

Advancing the Research Agenda of Interlanguage Pragmatics: The Role of Learner Corpora

Marcus Callies 4

1 Pragmatics in Second Language Acquisition Research: A Critical Assessment 5 6

1.1 Interlanguage Pragmatics and Its Scope of Inquiry 7

Broadly defined, pragmatics as a discipline can be conceived of as “the study of language from the point of view of the users, especially of the choices they make, the constraints they encounter in using language in social interaction, and the effects their use of language has on the other participants in an act of communication” (Crystal 2003: 364). Leech (1983: 10f.) distinguishes between two components of general pragmatics. First, he defines socio-pragmatics as “the sociological interface of pragmatics” that focuses on the conditions of language use which derive from the social situation, i.e. the social setting of language use, including variables such as cultural context, social status or social distance of speakers. Second, pragmalinguistics is “the more linguistic end of pragmatics”, considering the particular linguistic resources which a given language provides for conveying particular illocutions, i.e. the range of structural resources from which speakers can choose when using language in a specific communicative situation, e.g. speech act verbs, imperatives, politeness markers, pragmatic markers etc. 8
9
10
11
12
14
15
16
17
18
19
20
21

The study of pragmatics as a field of inquiry within Second Language Acquisition (SLA) research is usually referred to as Interlanguage Pragmatics (ILP). ILP is commonly defined as “the study of nonnative speakers’ comprehension, production, and acquisition of linguistic action in L2” (Kasper 2010: 141). While this suggests a relatively broad range of research topics as in pragmatics in general, ILP to date 22
23
24
25
26

M. Callies (✉)
English-Speaking Cultures, University of Bremen,
FB 10, Bibliothekstr.1, GW2, Bremen 28359, Germany
e-mail: callies@uni-bremen.de

27 has operated on a fairly narrow understanding of what constitutes linguistic action
28 in L2. One of the main reasons for this is that traditionally, ILP has been heavily
29 influenced by and largely modeled on cross-cultural pragmatics, adopting its
30 research topics, theories and methodologies (Kasper 2010: 141). Thus, it has pre-
31 dominantly been concerned with politeness phenomena by investigating foreign/
32 second language (L2) learners' comprehension and production of a variety of
33 speech act types such as requests, apologies, refusals, complaints, compliments and
34 compliment responses, and the use of internal and external modification to these
35 speech acts. The findings of these investigations have subsequently been compared
36 with native speaker performance.

37 In their review of research methods in ILP, Kasper and Dahl (1991) define the
38 field "in a narrow sense, referring to nonnative speakers' (NNSs') comprehension
39 and production of speech acts, and how their L2-related speech act knowledge is
40 acquired" (1991: 216). Studies addressing topics like conversational management,
41 discourse organization, or sociolinguistic aspects of language, e.g. address forms,
42 were explicitly left outside of the scope of this article. This narrow view has been
43 taken over in many overview articles and book chapters on ILP that have been pub-
44 lished since. For example, Ellis (2008: 160), explicitly referring to Kasper and Dahl
45 (1991), also adopts the narrow sense of ILP arguing that this aspect of pragmatics
46 has received the greatest attention in SLA research. Ellis even maintains that the
47 scope of pragmatics in ILP is "relatively well-defined. Researchers have investi-
48 gated what speakers accomplish when they perform utterances in terms of: (1) inter-
49 actional acts and (2) speech acts" (2008: 159). In sum, this perspective has led to a
50 narrow research focus and sociopragmatic bias in ILP where the dominant area of
51 investigation has been the speech act.

52 Almost 20 years after Kasper and Dahl's review paper, Bardovi-Harlig (2010)
53 provided a state-of-the-art meta-analysis of published research in ILP. Noting that
54 "the study of interlanguage pragmatics has not typically been as broad as the areas
55 outlined by the definition of pragmatics used in the handbook",¹ she states that
56 "within second language studies, work in pragmatics has often been narrower than
57 in the field of pragmatics at large" and that "there seems to be less agreement in the
58 field about the scope of *pragmatics*" (2010: 219f.; emphasis in original). Her meta-
59 analysis of a sample of 152 research articles published between 1979 and 2008
60 reveals that in 99 out of the 152 studies reviewed (65.1 %), pragmatic competence
61 was operationalized in terms of speech acts. This leads her to conclude that "the
62 dominant area of investigation within interlanguage pragmatics has been the
63 speech act" (2010: 219). Only few studies have investigated other pragmatic phe-
64 nomena, e.g. turn structure (sequencing of turns, repair, alignment, greeting and
65 leave taking), pragmalinguistic devices, i.e. grammatical and lexical devices

