

Corpus-based Teaching of Pragmatics

Francisco Alonso Almeida, *PhD, Universidad de Las Palmas de Gran Canaria*, Elena Quintana Toledo, *PhD, Universidad de Las Palmas de Gran Canaria*

Abstract—This paper is a proposal for teaching pragmatics following a corpus-based approach. Corpora have had a high impact on how linguistics is looked at these days. However, teaching linguistics is still traditional in its scope and stays away from a growing tendency of incorporating authentic samples in the theoretical classroom, and so lecturers perpetuate the presentation of the same canonical examples students may find in their textbooks or in other introductory monographs. Our view is that using corpus linguistics, especially corpora freely available in the World Wide Web, will result in a more engaging and fresh look at the course of Pragmatics, while promoting early research in students. This way, they learn the concepts but most importantly how to later identify pragmatic phenomena in real text. Here, we raise our concern with the methodology, presenting clear examples of corpus-based pragmatic activities, and one clear result is the fact that students learn also how to be autonomous in their analysis of data. In our proposal, we move from more controlled tasks to autonomy. This proposal focuses on students enrolled in the course *Pragmática de la Lengua inglesa*, currently part of the curriculum in *Lenguas Modernas*, Universidad de Las Palmas de Gran Canaria.

Index Terms— Pragmatics, corpus linguistics, world wide web, inferences, presupposition, perspectivization, lesson plan, peer work, autonomous learning, corpus-based teaching.

I. INTRODUCTION

THE teaching of Pragmatics as a discipline in Spanish universities has now a well established tradition, at least in the degree of English studies, and other related ones. In the case of the Universidad de Las Palmas de Gran Canaria, this course, i.e. *Pragmática de la lengua inglesa* ‘English Pragmatics’, is part of the Linguistics module in the Degree of *Lenguas Modernas*. The course is designed for last year students (4th year). One may well think that this course is finely integrated into a coherent set of preparatory set of courses focusing on general linguistics, or at the very least, students have already taken courses on Semantics, Morphology, Syntax, etc. The situation could not be more unsympathetic, though, as only two courses, English Phonetics

Francisco Alonso Almeida, Universidad de Las Palmas de Gran Canaria, Edificio de Humanidades. Instituto para el Desarrollo Tecnológico y la Innovación en Comunicaciones, Despacho 23. C/ Pérez del Toro, 1. E35017. Las Palmas de Gran Canaria, Spain (e-mail: francisco.alonso@ulpgc.es).

Elena Quintana Toledo, Universidad de Las Palmas de Gran Canaria, Edificio de Ciencias de la Educación. Despacho 1. C/ Santa Juana de Arco, 1. E35017. Las Palmas de Gran Canaria, Spain (e-mail: elena.quintan@ulpgc.es).

and Phonology and History of the English Language, complete the pack of compulsory courses students take. This situation obviously calls for a very specific methodology and course planning in which the teacher should try and present material in the easiest and most basic way while s/he also keeps their students interest high. Our point of departure necessarily demands awareness of our students’ linguistics background, truly below the standards, and this complicates teaching and learning. So much so that the instructor has to identify whether lack of understanding of a concept during the presentation and activation phases is due to the own complexity of the concept being presented, or is due to the lack of background knowledge of the metalanguage deployed at this stage, no matter how basic this is. In this context, we think that may represent an engaging alternative, so that students can actively cooperate in building their own learning.

The use of corpora in language studies have been explained earlier in Chen [1], Costas [2], Hidalgo et al. [3], Aijmer [4], and Casas-Pedrosa et al. [5], among many others. Articles and books dealing with corpus-based teaching of pragmatics and pragmatic functions do not really abound, but there are some representative samples, namely O’Keeffe [5], K. Bardovi-Harlig, S. Mossman, and H. E. Vellenga [6][7]. We follow these previous works in our proposal of activities, which will be based on the *Michigan Corpus of Academic Spoken English* (MICASE) [8]. This corpus contains spoken data, which can be retrieved according to certain selected sociological criteria. These selection criteria will benefit our design of activities since students may evaluate the use of forms and their functions considering different variables that may have an effect on language.

