LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 17, No: 1, January – June 2024

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Corpus-based Teaching of English Conversation and Potential Integration of Conversation Analysis (CA) for the Benefit of EFL Teachers and Learners

Patson Jaihow[a], Kemtong Sinwongsuwat[b,*]

[a] patson.j@psu.ac.th, Department of Foreign Languages, Faculty of Liberal Arts, Prince of Songkla University, Thailand
[b] ksinwong@gmail.com, Research Center for Language, Culture, and Human Development in Lower ASEAN (RC-LCHD), Faculty of Liberal Arts, Prince of Songkla University, Thailand
[*] Corresponding author, ksinwong@gmail.com

APA Citation:
Jaihow, P., & Sinwongsuwat, K. (2024). Corpus-based teaching of English conversation and potential integration of Conversation Analysis (CA) for the benefit of EFL teachers and learners. *LEARN Journal: Language Education and Acquisition Research Network, 17*(1), 264-285.

Received
16/07/2023

Received in revised form 12/09/2023

Accepted
24/09/2023

ABSTRACT

Even though corpora have transformed language education, the majority of corpus-related research focuses on the teaching of writing. Via a systematic review of previous studies and a survey of language corpora available, this study aims to ascertain whether and to what extent the teaching of speaking, especially of conversation skills, to EFL learners, has been informed by existing language corpora, to identify spoken English corpora available and discover whether they have been informed by insights from such an approach to studying conversation as Conversation Analysis (CA). Finally, the study suggests possibilities for incorporating CA insights such that CA-informed, corpus-driven language pedagogy can be materialized. Previous studies on the use of corpora for teaching speaking were examined and spoken language corpora available were identified along with how they have been recommended and applied to the teaching of speaking,

as well as possibilities for developing CA-informed corpora of spoken English for conversation teaching. The study revealed that conversation teaching in the EFL context remains to be informed by corpus linguistics. Accessible spoken English corpora are not yet geared towards language teachers and learners, and there remain issues to be resolved before employing the available corpus data and confirming its efficacy in teaching EFL conversation or speaking in general.

Keywords: conversation analysis, corpus-driven language pedagogy, corpus linguistics, teaching EFL speaking, teaching English conversation

Introduction

Communication skills in languages such as English are undoubtedly one of the essential attributes of global citizens in the 21st century. With the growing worldwide demand for English in international interactions across different arenas in the digital age, it is difficult to imagine how EFL learners could advance in their professional growth without effective English communication skills (Stauffer, 2022). Therefore, even with available technology for communication aids, teaching of L2 skills such as English conversation or speaking skills in general is becoming even more crucial for real-time communication, particularly interaction with speakers of other languages, especially those with little literacy.

While communicative approaches such as Communicative Language Teaching (CLT), which promote students' engagement in authentic conversations and materials, have always been touted in EFL settings such as Thailand, English conversation skills are often taught using prescribed commercial textbooks with invented written conversations as models and scripted talk as the target output (Pitaksuksan & Sinwongsuwat, 2020), leading to inadequacy of its value of serving learners' personal, cultural, and social needs (Nunan & Richards, 2015). EFL teachers reported several major challenges when teaching speaking, including their own and students' limited outside-of-class exposure to English, which led to a lack of teaching and speaking confidence, an inability to effectively teach conversation and properly assess students' performance, and a lack of understanding of the genuine nature of what they teach to be able to provide adequate explanations due to the complexity of speaking (Burns, 2017). Some EFL instructors even struggle to find the right training and materials designed especially for teaching conversation (Wessels et al., 2017). Without adequate professional

development and access to appropriate, well-structured materials that promote conversational skills, teachers may struggle to design effective conversation lesson plans and activities.

Learners, on the other hand, are also faced with difficulties when trying to develop their conversation skills. Most are reportedly reluctant to speak English due to limited exposure to authentic face-to-face communication in English, a high level of anxiety caused by evaluation, teaching and learning methods, and the face-saving, hierarchical culture which makes them reluctant to talk in public or in front of authority figures like their teachers (Roberts & Cooke, 2009; Savasci, 2013). Their lack of motivation and limited knowledge of vocabulary are also an issue (Ansari, 2015; Paneerselvam & Mohamad, 2019).

To help learners in particular to overcome these challenges, a variety of teaching methods were recommended in the literature. For example, Handayani (2016) and Mansor and Rahim (2017) suggested using Instagram to enhance learners' speaking skills in the classroom by asking them to talk about familiar topics such as their trips and celebrities. The flipped classroom was also introduced as a method to stimulate active discussions among learners (Handayani, 2016). Othman (2014) made use of debates in developing learners' critical thinking skills via speaking. Paneerselvam and Mohamad (2019) recommended the use of games to motivate learners to participate actively in speaking activities. Storytelling has also been reportedly used as a means of enhancing EFL learners' communication skills (Tambunan et al., 2018).

Amidst such an array of methods introduced to make teaching speaking more effective, corpus-based teaching has also been extensively discussed in previous literature. With its revolutionary impact on classroom practice (Flowerdew, 2009; Hunston, 2002; O'Keeffe et al., 2007), corpora have long played an important role in language education since they may be used to teach both written and spoken language. However, to date, most of the research on the use of corpora for language instruction has focused mainly on the application of corpora of written data and corpus techniques for teaching writing, and very little has been done to maximize its benefits for teaching speaking (Akkoyunlu & Kilimci, 2017; Cobb & Boulton, 2015; De Cock, 2010).

To address previously discussed challenges in teaching speaking, especially regarding limited exposure to authentic L2 input in and outside the classroom, the use of language corpora has been recommended when developing materials for teaching speaking. Burns (2017) recommended that corpora of spoken language be developed that can help learners deepen their understanding of how English is used to meet their real needs in both local and global contexts. Through the lens of Conversation Analysis (CA),

Lazaraton (2014) in particular observed that EFL speaking materials were largely based on invented, scripted dialogues free of natural features of spoken language and unidentifiable with any real speakers of English (see also McCarthy & O'Keeffe, 2004). Without authentic language data and engaging, meaningful tasks appropriate to learners in the classroom, it is difficult to imagine how students will be able to notice naturally occurring features of language use and successfully apply what they learn in real-life situations. Roberts and Cooke (2009) reported that the use of commercially created materials to teach English in the workplace failed to prepare adult learners for real-life workplace communication.