¹Bardovi-Harlig refers to the *Handbooks of Pragmatics* series published with DeGruyter Mouton. In the general preface to the series, the editors state that all the handbooks in the series share the same wide understanding of pragmatics as the scientific study of all aspects of linguistic behaviour.

including routines (e.g. modal particles, adverbials, formulas), and pragmatic interpretation (meta-pragmatic knowledge and assessment, e.g. in the form of ranking or rating). 66
67
68

In 2005, Müller provided one of the first comprehensive studies of discourse markers in learner English. While the use of discourse markers in native English has been studied extensively in pragmatics in the last decades, Müller concluded in her overview chapter on pragmatics in SLA that “there is little in the area of second language acquisition and applied linguistics which deals explicitly with discourse markers. The focus in this area is either on grammatical features or, as far as pragmatic competence goes, on speech acts” (2005: 23). 69
70
71
72
73
74
75

Callies (2009a) draws attention to the pragmlinguistic component of pragmatics and its interplay with grammar. He examined advanced L2 learners' comprehension and use of focus constructions, i.e. pragmatically-motivated variations of the basic word order. Outlining that knowledge of the principles of information organization in discourse, and the use of linguistic devices for information highlighting clearly relates to L2 pragmatic knowledge, Callies suggests that further research into L2 learners' abilities at the syntax-pragmatics interface may also be a rewarding enterprise with respect to the interplay of grammatical and pragmlinguistic knowledge, an important yet unresolved issue in ILP. 76
77
78
79
80
81
82
83
84

Dippold (2009) notes that ILP not only prioritizes research on the expression of L2 politeness and the acquisition of politeness strategies, but that it also does so in a decontextualized manner that takes little account of the situatedness of linguistic discourse. She argues that ILP should move away from its focus on politeness in a limited set of speech acts and focus also on self-presentation. 85
86
87
88
89

In sum, this clearly suggests that the significance of L2 pragmatic knowledge beyond the domain of speech acts has been neglected in ILP research to date. However, the scope of pragmatics in the context of SLA does not necessarily have to be a narrow one. In many broad definitions such as the one given by Kasper (2010: 141) (“the study of nonnative speakers' comprehension, production, and acquisition of linguistic action in L2”) the scope of research in ILP is not restricted to issues of politeness and the domain of speech acts. Kasper and Rose (2002) have proposed the concept of “pragmatics-as-perspective” which “has the advantage of being inclusive and open to study new research objects *as* pragmatics, without precluding them from being examined from a different angle as well” (2002: 5; emphasis in original). In fact, recent developments suggest that there is a growing awareness in the field that L2 pragmatics is more than speech acts and that the scope of inquiry needs to be adjusted accordingly. For example, LoCastro (2011: 333) observes “a movement away from an almost exclusive focus on speech acts, particularly apologies, requests, refusals, and compliments, and formulaic language to a much broader view of language in use”, pointing to studies that have examined topic marking, negation strategies, referent introduction and maintenance, self-qualification, discourse markers, modal particles, definiteness, and text organization. LoCastro also notes that “many of these studies delve into complexities in signaling pragmatic meaning beyond the more commonplace comparisons of a speech act in learners' L2 production and the native speaker enactment of the same speech act” (2011: 333). 90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110

111 **1.2 Modeling L2 Pragmatic Knowledge**

112 In this section, I argue that pragmatic knowledge in an L2 clearly includes more
113 than the sociopragmatic and pragmalinguistic abilities for understanding and per-
114 forming speech acts and propose a more encompassing definition of L2 pragmatic
115 knowledge. Standard descriptions of ILP frequently use notions like “linguistic
116 action in L2” (Kasper 2010: 141) and “L2 pragmatic knowledge” (Kasper and Rose
117 1999: 81; Gass and Selinker 2008: 287) respectively to refer to the general domain
118 of inquiry. But what exactly constitutes L2 pragmatic knowledge? Definitions of
119 pragmatic knowledge or competence² range from rather broad and general ones, e.g.
120 “the ability to use language appropriately in a social context” (Taguchi 2009: 1) to
121 more detailed ones, e.g. “the knowledge of the linguistic resources available in a
122 given language for realizing particular illocutions, knowledge of the sequential
123 aspects of speech acts and finally, knowledge of the appropriate contextual use of
124 the particular languages’ linguistic resources” (Barron 2003: 10). While Barron’s
125 proposal draws a useful distinction between pragmalinguistic and sociopragmatic
126 knowledge, it reflects the bias in mainstream ILP in that it centers around the con-
127 cept of illocutionary acts, thus narrowing down the scope of pragmatic knowledge
128 to sociopragmatics.