The structure of our paper is, as follows. We describe approaches to the study of pragmatics from a corpus-based perspective. We then describe very briefly MICASE and its retrieval possibilities. After this we present a set of activities, all of them based on the use of MICASE to develop and practice new pragmatic knowledge. This is preceded by a short description of the syllabus designed for the course *Pragmática de la lengua inglesa* at the Universidad de Las Palmas de Gran Canaria. The last section contains some thoughts concerning the use of corpus linguistics in the linguistics classroom.

II. CORPUS LINGUISTS AND PRAGMATICS

It seems obvious that corpus linguistics has a lot to offer to the linguistics researcher because of the extent to which this

methodological approach has facilitated the empirical analysis, both quantitative and qualitative, of language variation and use. This relatively young methodology may be also applied to other academic contexts like language pedagogy, as evinced in the large number of publications generated after every biennial international conference on *Teaching and Language Corpora* has taken place [9]. Some of the most recent works include Campoy-Cubillo et al. [10], Aijmer [11], or Hidalgo, Quereda and Santana [12]. As noted by Leech [13], corpora may contribute to language teaching in three different ways, namely indirect, direct and teaching-oriented corpus development. Indirect uses include “reference publishing, materials development and language testing”. Direct uses encompass “teaching about, teaching to exploit and exploiting to teach”. Teaching-oriented corpus development has to do with “LSP corpora, L1 developmental corpora and L2 learner corpora” [14].

Works dealing with corpus-based approaches to teaching pragmatics specifically are scarce. One case in point is that by Bardovi-Harlig, Mossman and Vellenga (2015) [6], which offers a description about the use of a spoken corpus in order to develop materials to teach how to identify certain speech acts. One of the issues raised in this paper is the adequacy of deploying real language samples when teaching pragmatic competence. We share with the authors the view that an emphasis should be placed on language authenticity as only exposure to real language use can allow students to gain awareness of the appropriateness of some expressions in certain contexts.

The context of a communicative exchange both determines the pragmatic choice(s) made by the participants and shapes the interpretative process in fairly specific ways. Clancy and O’Keeffe [15] point out that “corpus linguistics has allowed for the comparison of this pragmatic choice at a number of levels” like language variety, medium and discourse domain. In the same line as Bardovi-Harlig, Mossman and Vellenga [6], these authors also highlight the usefulness of corpora in a language learning environment since they may provide language learners with an opportunity to learn from the pragmatic choices made by the producers of authentic texts and so develop their pragmatic competence.

In spite of the fact that publications on corpus-based approaches to teaching pragmatics do not abound, there has been a rapidly growing body of data-driven pragmatic research in general in the last decades. The pragmatic functions of modal verbs and adverbs are frequently addressed in corpus-based studies. Within the English Language Teaching (ELT) context, Römer [16] analyses the occurrence of modals in the spoken part of the British National Corpus (BNC) and compares the results with their occurrence in six ELT textbooks and a reference grammar. Her results show that their distribution in real language use differs significantly from that in textbooks, pointing to a need for improving teaching materials. Farr and O’Keeffe [17] concentrate on the occurrence of the hedging devices *I would say* and *I’d say* in three varieties of English, namely, British (the Cambridge and the Nottingham Corpus of Discourse in English), Irish (the

Limerick Corpus of Irish English) and American (the Cambridge International Corpus). These hedging devices are more frequently used in Irish English and their findings put forward, again, a need for considering language varieties when designing teaching materials.

The use of discourse markers has been also the focus of attention of several corpus-based studies. They have been predominantly looked at in oral rather than in written discourse, and very often in contrastive studies. One example is Fung and Carter [16] whose work compares the use of discourse markers by native speakers of English and learners using a corpus of spoken British English and a corpus of classroom discourse gathered from students in Hong Kong. Similarly to Römer [17] and Farr and O’Keeffe [18], they draw some pedagogical implications related to the improvement of the learner’s interactional competence. Though to a lesser extent, the occurrence and pragmatic functions of discourse markers have been investigated in written discourse in works like Bondi [19], where the author examines the discursive role of *however* in abstracts from academic journals belonging to different disciplines. She observes that *however* is indeed an interpersonal marker signalling that there is a common ground between writer and reader.

Another well-researched feature of pragmatics is deixis. Rühelman [20], for instance, looks into personal deixis in the spoken part of the BNC; particularly he searches for the personal pronouns *I* and *you* and compares their occurrences with written discourse. In the light of the author’s results, it seems that personal deixis is less important in the written mode. O’Keeffe [21][22] performs queries in a small corpus of media discourse in order to find out examples of othering and centring as manifested through the use of *they* and *we*, respectively.