Corpora of spoken English that offer input from various natural or near-natural speech settings are therefore essential for teachers to create materials to develop their learners' speaking skills appropriate for real-life interaction. Such input will provide learners a greater chance to be exposed to natural features of the target language and make them aware of these features. Markee (2005) and Seedhouse (2005) concurred that explicit knowledge of interactional practices in constructing conversation as revealed via Conversation Analysis (CA) is crucial for teaching and conversing in a second language, and their ideas of CA-informed language pedagogy has been resonated in works by several scholars in numerous contexts (see, e.g., Barraja-Rohan, 2011; Fujii, 2012; Sinwongsuwat et al., 2018; Wong & Waring, 2020). Such knowledge has been proven to help learners navigate through L2 conversations with more confidence and equips them with the ability to deal with common communicative problems more effectively. CA transcription of genuine talk-in-interactions can help learners understand spoken language features and interactional practices such as turn-taking, sequencing, and repair. If used properly, CA-informed transcription also has the potential to enhance transcribed features of existing spoken language corpora for pedagogical purposes. Despite the scarcity of research that delves into the new perspective on integrating conversation analysis into corpus linguistics, O'Keeffe and Walsh (2012) argued that adopting such an integrated approach in a speaking classroom can provide better insights into the connections between interaction patterns, language use, and learning than using only either one of them.

While it is expected that utilizing the wealth of input from empirical corpus data will make the creation of today's ELT materials even more effective and while there are several other recommendations for teachers to incorporate insights into natural talk-in-interaction from CA into speaking pedagogy (Burns, 2017; Lazaraton, 2014; McCarthy & O'Keeffe, 2004; O'Keeffe & Walsh, 2012), it is still unclear how and to what extent these recommendations have been implemented over the years when it comes to teaching speaking. Therefore, before conducting any further research into the

efficacy of CA-informed corpus-driven pedagogy in improving EFL learners' speaking performance, this paper aims to explore the existing body of research investigating corpus-data-driven pedagogy in teaching speaking and explore the possibilities of incorporating CA insights into it.

Objectives (OBJ)

Accordingly, the purposes of this manuscript, driven by documentary research, are threefold.

1. To determine whether the teaching of conversation to EFL learners has been informed by corpus linguistics. If so, how and to what extent?
2. To identify spoken language corpora available for teaching speaking, describe their characteristics, and determine whether they have been informed by CA.
3. To suggest possibilities for CA integration into corpus development to enhance the efficacy of corpus-based teaching in improving EFL learners' conversation skills.

Methodology

Since this is a primarily documentary research study, a systematic approach to reviewing previous literature was adopted generally following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines explained in Page et al. (2021) and the weaker version of systematic reviews as defined in Sataloff, R. T. et al. (2021).

To fulfill OBJ1, in adherence to PRISMA items 6-9, concerning eligibility criteria, information sources, search strategy, selection, and data collection process, the data examined in this study were obtained from the authors' internet keyword search, including publications from 2013 onwards in the top three academic research databases most institutionally recognized in humanities and social sciences, namely Scopus, Web of Science, and ERIC. The keywords used in the search were *corpus linguistics, spoken English corpora, teaching speaking,* and *EFL*. The focus was on research/review articles, conference proceeding papers, and book chapters containing these keywords. Since the search of the ERIC database bore no target publications, the investigation was scoped down to two databases: Web of Science and Scopus. The Web of Science search in April 2023 rendered 17 publication titles. In Scopus, the search produced only two titles. Given the limited number of titles retrieved from these databases, attempts were also made to conduct a free search on Google Scholar for additional titles related to the teaching of

conversation or speaking to EFL learners that have been informed by corpus linguistics, and only one matching title was found, namely, Şahin Kızıl and Savran (2018), as shown in Appendix A.

Following PRISMA items about the inclusion and exclusion criteria of studies for the review, all the retrieved titles were skimmed to determine whether they report the use of corpora in teaching speaking. Those papers not doing so were discarded while those relevant were thoroughly examined to discover how a corpus or corpus data was used in the process of teaching speaking. At this stage, a preliminary analysis was performed by the first author to decide whether or not the papers met the inclusion criteria. Then, a summary of each study was prepared, followed by a discussion with the second author to confirm the data and the process of screening the papers following PRISMA item 10a. (See Appendix A for a list of publications obtained from the search). Finally, the summaries of the studies were synthesized following PRISMA item 13 to address the question of whether the teaching of speaking, especially conversation, to EFL learners has been informed by corpus linguistics.

To address the second research objective, the spoken language corpora available for EFL teaching were identified and how they have been recommended and applied to the teaching of speaking were then determined. Listing spoken corpora of all languages, the Common Language Resources and Technology Infrastructure (CLARIN) database, available at https://www.clarin.eu/resource-families/spoken-corpora (accessed in June 2023), was screened for the corpora of spoken English only along with the list of well-known and influential corpora prepared by Xiao (2008). Around 35 corpora of spoken English were found in these two sources. Based on the descriptions given in both of these sources, each corpus was then examined concerning the language represented, size, and tags or annotation. Of these 35 corpora, only nine were made available to the public on the web or via download; the rest were not accessible (See Table 1). Additionally, only one of the nine, i.e., Griffith Corpus of Spoken Australian English (GCSAusE), adopts the annotation system with a CA transcription convention.

The findings from OBJ1 and OBJ2 will reveal possibilities for the construction and integration of CA-informed corpora of spoken English into conversation teaching to enhance its effectiveness in developing EFL learners' conversation skills, which is OBJ3 of this paper.

Findings and discussion

Has the Teaching of Conversation to EFL Learners Been Informed by Corpus Linguistics?

Despite the availability of both written and spoken English represented in a corpus, research studies involving the use of corpora in teaching the written language have outnumbered those applying corpus data to the teaching of speaking despite the latter being the basic skill of learning a language. Out of 19 publication titles retrieved from 2013-2023, four papers related mainly to the use of corpora in writing. One title by Yin (2015) shows how data from spoken language corpora can inform the use of cohesive devices in a news writing class. Three others that appeared in the search results, i.e., Babanoglu (2014), Gu and Xu (2021), and Kayumova et al., (2017), were oriented towards understanding English as a second or foreign language writing by examining learner written corpora or comparing EFL-L1 written corpus data rather than using corpora for the benefit of teaching speaking.