129 There are a number of models of language proficiency that aim to capture the
130 ability of L2 learners to use language in social interaction, all of which acknowl-
131 edge to some degree the importance to acquire pragmatic competence in L2
132 learning. The two most influential constructs, communicative competence and
133 communicative language ability, will be discussed briefly in turn. In general
134 terms, communicative competence can be defined as “the fundamental concept
135 of a pragmalinguistic model of linguistic communication: it refers to the reper-
136 toire of know-how that individuals must develop if they are to be able to com-
137 municate with one another appropriately in the changing situations and
138 conditions” (Bußmann 1996: 84). In reaction to Chomsky’s dichotomy of com-
139 petence and performance, in which the notion of linguistic competence only
140 includes knowledge of abstract grammatical rules and sets aside contextual fac-
141 tors of language use, Hymes (1972) introduced the concept of communicative
142 competence, containing both grammatical competence and knowledge of the
143 sociocultural rules of language use. Canale (1983), building on Canale and Swain
144 (1980), suggested a model of communicative competence that includes four
145 major components:

- 146 • **GRAMMATICAL COMPETENCE** (knowledge of the language code: vocabu-
147 lary, phonology, spelling, morphology, and syntax needed to produce and under-
148 stand well-formed sentences);
- 149 • **SOCIOLINGUISTIC COMPETENCE** (knowledge of appropriate use and
150 understanding of language in different sociolinguistic contexts, with emphasis
151 on appropriateness of both meanings and forms);

²The two terms are frequently used interchangeably in the literature.

- DISCOURSE COMPETENCE (knowledge of how to combine and interpret grammatical forms and meanings to achieve unified texts in different modes by using cohesion devices and coherence rules);
- STRATEGIC COMPETENCE (knowledge of the verbal and non-verbal strategies used to compensate for breakdowns in communication and to enhance the rhetorical effect of utterances).

Although these four components are described separately in Canale's model, it should be made clear that they interact with each other and also partly overlap. Pragmatic competence is not recognized separately here, but implicitly included in the sociolinguistic component in a predominantly sociopragmatic, that is speech-act based sense. In addition, Canale sees discourse competence as bridging the gap between grammatical and sociolinguistic competence and includes it as a separate component, predominantly understood in a textlinguistic sense (hence the focus on coherence and cohesion).

Building on the work of Hymes and Canale, Bachman (1990) introduces the model of communicative language ability which is composed of three components:

- LANGUAGE COMPETENCE, "a set of specific knowledge components that are utilized in communication via language";
- STRATEGIC COMPETENCE, "the mental capacity for implementing the components of language competence in contextualized communicative language use", and
- PSYCHOPHYSIOLOGICAL MECHANISMS, "the neurological and physiological processes involved in the actual execution of language as a physical phenomenon" (1990: 84).

Particularly interesting is the component of language competence which is further subdivided into

- ORGANISATIONAL COMPETENCE, which contains the modules of GRAMMATICAL COMPETENCE (the knowledge of vocabulary, morphology, syntax, and phonology), and TEXTUAL COMPETENCE, which "includes the knowledge of the conventions for joining utterances together to form a text, which is essentially a unit of language – spoken or written – consisting of two or more utterances or sentences that are structured according to rules of cohesion and rhetorical organisation" (1990: 88), and
- PRAGMATIC COMPETENCE, which intends to capture the speaker's or writer's ability to achieve his or her communicative intentions through the use of language, subsuming ILLOCUTIONARY COMPETENCE (knowledge of expressing and interpreting language functions and speech acts) and SOCIOLINGUISTIC COMPETENCE, or "sensitivity to, or control of the conventions of language use that are determined by the features of the specific language use context" (1990: 94).

Bachman's construct thus explicitly includes pragmatic competence, which is, however, described primarily in a sociopragmatic sense.

195 A more detailed model of discourse competence building on Canale's construct
196 of communicative competence has been proposed by Archibald (1994: 59f.). It includes
197 four components:

- 198 • COHESION: knowledge of how the lexico-grammatical structures of language
199 may be used to produce connectedness in text;
- 200 • COHERENCE: knowledge of the principles of relevance and cooperation and
201 the illocutionary functions of language;
- 202 • SITUATIONALITY: knowledge of how a text is related to discourse context, and
203 the role of background knowledge;
- 204 • INFORMATION STRUCTURE: knowledge of thematic structure, the ordering
205 of given and new information.

