.

BACHMAN, Lyle. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 1990.

BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey; CONRAD, Susan; FINEGAN, Edward. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited, 1999.

BROWN, Gillian; YULE, George. *Discourse Analysis*. Cambridge: Cambridge University Press, 1983.

CANALE, Michael. On some Dimensions on Language Proficiency. In: OLLER, John W. (Ed.). *Issues in Language Testing Research*. Rowley: Newbury House, 1983. p. 333-342.

CANALE, Michael; SWAIN, Merrill. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, v. 1, n. 1, p. 1-47, 1980.

CARTER, Ronald; MCCARTHY, Michael. *The Cambridge Grammar of English*: A Comprehensive Guide to Grammar and Usage. Cambridge: Cambridge University Press, 2006.

Corpus Query Processor at Lancaster. Disponível em: <https://cqpweb.lancs.ac.uk/>. Acesso em: 6 abr. 2014.

CORSETTI, Cristiane. *Conversational competence in English as a second language*: a study of pragmatic markers. 2015. 328 fl. Tese (Doutorado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2015.

COUNCIL OF EUROPE. *Common European Framework of Reference for Languages*: learning, teaching, assessment. Cambridge: Cambridge University Press, 2001.

CULPEPER, Jonathan; KYTÖ, Merja. *Early Modern English Dialogues*: Spoken Interaction as Writing. Cambridge: Cambridge University Press, 2010.

JUCKER, Andreas. Corpus Pragmatics. In: OSTAM, Jan-Ola; VERSCHUEREN, Jef. *Handbook of Pragmatics*: 2013. Installment. John Benjamins Publishing Company, 2013. v. 17, p. 1-18.

LEVINSON, Stephen. *Pragmatics*. New York: Cambridge University Press, 1983.

NIKULA, Tarja. *Pragmatic Force Modifiers:* A Study in Interlanguage Pragmatics. Jyväskylä: University of Jyväskylä, v. 39, 1996.

O'KEEFFE, Anne; ADOLPHS, Svenja. Response tokens in British and Irish discourse: Corpus, context and variational pragmatics. In: SCHNEIDER, Klaus; BARRON, Anne. *Variational Pragmatics:* a focus on regional varieties in pluricentric languages. John Benjamins Publishing Company, 2008. p. 69-98.

O'KEEFFE, Anne; MCCARTHY, Michael; CARTER, Ronald. *From Corpus to Classroom*: language use and language teaching. Cambridge: Cambridge University Press, 2007.

PARADIS, Carita. Between epistemic modality and degree: The case of *really.* In: FACCHINETTI, Roberta; KRUG, Manfred; PALMER, Frank Robert. *Modality in Contemporary English*. Mouton de Gruyter, 2003. v. 44, p. 191-220.

O'KEEFFE, Anne; WALSH, Steve. Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. In: *Corpus Linguistics and Linguistic Theory*, v. 8, n. 1, p. 159-181, 2012.

The BNC Sampler, CML version. Distributed by Oxford University Computing Services on behalf of the BNC Consortium, 2005. Disponível em: <http://www.natcorp.ox.ac.uk/>. Acesso em: 6 abr. 2014.

The Diachronic Corpus of Present-Day Spoken English (DCPSE). Distributed by University College London. Disponível em: <http://www.ucl.ac.uk/english-usage/projects/dcpse/>. Acesso em: 2 maio 2014.

The Louvain International Database of Spoken English Interlanguage (LINDSEI). Distributed by Université Catholique de Louvain. Disponível em: <http://www.uclouvain.be/en-cecl-lindsei.html>. Acesso em: 15 maio 2014.

THOMAS, Jenny. *Meaning in Interaction*: An Introduction to Pragmatics. London: Longman, 1995.

THORNBURY, Scott; SLADE, Diana. *Conversation*: From Description to Pedagogy. Cambridge: Cambridge University Press, 2006.

UNIVERSITY OF CAMBRIDGE ESOL EXAMINATIONS. *First Certificate in English Handbook for Teachers*: For examinations from December 2008. Cambridge: UCLES, 2007.

Contato:
Cristina Lopes Perna <cperna@pucrs.br>
Cristiane Ruzicki Corsetti <cris.corsetti@gmail.com>