The other papers were found more relevant to the use of corpora in teaching speaking. Studies such as Khojasteh and Shokrpour (2014) highlighted the pedagogical benefits of language corpora, revealing the mismatch between language use in corpora and textbooks and recommending the development of corpora-informed materials for language teaching. Geluso and Yamaguchi (2014) used the Corpus of Contemporary American English (COCA) in an EFL speaking class to enhance Japanese learners' speaking fluency by having them search for interesting words or phrases in reading or audiovisual materials and uncover their various patterns of use in the corpus that would aid them in subsequent speaking-related tasks. Similarly, Alfehaid (2018) discovered a positive impact of the use of online resources such as COCA and the British National Corpus (BNC) on the speaking fluency of EFL learners in Saudi Arabia. A corpus has additionally been used as a guide for Russian EFL learners to discover pronunciation patterns of connected speech (Khoroshilova et al., 2015).

Other studies began to evaluate different corpora in terms of their benefits for L2 teaching and learning purposes. Dang et al., (2022) investigated teachers' and learners' perceptions of the usefulness of wordlists and the knowledge gained from mainstream corpora COCA2000/the British National Corpus (BNC) compared to other corpora, suggesting the greater coverage and merits of these corpora for learning English lexical items. Attempts were also made to validate the use of a corpus of spoken English in the media such as Internet TV against BNC for teaching speaking. Via a comparison of the distribution of formulaic sequences (FSs) in a corpus of internet television (the iTV corpus) with the everyday spoken English component of the BNC, Lin (2014) confirmed the validity of internet television as a resource for EFL learners' FS acquisition, as the frequency of FSs appears directly proportional to everyday speech. Chen et al., (2014) confirmed the benefit of using available technology such as GRASP, an

automatic reference aid for formulaic expressions, to improve EFL learners' fluency in productive skills.

The remaining three titles were related to building L1/L2 corpora. Two, i.e., Kwon (2022) and Genç-Yöntem and Eveyik-Aydın (2022), tried to build corpora of EFL learners' spoken English while the other attempted to build a corpus of spoken L1 other than English for pedagogical and linguistic purposes, i.e., Garrote et al., (2015).

Five papers, i.e., Roca-Varela (2013) and Sung and Kim (2016), used a corpus-based method to analyze linguistic items used in EFL learners' spoken English such as high-frequency words and phrasal verbs to identify problems that need to be addressed in teaching. Neves (2020) used corpus analysis to understand learners' construction of L2 discourse. The other studies also used the corpus analysis method and offered findings that can inform teaching speaking. Namely, Huang and Graf (2020) compared the speech rate and pausing between L2 and L1 English speakers of the same proficiency levels to understand fluency phenomena. Finally, Wang and Zou (2019) used the method to conduct a comparative analysis of the prosodic use to understand the prosody of *feel* in Chinese EFL learners' public speaking compared with native speakers.

Like Gabrielatos (2005), to motivate learners and enable them to develop skills in recognizing language patterns from corpus data, it is essential to make a pedagogical shift from focusing mainly on rote learning of fixed rules and exceptions or only on one correct answer based on examples taken out of context to recognizing patterns and alternatives of actual language use in genuine contexts. Based on the additional Google Scholar search, Şahin Kizil and Savran (2018) agreed that this is also true for speaking where learners' motivation to improve their speaking skills can increase when they notice features of spoken language such as hesitations, pauses, and narrow range of vocabulary used in naturally occurring conversation of native speakers. The study shows that corpus-informed approaches significantly impact lower-level students' language learning, contrasting the belief that mostly advanced English learners benefit from corpus-based instructional sources (Granath, 2009; Hunston, 2002). Both upper and lower-level students have a positive attitude towards corpus-based learning since it can motivate their confident participation in speaking activities. This is similar to Boulton (2009), who contends that even lower proficiency level learners could benefit from data-driven learning.

Several benefits of corpus-informed materials such as commercial textbooks and dictionaries have been pointed out in the literature. McCarthy (2004), in particular, promoted the genuine value of these materials as they rely on actual usage of language, frequency information, and authentic contexts, and can highlight the differences between spoken and written

language. Examples used in corpus-informed materials are adapted from real usage, and a corpus-informed syllabus prioritizes grammar and vocabulary that is most useful for learners, incorporating language previously overlooked in teaching communication. Via corpus data, subtle meanings behind words and expressions can also be revealed. Specialized corpora can additionally be analyzed to cater to specific learners' needs, such as academic or business contexts, to provide tailored materials for learners studying abroad or working in a second language, for example. Through these materials including coursebooks and dictionaries, teachers can predict common errors and provide their learners with authentic language experience without requiring target language immersion.

Learners should also become more motivated to learn as these materials provide them with modern, everyday language that aligns with real conversations, movies, and media. This non-artificial language, which includes widely used words, phrases, and grammar, can ensure faster and more efficient learning experiences. Such a corpus-informed pedagogical approach also helps learners understand essential language for basic communication and improve their listening and speaking skills while promoting social communication and enhancing their communication effectiveness (McCarthy, 2004).

Decades later, while maintaining that real-world social interaction cannot be represented in exemplary single sentences, McCarthy and McCarten (2022) still stress the need for sample utterances in a wider context and the necessity of using corpus data in teaching conversation. According to previous literature examined, although spoken corpora have been strongly advocated by scholars for EFL teaching (see Caines et al., 2016), there has hardly been any thorough examination into how they might be used in the classroom to teach everyday conversation skills and whether they are helpful in doing so.

To What Extent Have Spoken Corpora Available for Teaching Speaking Been Informed by CA?

Regarding OBJ 2, based on the survey of spoken language corpora available, it was found that the majority are not readily accessible online despite attempts with either the corpus title search adopted in this study or recommended links provided in the literature (see Xiao, 2008). Shown in the table below are nine out of over 35 corpora with transcribed spoken texts with or without audio-video recordings which are accessible for download or for display via web search engines.