III. MICASE

The Michigan Corpus of Academic Spoken English (MICASE) is a collection of transcribed speech recorded at lectures, discussions, seminars and other academic sessions taking place at the University of Michigan, and involving students, faculty members and other staff, both native and non-native. The corpus comprises texts dating from 1997 to 2001, and its size reaches 1.8 million words.

MICASE is freely available for teaching and research purposes at <http://quod.lib.umich.edu/m/micase/>. The website in which it is hosted offers users a search interface that allows them to “browse the corpus according to specified speaker and speech attributes, returning quick file references” (Figure 1). The speaker attributes that may be used to filter results include academic position/role, native speaker status and first language. As for the transcript attributes, these are speech event type, academic division, academic discipline, participant level and interactivity rating. The search interface also allows users to “search the corpus for words or phrases in specified contexts, returning concordance results with references to files, full utterances, and speakers” (Figure 2). The search page also shows transcript and speaker attributes to filter out

results. In this case, the transcript attributes remain the same, but the speaker attributes further involve gender and age.

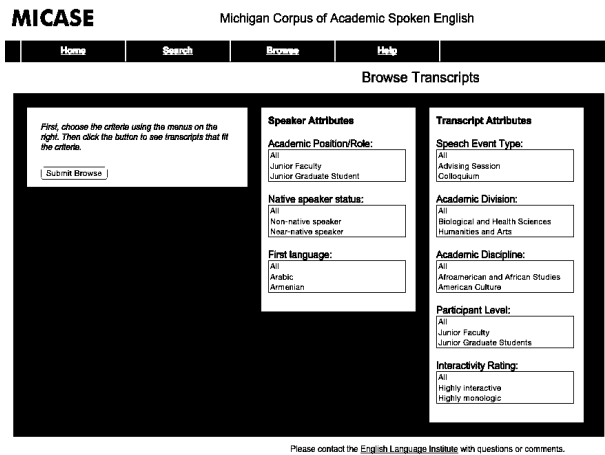


Figure 1. Browse corpus frontend.

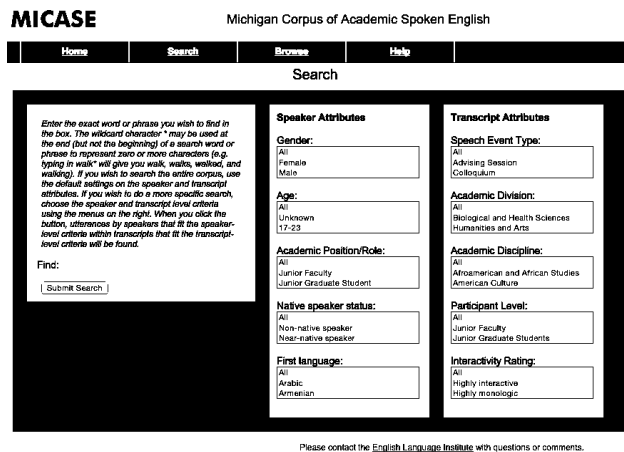


Figure 2. Search corpus frontend.

IV. CORPUS-BASED ACTIVITIES

A. Description of the syllabus

The course *Pragmática de la lengua inglesa* is developed in sixty contact hours, and this is completed with personal study to reflect on pragmatic phenomena and practice this new learning. Part of this autonomous learning is devoted to prepare oral presentations concerning one or several aspects dealt with in the course. While one of this is completely guided, as they only summarise one lesson in their textbooks, the second is autonomous in that they choose the topic and form of their presentation. The only requirement they need to comply with is the use of technological media to present their work: video, audio, laser beams, webcast... This is group work, and so they will monitor and evaluate their own material taking decisions concerning contents and form. This type of work allows assessing the students' competences acquired during the course, and the teacher is then able to provide feedback. These competences concern the following: (a) knowledge and ability to identify language functions, (b) the relationship of Pragmatics with other disciplines, (c) understanding of the process of linguistic communication, including the ability to identify the limitations of the

traditional model and the contributions of the new proposals (e.g. inferential model, ostensive-inferential model), and (d) knowledge of various principles governing the conversation and phenomena that arise in the process of language production and interpretation.