206 In sum, an integration of Canale's and Archibald's modules of discourse compe-
207 tence, largely covering the pragma- and textlinguistic component of pragmatics, and
208 Bachman's definition of pragmatic competence, reflecting the sociopragmatic com-
209 ponent, seems to account best for the complex nature of L2 pragmatic competence.
210 I thus propose the following definition of pragmatic knowledge: L2 pragmatic
211 knowledge is the knowledge of the (pragma-) linguistic resources available in a
212 particular language for realizing communicative intentions, and the knowledge of
213 the appropriate socio-contextual use of these resources. Pragmalinguistic knowl-
214 edge is a component of L2 pragmatic knowledge which relates to learners' knowl-
215 edge of the structural linguistic resources available in a given language for realizing
216 particular communicative effects, and knowledge of the appropriate contextual use
217 of these resources.

218 **2 Going Beyond Speech Acts: The Role of Learner Corpora**

219 Research in ILP has largely relied on elicited assessment and production data, most
220 typically in the form of pseudo-oral discourse completion or production tasks.
221 According to Bardovi-Harlig's meta-analysis, only 27 % of the studies she surveyed
222 collected and analyzed authentic language samples (2010: 241). Despite the firm
223 belief that the most authentic data in pragmatic research is provided by spontaneous
224 speech gathered through observation, the discourse completion task (DCT) has
225 become almost the standard technique due to the manifold administrative advan-
226 tages of using written questionnaires.³ The DCT is a data collection technique
227 widely used to elicit production data about sociopragmatic behaviour in a specific
228 communicative context. DCTs are usually administered in the form of written ques-
229 tionnaires that contain several contextualized descriptions designed to create com-
230 municative situations. Informants are then asked to provide direct speech in a
231 written response to a stimulus, e.g. a first turn provided to them. DCTs come in

³LoCastro (2011: 331) sees this as another reason for the dominance of speech act research in ILP.

various formats. The classic format, in which informants have to fill in only one turn at talk, consists of an open turn for the required response (sometimes prefaced by an initiation of a fictitious interlocutor), and a rejoinder to the turn to be provided by the informant. The free DCT, also called dialogue construction task, has an open response format. It can be introduced by a first pair part, but includes no rejoinder to the required response. The response can be verbal, non-verbal, or the informant is given the possibility to opt out, i.e. to provide no response at all. Another type is the discourse production task in which participants are only provided with a contextualized situational description and have to construct a short dialogue sequence involving two or more participants.

The benefits and disadvantages of using elicitation data are widely recognized and discussed in the field, and there is by now a considerable amount of literature on various issues of research methodology in ILP.⁴ Obviously, DCTs make it possible to collect large amounts of data in relatively short time and with comparatively little effort. Moreover, the context and situational descriptions can be manipulated to constrain the response so that the required, often highly specific linguistic structures can successfully be elicited. Also, social variables can be controlled much more systematically than in naturally-occurring situations. But there are also several disadvantages. The DCT is a pseudo-oral format, because despite its oral setting, it is more likely to elicit written than spoken language. Apparently, informants do not write as spontaneously as they would speak, and do not necessarily write down what they would say, but rather what they imagine is expected or should be said. Thus, data elicited in such a way are more likely to reflect interactive norms and underlying social and cultural values acquired in communication or learnt in the process of socialization. While the recording of naturally occurring talk enables the researcher to study the organization and realization of talk-in-interaction in natural settings, elicited data from DCTs indirectly reflect prior experience with language. Several studies have compared various formats of DCTs with other common data collection methods to investigate the effects of the instrument on the results (e.g. Sasaki 1998; Yuan 2001; Golato 2003). While oral formats, e.g. role plays, due to their interactive nature, induce longer responses and a larger number and greater variety of strategies/formulas than questionnaires, written formats produce more direct responses.

The compilation and accessibility of computer corpora and software tools for corpus analysis has revolutionized (applied) linguistics in the last two decades. Corpus linguistics and pragmatics can be considered related, but historically distinct disciplines in that the latter is a subfield of linguistics while the former is often considered a methodological approach to carrying out linguistic research (Andersen 2011: 588). Nevertheless, corpus linguistics and pragmatics can be said to form a “mutualistic entente” (Romero-Trillo 2008) in that they are joint forces in the common cause to work with real usage data, thus more convincingly addressing some specifics of language usage by combining the methodologies

⁴See e.g. the overviews by Kasper (2008) and Ellis (2008: 163–169). Callies (2012b) summarizes the advantages and disadvantages of the DCT.