Table 1

Corpora with Transcribed Spoken Texts Accessible for Download or Display via the Web Search Engines

Corpus	Description	Transcription and Annotation
1. The British National Corpus (BNC)	represents British English of the late 20th century (from 1960-1993); contains face-to-face conversations between L1 speakers of British English; covers informal everyday conversation	speech orthographically transcribed not following any particular conversation analysis conventions; transcripts annotated with word-class information and lemmatized; texts detailed with metatextual information; concordance software allowing both keyword-in-context (KWIC) and sentence views (See examples in Appendix B.); larger, extended contexts of the keyword provided
2. Spoken BNC2014	contains contemporary British English conversations between friends and family members recorded in casual settings (mostly at home); the speakers self-recorded these conversations using the built-in audio recording feature on their smartphones.	
3. The Longman British Spoken Corpus (LBSC)	contains natural conversations from a representative sample, including lectures, business meetings, and chat shows, being the first systematic large-scale collection of spoken data, part of the British National Corpus.	
4. Buckeye Corpus of Conversational Speech (BCCS)	contains high-quality recordings from 40 speakers of American English in Columbus, Ohio conversing freely in sociolinguistic interviews	orthographically transcribed and phonetically labeled especially for research and acoustic training related to variation in English pronunciation; concordance software not accessible online
5. Griffith Corpus of Spoken Australian English (GCSAusE)	a collection of 40 audio recordings and transcriptions of spoken interaction amongst Australian speakers of English, and users of English in Australia more generally	recordings transcribed using Conversation Analysis transcription conventions along with orthographical annotation; no concordance software provided
6. The Michigan Corpus of Academic Spoken English (MICASE)	covers contemporary university speech at the University of Michigan (large and small lectures, dissertation defences, presentations and discussions, lab sections, seminars, advising	orthographically transcribed using conventions and mark-up system allowing for ease of readability; standard punctuation not used; pauses of varying lengths marked with commas,

Corpus	Description	Transcription and Annotation
	sessions, study groups and meetings) from 1997-2001, with 37,000 students and diverse speakers, including faculty, staff, and non-native speakers	periods, and ellipses; question marks used to indicate utterances serving as questions; backchannel cues and hesitation transcribed using normalized orthographic representations, disregarding minor phonetic variations; speakers' overlaps, and interruptions transcribed in the sequences where they occur; concordance software allowing only keyword-in-context (KWIC) views; larger, extended contexts of the keyword provided
7. The Corpus of Contemporary American English (COCA)	includes spoken and written language, and multimedia content, such as fiction, magazines, newspapers, academic texts, movie and TV subtitles, web pages, and blogs	speech only orthographically transcribed, including unscripted conversation from TV and radio programs and movies; concordance software allowing only keyword-in-context (KWIC); larger, expanded contexts of the keyword provided as in connected text without any line identifications, making it relatively difficult to recognize speakers' turns
8. The Santa Barbara Corpus of Spoken American English (SBCSAE)	contains spontaneous speech recordings from various regions, ages, occupations, and backgrounds, reflecting various language usage in various settings, including conversation, gossip, arguments, and speeches	speech transcribed following conversation analysis transcription conventions; transcripts time-stamped and altered to preserve speaker anonymity; representation using a wide range of symbols suitable more for research; no concordance software provided
9. The Hong Kong Corpus of Spoken English (HKCSE)	recorded in the mid-1990s, representing Hong Kong's main spoken English discourses; accompanied with an annotated speech act corpus	transcribed recordings with prosody being annotated orthographically, being the largest English corpus with prosodic details; transcripts not following any particular conversation analysis transcription conventions; spoken features such as laughter described in words; concordance software allowing only KWIC views; extended contexts of the keyword provided

Out of the nine corpora examined, six contained face-face conversations among speakers of different varieties of English across various settings, i.e., the British National Corpus (BNC), the Spoken BNC 2014, The Longman British Spoken Corpus, the Griffith Corpus of Spoken Australian English (GCSAusE), the Santa Barbara Corpus of Spoken American English (SBCSAE), and The Hong Kong Corpus o Spoken English (HKCSE). One contained a monologue talk taken from sociolinguistic interviews, i.e., Buckeye Corpus of Conversational Speech, while The Michigan Corpus of Academic Spoken English (MICASE) is well-known for its large collection of academic spoken English from various settings. The Corpus of Contemporary American English (COCA) also includes spoken texts from TV and movies.

Among those corpora where conversation samples might be used for teaching EFL learners contemporary, everyday English conversation, namely, Spoken BNC2014, LBSC, BCCS, GCSAusE, COCA, and SBCSAE, half are orthographically transcribed and annotated, while the other half offer transcripts following a particular conversation analysis convention. Transcribed spoken texts in these corpora transcribed are more oriented towards linguists, language researchers and scholars rather than teachers, practitioners, or EFL learners. Although adopting conversation analysis transcription conventions, GCSAusE and SBCSAE, which represent conversations in an easily discernible sequential order of turns together with audio or video recordings, do not provide any concordance software to allow users to search for words or expressions in contexts with different concordance view results. GCSAusE also did not appear to have made transcription symbols readily available. Despite concordance software provided, those corpora containing spoken English with mainly orthographical transcriptions such as HKCSE are relatively difficult to process since they did not represent talks turn by turn as they occurred. This could perplex non-specialists, especially EFL teachers and learners, most of whom are familiar with model conversations represented with speakers' turns.

Consequently, despite the availability of these spoken corpora online and their recommendations for using them to replace made-up conversations often used in textbooks, using these corpora for conversation teaching and learning can still be challenging at this point. While encouraging the use of spoken corpora such as the spoken BNC, McCarthy and McCarten (2022) admit that understanding the transcripts of real conversations appearing in these corpora can be difficult as the annotations added to represent its features can be distracting and challenging to understand. To be used for language teaching and learning, it is essential to make them more easily accessible and user-friendly. As remarked by Şahin Kızıl and Savran (2018),

while showing positive attitudes towards learning linguistic expressions from corpus data, their learners were indecisive when asked about the user-friendliness of the corpus used.