The course contents include a description of the origins of the discipline, the functions of language, the models of linguistic communication, deixis, meaning and meaning relationships, Gricean cooperative principle, relevance theory, politeness, speech act theory, and conversation analysis. In the following section, we will present a series of corpus-based activities, which might be used to present and practice pragmatics.

B. Activities

These activities are only examples of how the use of free corpora available via the Internet, as it is MICASE, helps both teaching and learning of pragmatic phenomena. Due to the nature of the corpus, multimodal analysis is not intended in these exercises.

• DEIXIS

a) Use MICASE browse facility and select transcript ID: LEL300SU076. Find out cases of symbolic and gestural deixis, and describe them.

b) The word *stuff* is frequently used in conversation in the place of more specific referents for different reasons. The following excerpt contains two instances of this device: “and for some magic reason elements like to have a complete shell of electrons, which means, if they have less than eight, they will attempt to gain electrons to get eight if they're a nonmetal. okay? now some of you i can see are smiling and saying well probably i know this **stuff** already but fine be patient, this is the difference in background here too, gotta know this **stuff**” (MICASE, transcript ID: LEL200JU105). Now, interrogate MICASE for similar authentic cases, and

b1) say whether you can retrieve the reference of *stuff* in all, in some, or in none of the cases, and

b2) find out whether there is variation concerning linguistic background (native vs. non-native speakers).

• POLITENESS

a) Modal verbs may have a downtoning effect of the propositional content in a given speech act, i.e. positive politeness. Other times modal verbs appear embedded in structures used to express negative politeness. Find out one example of each category in MICASE, and describe them.

b) Although speakers generally tend to convey politeness in discourse, it may happen that they use other devices, which are clearly impolite. In academic speech, impolite language might not be very easily identified, especially in the written medium. Explore lectures in MICASE for cases of impolite language. Give at least four examples covering male and female speech from two different linguistic backgrounds.

• THE COOPERATIVE PRINCIPLE

a) The Cooperative Principle indicates that four maxims should be observed in successful communication: the maxim of quantity, the maxim of quality, the maxim of relation, and

the maxim of manner. Using examples from MICASE, describe the way in which politeness and these maxims relate in speech.

- **SPEECH ACTS**

a) Focus on instances of *I will...* and *you will...* in MICASE, and decide the types of speech acts they exemplify. Justify your decision.

b) Analyse the use of *it's true + proposition* in varied speech events in MICASE.

c) Find out cases of directives in MICASE, and

c1) summarise main ways of expressing this type of speech act,

c2) select one example and illustrate Searle's felicity conditions [23].

- **CONVERSATION ANALYSIS**

a) Find out examples of *repair* in seminar and meeting events in MICASE.

b) Search MICASE for examples of *turn-taking strategies* in native and non-native speakers. Are there significant differences in how these devices are deployed by these two groups?

c) Analyse the function of *i mean* in interviews in MICASE.

d) Analyse the use of *we* in dissertation defenses in the Humanities and in the Biological and Health Sciences domains. Compare your findings and say whether there are major differences in use.

V. FINAL REMARKS

In this paper, we have exemplified the use of corpus linguistics for the Pragmatics classroom. We have seen that, while there has been an intention to provide teachers with corpus-based material for developing pragmatic competence in the English language classroom, this is not the case for the teaching of Pragmatics as a discipline. A set of activities has been given here to illustrate the way in which students may benefit from freely available corpora in the Internet to practise the new pragmatic knowledge they have acquired. Obviously, the activities must be graded according to the students' command of the discipline. They may start off by analysing simple collocations and, then, they can move towards more complex issues, such as functional and pragmatic interpretations of authentic data.

REFERENCES

- [1] Y. H. Chen. (2004). The use of corpora in the vocabulary classroom. *The Internet TESL Journal*, vol. X, no. 9. Available: <http://iteslj.org/Techniques/Chen-Corpora.html>.
- [2] G. Costas. (2005) Corpora and language teaching: Just a fling or wedding bells? *Teaching English as a Second or Foreign Language*, vol. 8. no. 4. Available: <http://tesl-ej.org/ej32/a1.html>.
- [3] E. Hidalgo, L. Quereda and J. Santana (eds.). 2007. *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi.
- [4] A. V. Casas-Pedrosa, J. Fernández-Domínguez and A. Alcaraz-Sintes. (2013). Introduction: the use of corpora for language teaching and learning. *Research in Corpus Linguistics*, vol. 1, 1–5.
- [5] A. O'Keeffe, B. Clancy, S. Adolphs. (2011). *Introducing Pragmatics in Use*. Oxford: Routledge.
- [6] K. Bardovi-Harlig, S. Mossman, and H. E. Vellenga. (2015). Developing corpus-based materials to teach pragmatic routines. *TESOL Journal*, vol. 6,

no. 3, 499-526.