In line with McCarthy and McCarten (2022), language learners, especially those in EFL contexts, need even more exposure to authentic talk and it is necessary to reinforce their experience by getting them to pay attention to what they have listened to or watched by getting them to notice what talk is delivered and more importantly how it is delivered. The symbols used in the representation therefore need to be very selective and it is necessary to represent only the main features of natural talk that may be challenging for learners to master. For example, Thai EFL learners whose first language is syllable-timed may benefit more from corpora that depict important prosodic aspects like stress and intonation to master these features barely addressed in conversation lessons more quickly.

Additionally, similar to McCarthy and McCarten (2022), it likely takes a lot of searching for teachers to get the right materials containing target items out of raw corpus data of real conversation available. Therefore, to get the most out of corpus data technology today for learning everyday English conversation, language corpora still need to be developed, especially for learners of different language proficiencies. The transcription and annotation conventions used also need to be simplified, highlighting mainly those essential for mastering conversation in the target language in various settings.

Possibilities for Integrating CA into Corpus Development for Corpus-based Conversation Teaching

Based on the findings discussed so far and to address OBJ3, there are possibilities for integration of insights, especially from conversation analysis, into the creation of spoken language corpora for corpus-based conversation teaching to enhance its efficacy in improving EFL learners' conversation skills. Given the established CA view on language as action performed by talk participants as they engage in everyday social activities, CA-informed pedagogical corpora can be created that are oriented towards these activities. These pedagogical corpora do not need to include extraneous annotating elements that can be distracting for learners unfamiliar with nuances of the spoken target language, but only those essential for constructing and allocating turns in real-time talk such as pitch (i.e., stress and intonation), pauses, and lengthening. Particularly, the transcripts provided in these corpora may just be annotated only with symbols necessary for noticing those spoken language features which are marked, infrequent forms in their L1, and concordance software should be provided to offer sentence or turn views and

views of key words or expressions searched in extended contexts in which they emerge.

In line with Wong and Waring (2010), Barraja-Rohan, (2011), and Sinwongsuwat et al., (2018), to advance in teaching and learning L2 conversational skills, both teachers and learners need to be cognizant of interactional practices involved, and conversational corpora can help with this. Embracing the pedagogical merits of CA insights into these practices, Sinwongsuwat and Nicoletti (2020) attempted to introduce a teaching model for English conversation teachers to develop their English conversation lessons with explicit knowledge of interactional mechanisms. According to the model, learners need to be exposed to recordings of naturally occurring or near-natural talk by English speakers, while teachers are also recommended to use a corpus of conversation data developed for creating their lessons and also for investigating and diagnosing their learners' interactional challenges. The study argued for English conversation corpora that can meet EFL learners' and teachers' needs.

As previously discussed, despite several spoken English corpora available today, most are of enormous size and contain the language that may not fit the proficiency level of most EFL learners. There is a need, therefore, to provide them with smaller, more manageable corpora with the language suitable to their proficiency levels (Boulton, 2017). Additionally, to date, there have hardly been any CA-informed English conversation corpora that are action-driven and made suitable for primary and secondary school teachers to design lessons for their target learners.

Further research therefore should aim at engaging English language educators and specialists in corpus linguistics and CA in co-creating English conversation corpora with concordance software for EFL teachers and learners. These corpora need also to be tested among the target users, and studies should be conducted to investigate their utilization of these corpora and its effectiveness in enhancing learners' communication skills.

Conclusion

Via a systematic review of previous studies and an investigation of accessible corpora, this study aimed to determine whether EFL conversation teaching has been informed by corpus linguistics, identify available spoken language accessible to EFL teachers and learners and suitable for teaching speaking, and explore potential integration of conversation analysis into corpus-based conversation teaching to improve EFL learners' conversation skills.

Based on the review, conversation teaching in the EFL context has not optimally been informed by corpus linguistics. Only a few studies reported

the use of spoken corpora for teaching some features of spoken language. Existing spoken English corpora have been designed mainly to serve the research community and language publishers and are hardly accessible to teachers and learners. Those accessible for the examination were not presented in ways that support classroom use. Most of these corpora represented the talk that learners or even teachers in the EFL context, especially young learners, are less likely to encounter in everyday life, countering what is normally prescribed in an English conversation syllabus.

Therefore, challenges faced by teachers wishing to use corpus data of spoken English remain to be overcome, opening the gap to be filled by integration of insights from CA into corpus-based conversation teaching. If spoken corpora are to be used for enhancing EFL learners' conversation skills, corpus creators and providers need to provide a corpus suitable for the learners in terms of language proficiency, size, and ease of technical use. Features of talk necessary for learners to learn need to also be represented in ways that will not confuse teachers and learners while making them aware of both linguistic and interactional resources essential for constructing and allocating turns at talk. Teachers also need training in how to use language corpora for pedagogical purposes. Further research then can attempt to verify the effectiveness of utilizing language corpora in EFL teaching of speaking skills such as everyday conversation.

Acknowledgements

The authors thank Prince of Songkla University, Hat Yai, Thailand, for the research grant #LIA6505185S under the 2022 Fundamental Fund of National Science, Research and Innovation Fund. Our gratitude is also extended to the Department of Foreign Languages and the Faculty of Liberal Arts at Prince of Songkla University for the support of the development of this paper.

About the Authors

Patson Jaihow: A lecturer at the Department of Foreign Languages, Faculty of Liberal Arts, Prince of Songkla University, Thailand. He received a doctoral degree in English Language and Applied Linguistics from the University of Birmingham, UK. His research interest lies in applied corpus linguistics, language pedagogy, and teacher education.

Kemtong Sinwongsuwat: An Associate professor of English linguistics and a board member of the Research Center for Language, Culture, and Human Development in Lower ASEAN at the Faculty of Liberal Arts, Prince of Songkla University, Hat Yai, Songkhla, Thailand. Her research

primarily focuses on developing Conversation-Analysis (CA)-informed language pedagogy to support teachers at all levels in helping their Thai EFL learners build oral communication skills.

References

Akkoyunlu, A. N., & Kilimci, A. (2017). Application of corpus to translation teaching: Practice and perceptions. *International Online Journal of Education and Teaching, 4*(4), 369–396.

Ansari, M. S. (2015). Speaking anxiety in ESL/EFL classrooms: A holistic approach and practical study. *International Journal of Educational Investigations, 2*, 38-46.

Barraja-Rohan, A-M. (2011). Using conversation analysis in the second language classroom to teach interactional competence. *Language Teaching Research*, *15*(4), 479 - 507. https://doi.org/10.1177/1362168811412878

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics, 35*(1), 81–106.

Boulton, A. (2017). Data-driven learning and language pedagogy. In S. Thorne & S. May (Eds.), *Language, education and technology: Encyclopedia of language and education.* Springer. https://doi.org/10.1007/978-3-319-02328-1_15-1

Burns, A. (2017). Research and the teaching of speaking in the second language classroom. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning,* (pp. 242-256). Routledge.

Caines, A., McCarthy, M. J., & O'Keeffe, A. (2016). Spoken language corpora and pedagogical applications. In F. Farr & L. Murray (Eds.), *Routledge handbook of language and technology* (pp. 348-361). Routledge.

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478-497). Cambridge University Press.

De Cock, S. (2010). Spoken learner corpora and EFL teaching. In M. C. Campoy, B. BellésFortuño & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 123-137). Continuum.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*(3), 393–417. https://doi.org/10.1075/ijcl.14.3.05flo

Fujii, Y. (2012). Raising awareness of interactional practices in L2

conversations: Insights from conversation analysis. *International Journal of Language Studies, 6*(3), 99-126.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ, 8*(4), 1-37. http://www.tesl-ej.org/ej32/a1.html

Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47-65). John Benjamins.

Handayani, F. (2016). Instagram as a teaching tool? Really? *Proceedings of the 4th International Seminar on English Language & Teaching (ISELT)*, *4*(1), 320-327. https://ejournal.unp.ac.id/index.php/selt/article/view/6942

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Lazaraton, A. (2014). Second language speaking. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (pp.106-120). National Geographic Learning/Heinle Cengage Learning.

Mansor, N., & Rahim, N. (2017). Instagram in ESL classroom. *Man in India, 97*, 107-114.

Markee, N. (2005). Conversation analysis for second language acquisition. In E. Hinkel (Ed.) *Handbook of research in second language teaching and learning* (pp. 355-374). Taylor and Francis.

McCarthy, M. (2004). *Touchstone: From corpus to coursebook*. Cambridge University Press.

McCarthy, M., & McCarten, J. (2022). Corpora for teaching social conversation. In R. R. Jablonkai, & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (1st ed., pp.102-115). Routledge. https://doi.org/10.4324/9781003002901

McCarthy, M., & O' Keeffe, A. (2004). Research in the teaching of speaking. *Annual Review of Applied Linguistics, 24*, 26-43. https://doi.org/10.1017/S0267190504000029

Nunan, D., & Richards, J. C. (Eds.). (2015). *Language learning beyond the classroom.* Routledge.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge University Press.

O'Keeffe, A., & Walsh, S. (2012). Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. *Corpus Linguistics and Linguistic Theory, 8*(1), 159-181.

Othman, M. (2014). Students' perception toward using classroom debate to develop critical thinking and oral communication ability. *Asian Social Science, 11*, 158-170. https://doi.org/10.5539/ass.v11n9p158

Paneerselvam, A., & Mohamad, M. (2019). Learners' challenges and English educators' approaches in teaching speaking skills in an ESL classroom: A literature review. *Creative Education, 10*(13), 3299-3305.

Pitaksuksan, N., & Sinwongsuwat, K. (2020). CA-informed interactional feature analysis of conversations in textbooks used for teaching English speaking in Thai secondary schools. *English Language Teaching, 13*(7), 140-151.

Page, M.J., McKenzie, J.E., Bossuyt, P.M. et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Review, 10*, 89. https://doi.org/10.1186/s13643-021-01626-4

Roberts, C., & Cooke, M. (2009). Authenticity in the adult ESOL classroom and beyond. *TESOL Quarterly, 43*(3), 620-642.

Şahin Kızıl, A., & Savran, Z. (2018). The integration of corpus into EFL speaking instruction: A study of learner perceptions. *International Online Journal of Education and Teaching (IOJET), 5*(2), 376-389.

Sataloff, R. T., Bush, M. L., Chandra, R., Chepeha, D., Rotenberg, B., Fisher, E. W., Goldenberg, D., Hanna, E. Y., Kerschner, J. E., Kraus, D. H., Krouse, J. H., Li, D., Link, M., Lustig, L. R., Selesnick, S. H., Sindwani, R., Smith, R. J., Tysome, J., Weber, P. C., & Welling, D. B. (2021). Systematic and other reviews: Criteria and complexities. *Annals of Otology, Rhinology and Laryngology, 130*(7), 649-652. https://doi.org/10.1177/00034894211004324

Savasci, M. (2013). Why are some students reluctant to use L2 in EFL speaking classes? An action research in tertiary level. *Proceedings of the 5th World Conference on Educational Sciences, 116,* 2682–2686. https://doi.org/10.1016/j.sbspro.2014.01.635

Seedhouse, P. (2005). Conversation analysis and language learning. *Language Teaching, 38*(4), 165-187. https://doi.org/10.1017/S0261444805003010

Sinwongsuwat, K., & Nicoletti, K. (2020). Implementing CA-T Model lessons in schools: A preliminary study in southern border provinces of Thailand. *English Language Teaching, 13*(11), 15-29.

Sinwongsuwat, K., Nicoletti, K., & Teng, B. (2018). CA-informed conversation teaching: Helping Thai students unpack English

conversation to become conversationally competent. *The Journal of Asia TEFL, 15*(3), 700-720. https://doi.org/10.18823/asiatefl.2018.15.3.9.700

Stauffer, B. (2022, January 10). What are 21st century skills? https://www.aeseducation.com/blog/what-are-21st-century-skills

Tambunan, V. D., Raja, V. L., & Sari, A. S. P. (2018). Improving students' speaking skill through story telling technique to the eleventh grade students of SMA Swasta Katolik Budi Murni 2 Medan. *Kairos English Language Teaching Journal, 2*(1), 28-46.

Wessels, S., Trainin, G., Reeves, J., Catalano, T., & Deng, Q. (2017). Pre-service teachers' confidence and attitudes toward teaching English learners. *Faculty Publications: Department of Teaching, Learning and Teacher Education, 30*(3), 443-461.