[7] K. Bardovi-Harlig, S. Mossman, and H. E. Vellenga. (2015). The effect of instruction on pragmatic routines in academic discussion. *Language Teaching Research*, vol. 19, 324-350.

[8] R. C. Simpson, S. L. Briggs, J. Ovens, and J. M. Swales. (2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. Available: quod.lib.umich.edu/m/micase/

[9] McEnery, T. and R. Xiao. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (ed.). *Handbook of research in second language teaching and learning*. Vol. 2. New York: Routledge.

[10] Campoy-Cubillo, M.C. et al. (eds.). (2010). *Corpus-based approaches to ELT*. London/New York: Continuum.

[11] K. Aijmer. (2009) *Corpora and Language Teaching*. Amsterdam: John Benjamins.

[12] Hidalgo, E., L. Quereda and J. Santana (eds.). (2007). *Corpora in the foreign language classroom. Selected papers from the Sixth International Conference on Teaching and Language Corpora (TaLC6)*. Amsterdam: Rodopi.

[13] Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds.). *Teaching and language corpora*. New York: Addison Wesley Longman.

[14] McEnery, T., R. Xiao and Y. Tono. (2006). *Corpus-based language studies. An advanced resourcebook*. London/New York: Routledge, p. 97.

[15] Clancy, B. and A. O'Keeffe. (2013). Using a corpus to enhance pragmatic awareness. In M. D. Garcia-Pastor (ed.) *Teaching English as a foreign language: Proposals for the language classroom*. Periferic Edicions.

[16] Fung, L. and R. Carter. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics* 28.3: 410-439.

[17] Römer, U. (2004). A corpus-driven approach to modal auxiliaries and their didactics. In J. Sinclair (ed.). *How to use corpora in language teaching*. Amsterdam: John Benjamins, 185-199.

[18] Farr, F. and A. O'Keeffe. (2002). *Would* as a hedging device in an Irish context: An intra-varietal comparison of institutionalised spoken interaction. In R. Reppen, S. Fitzmaurice and D. Biber (eds.). *Using Corpora to Explore Linguistic Variation*. Amsterdam: John Benjamins, 25-48.

[19] Bondi, M. (2004). The discourse function of contrastive connectors in academic abstracts. In K. Aijmer (ed.). *Discourse patterns in spoken and written corpora*. John Benjamins. 139-156.

[20] Rühlemann, C. (2007). *Conversation in context: A corpus-driven approach*. London: Continuum.

[21] A. O'Keeffe. (2002) Exploring indices of national identity in a corpus of radio phone-in data from Irish radio. In A. Sánchez Macarro (ed.), *Windows on the World: Media Discourse in English*. Valencia: Universidad de Valencia.

[22] A. O'Keeffe. (2006) *Investigating Media Discourse*. London: Routledge.

[23] Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, 1969.

Francisco Alonso Almeida Francisco Alonso-Almeida is Senior Lecturer of English Philology at the University of Las Palmas de Gran Canaria. His research interests focus on (historical) pragmatics, (historical) discourse analysis, specialised English discourse and corpus linguistics. He has published empirical articles and chapters on corpus linguistics, and the evolution of text patterns and functions. He has published in international refereed journals, such as *Neophilologus*, *Neophilologische Mitteilungen*, *Studia Neophilologica*, *Folia Linguistica Historica*, *Discourse Studies*, *Intercultural Pragmatics*, and *Journal of Historical Pragmatics*, among others. He serves as referee for international and national academic journals, and he is the Editor-in-Chief of *Revista de lenguas para fines específicos*.

Elena Quintana Toledo is part-time lecturer of English at the Department of Special Didactics at the University of Las Palmas de Gran Canaria. Her research interests include

English scientific discourse, (historical) pragmatics, (historical) discourse analysis and corpus linguistics. She has been recently engaged in the analysis of contemporary academic research papers from a pragmatically oriented perspective by focusing on their interpersonal features.