Wong, J., & Waring, H. Z. (2020). *Conversation analysis and second language pedagogy: A guide for ESL/EFL teachers*. Routledge.

Xiao, R. Z. (2008). Well-known and influential corpora. In A. Ludeling, & M. Kyto (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 383-457). Walter de Gruyter.

Appendix A

List of Publication Titles Retrieved and Used as Data for the Study

Alfehaid, A. (2018). Integrating CALL with analytical rubrics for developing speaking skills. *CALL-EJ, 19*(2), 166-186.

Babanoglu, M. P. (2014). A corpus-based study on the use of pragmatic markers as speech-like features in Turkish EFL learners' argumentative essays. *Global Conference on Linguistics and Foreign Language Teaching (LINELT), 136*, 186-193. https://doi.org/10.1016/j.sbspro.2014.05.312

Chen, M., Huang, C., Huang, S., Chang, J. S., & Liou, H. (2014). An automatic reference aid for improving EFL learners' formulaic expressions in productive language use. *IEEE Transactions on Learning Technologies, 7*(1), 57-68. https://doi.org/10.1109/TLT.2013.34

Dang, T., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research, 26*(4), 617-641. https://doi.org/10.1177/1362168820911189

Garrote, M., Kimura, C., Matsui, K., Moreno-Sandoval, A., & Takamori, E.

(2015). C-ORAL-JAPON: Corpus of spontaneous spoken Japanese. *Corpus Linguistics and Linguistic Theory, 11*(2), 373-392. https://doi.org/10.1515/cllt-2013-0004

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *Recall, 26*(2), 225-242. https://doi.org/10.1017/S0958344014000044

Genç-Yöntem, E., & Eveyik-Aydın, E. (2022). The compilation of a developmental spoken English corpus of Turkish EFL learners. *Research in Corpus Linguistics, 10*(1), 45-62.

Gu, X., & Xu, Z. (2021). Sustainable development of EFL learners' research writing competence and their identity construction: Chinese novice writer-researchers' metadiscourse use in English research articles. *Sustainability, 13*(17), 9523. https://doi.org/10.3390/su13179523

Huang, L., & Graf, T. (2020). Speech rate and pausing in English: Comparing learners at different levels of proficiency with native speakers. *Taiwan Journal of TESOL, 17*(1), 57-86. https://doi.org/10.30397/TJTESOL.202004_17(1).0003

Kayumova, A. R., Sadykova, G. V., Anthony, N. & Galimullina, Z. F. (2017). English-Russian academic discourses: Points of convergence and divergence. *Modern Journal of Language Teaching Methods, 7*(9), 324-332.

Khojasteh, L., & Shokrpour, N. (2014). Corpus linguistics and English language teaching materials: A review of recent research. *Khazar Journal of Humanities and Social Sciences, 17*(3), 5-17.

Khoroshilova, S., Kostina, E., & Ovechkina, I. (2015). Examining Russian tertiary-level students' attitude to the use of a corpus-based approach in language classes. *Psychology and Psychiatry, Sociology and Healthcare, Education, 2*, 415-422.

Kwon, H. (2022). English learner corpora and research in Korea. *Corpora*, 17, 5-22. https://doi.org/10.3366/cor.2022.0244

Lin, P. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, *42*, 164-176. https://doi.org/10.1016/j.system.2013.11.010

Neves, M. (2020). Effects of imaginary identification with the English language in excerpts of open responses about university education. *Revisits De Estudos Da Linguagem, 28*(3), 1291-1308. https://doi.org/10.17851/2237-2083.28.3.1291-1308

Roca-Varela, M. (2013). High-frequency English words in spoken learner language: Actual, career and pretend as a case in point. Corpus resources for descriptive and applied studies. *Proceedings of the*

Current Challenges and Future Directions: 5th International Conference on Corpus Linguistics, 95, 557-562. https://doi.org/10.1016/j.sbspro.2013.10.682

Şahin Kızıl, A., & Savran, Z. (2018). The integration of corpus into EFL speaking instruction: A study of learner perceptions. *International Online Journal of Education and Teaching (IOJET), 5*(2), 376-389.

Sung, M., & Kim, H. (2016). Tracing developmental changes in L2 learners' structuring of phrasal verbs: A corpus study of native and non-native argumentative essays. *3L-Language Linguistics Literature-The Southeast Asian Journal of English Language Studies, 22*(2), 151-166.

Wang, H., & Zou, Y. (2019). A corpus-based study of semantic collocations of the verb "feel" in English public speaking setting: Chinese EFL VS native speakers. *International Journal of English Linguistics, 9*(1), 251-260. https://doi.org/10.5539/ijel.v9n1p251

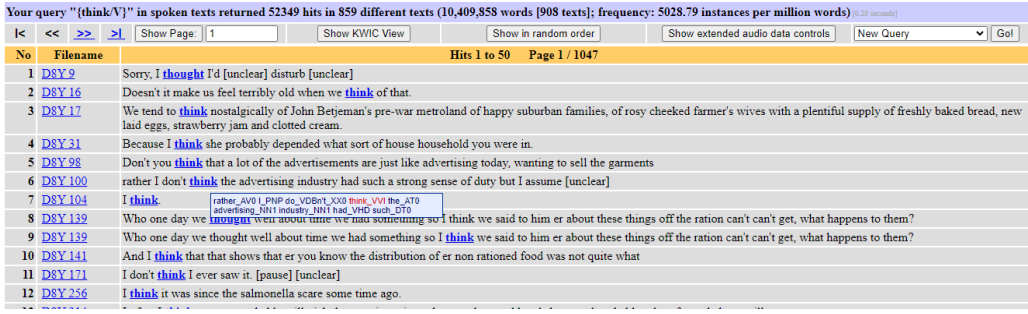
Yin, Z. H. (2015). The use of cohesive devices in news language: Overuse, underuse or misuse? *RELC Journal: A Journal of Language Teaching and Research, 46*(3), 309-326. https://doi.org/10.1177/0033688215597578

Appendix B

Concordances for the Lemma *'think'* Retrieved from the BNC

Figure 1

Concordances for the Lemma 'think' in a Sentence View Mode



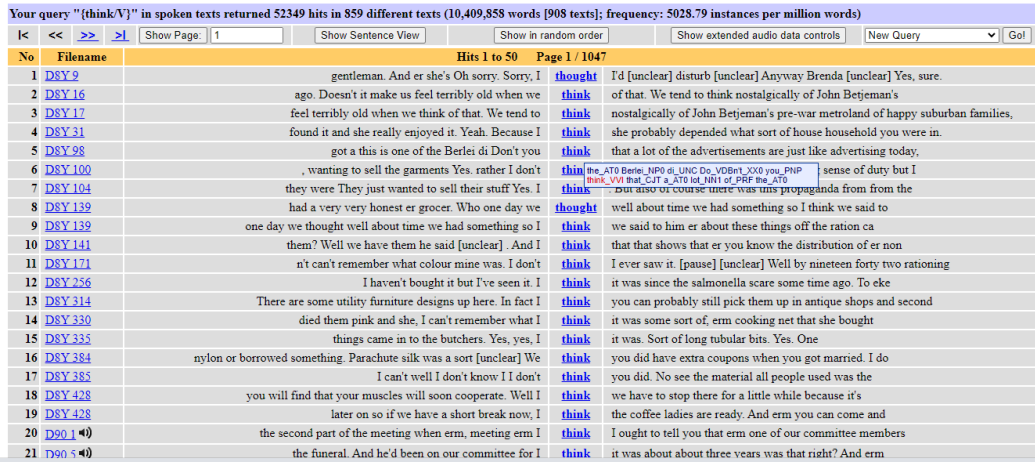
Your query "{think/V}" in spoken texts returned 52349 hits in 859 different texts (10,409,858 words [908 texts]; frequency: 5028.79 instances per million words) [0.29 seconds]

|< << >> >| Show Page: 1 Show KWIC View Show in random order Show extended audio data controls New Query Go!

No	Filename	Hits 1 to 50 Page 1 / 1047
1	D8Y 9	Sorry, I **thought** I'd [unclear] disturb [unclear]
2	D8Y 16	Doesn't it make us feel terribly old when we **think** of that.
3	D8Y 17	We tend to **think** nostalgically of John Betjeman's pre-war metroland of happy suburban families, of rosy cheeked farmer's wives with a plentiful supply of freshly baked bread, new laid eggs, strawberry jam and clotted cream.
4	D8Y 31	Because I **think** she probably depended what sort of house household you were in.
5	D8Y 98	Don't you **think** that a lot of the advertisements are just like advertising today, wanting to sell the garments
6	D8Y 100	rather I don't **think** the advertising industry had such a strong sense of duty but I assume [unclear]
7	D8Y 104	I **think**. rather_AV0 I_PNP do_VDBnt_XX0 think_VVI the_AT0 advertising_NN1 industry_NN1 had_VHD such_DT0
8	D8Y 139	Who one day we thought well about time we had something so I think we said to him er about these things off the ration can't can't get, what happens to them?
9	D8Y 139	Who one day we thought well about time we had something so I **think** we said to him er about these things off the ration can't can't get, what happens to them?
10	D8Y 141	And I **think** that that shows that er you know the distribution of er non rationed food was not quite what
11	D8Y 171	I don't **think** I ever saw it. [pause] [unclear]
12	D8Y 256	I **think** it was since the salmonella scare some time ago.

Figure 2

Concordances for the Lemma 'think' in a KWIC Mode



Your query "{think/V}" in spoken texts returned 52349 hits in 859 different texts (10,409,858 words [908 texts]; frequency: 5028.79 instances per million words)

|< << >> >| Show Page: 1 Show Sentence View Show in random order Show extended audio data controls New Query Go!

No	Filename	Hits 1 to 50 Page 1 / 1047		
1	DSY 9	gentleman. And er she's Oh sorry. Sorry, I	thought	I'd [unclear] disturb [unclear] Anyway Brenda [unclear] Yes, sure.
2	DSY 16	ago. Doesn't it make us feel terribly old when we	think	of that. We tend to think nostalgically of John Betjeman's
3	DSY 17	feel terribly old when we think of that. We tend to	think	nostalgically of John Betjeman's pre-war metroland of happy suburban families,
4	DSY 31	found it and she really enjoyed it. Yeah. Because I	think	she probably depended what sort of house household you were in.
5	DSY 98	got a this is one of the Berlei di Don't you	think	that a lot of the advertisements are just like advertising today,
6	DSY 100	, wanting to sell the garments Yes. rather I don't	think	sense of duty but I
7	DSY 104	they were They just wanted to sell their stuff Yes. I	think	But also of course there was this propaganda from from the
8	DSY 139	had a very very honest er grocer. Who one day we	thought	well about time we had something so I think we said to
9	DSY 139	one day we thought well about time we had something so I	think	we said to him er about these things off the ration ca
10	DSY 141	them? Well we have them he said [unclear] . And I	think	that that shows that er you know the distribution of er non
11	DSY 171	n't can't remember what colour mine was. I don't	think	I ever saw it. [pause] [unclear] Well by nineteen forty two rationing
12	DSY 256	I haven't bought it but I've seen it. I	think	it was since the salmonella scare some time ago. To eke
13	DSY 314	There are some utility furniture designs up here. In fact I	think	you can probably still pick them up in antique shops and second
14	DSY 330	died them pink and she, I can't remember what I	think	it was some sort of, erm cooking net that she bought
15	DSY 335	things came in to the butchers. Yes, yes, I	think	it was. Sort of long tubular bits. Yes. One
16	DSY 384	nylon or borrowed something. Parachute silk was a sort [unclear] We	think	you did have extra coupons when you got married. I do
17	DSY 385	I can't well I don't know I I don't	think	you did. No see the material all people used was the
18	DSY 428	you will find that your muscles will soon cooperate. Well I	think	we have to stop there for a little while because it's
19	DSY 428	later on so if we have a short break now, I	think	the coffee ladies are ready. And erm you can come and
20	D90 1	the second part of the meeting when erm, meeting erm I	think	I ought to tell you that erm one of our committee members
21	D90 5	the funeral. And he'd been on our committee for I	think	it was about about three years was that right? And erm

the_AT0 Berlei_NP0 di_UNC Do_VDBnt_XX0 you_PNP think_VVI that_CJT a_AT0 lot_NN1 of_PRF the_AT0