273 that underlie both disciplines.⁵ In fact, the marriage of corpus linguistics and
274 pragmatics has more recently given rise to a new hybrid subfield referred to as
275 “corpus pragmatics”.⁶

276 In ILP, learner corpora – due to their very nature of being large systematic collec-
277 tions of authentic, continuous and contextualized language use (spoken or written)
278 by L2 learners stored in electronic format – can help overcome several problems
279 and limitations posed by the dominance of data elicitation techniques in ILP to date.
280 Not only do learner corpora enable researchers to study a much broader range of
281 different phenomena, but they can also provide results that may be viewed as more
282 reliable, valid, and generalizable across populations without the lack of authenticity
283 and replicability that often arises from the use of other types of data. Learner cor-
284 pora also make it possible to abstract away from individual learners and identify a
285 corpus-based, supra-individual description of a specific learner group while at the
286 same time providing insights into intra-group variability. Such variability and indi-
287 vidual differences have important implications for learner corpus analysis and com-
288 pilation that will be addressed in detail in the case studies in Sect. 3. Additionally,
289 learner corpora can be the basis for quantitatively oriented studies that are subjected
290 to statistical analyses and create an opportunity for between-methods triangulation
291 and alternative views to qualitative, ethnographic studies that have been common in
292 pragmatics in general.

293 In particular, the availability of spoken learner corpora such as the *Louvain*
294 *International Database of Spoken English Interlanguage* (LINDSEI, Gilquin et al.
295 2010) has enabled researchers to study a wider range of pragmatic features of
296 learner language in the spoken mode.⁷ The LINDSEI was compiled by an interna-
297 tional research team and consists of spoken data, i.e. transcripts of interviews
298 between learners of English as a foreign language (EFL) and English native-
299 speaker or non-native-speaker interviewers. The learners are university undergrad-
300 uates in their twenties whose proficiency level ranges from higher intermediate to
301 advanced (being assessed on external criteria, most importantly their institutional
302 status, e.g. the time they spent learning English at school and university and the
303 fact that they are university undergraduates in English). The LINDSEI includes
304 subcorpora of learners from 11 mother tongue backgrounds (e.g. German, French,
305 Italian, Japanese, Polish, and Spanish) with 50 interview transcripts per subcorpus,
306 i.e. a total of about 100,000 words per component. Each interview lasts approxi-
307 mately 15 min and involves three tasks: (1) a warm-up sequence in which inter-
308 viewer and interviewee talk about a set topic, (2) a free discussion, and (3) a picture
309 description.

⁵See Andersen (2011) and Rühlemann (2011) for recent overviews of the interrelation of the two fields.

⁶See e.g. the titles of the recent/upcoming publications by Felder et al. (2011) and Aijmer and Rühlemann (forthcoming).

⁷See e.g. the papers in Romero-Trillo (2008) and the studies on the list of publications based on the LINDSEI provided by the Centre for English Corpus Linguistics in Louvain-al-Neuve, Belgium, at <http://www.uclouvain.be/en-cecl-lindsei-biblio.html>.

Using data from corpora of spoken interlanguage, it is now possible to systematically examine lexico-grammatical patterns and syntactic structures that are part of the grammar of conversation on a broad empirical basis (see e.g. Mukherjee 2009 for a study along these lines). Recent studies have investigated individual pragmalinguistic units, e.g. discourse markers (e.g. Müller 2004, 2005; Aijmer 2004, 2009, 2011), modal particles (e.g. Belz and Vyatkina 2005) and tag questions (Ramirez and Romero-Trillo 2005), as well as other features of turn- and discourse structure, e.g. performance phenomena like hesitations, repetitions and disfluencies (Götz 2007; Gilquin 2008) or filled and unfilled pauses (see e.g. Brand and Götz 2011 and Götz 2013 for studies that examine and operationalize these features as measures of fluency). The present chapter makes a contribution to research on the grammar of conversation in learner English and focuses on the pragmalinguistic component of L2 pragmatic knowledge, in particular as it relates to information highlighting in discourse.

3 Case Studies

An area where pragmalinguistic devices abound and are of crucial importance is discourse pragmatics, the “general domain of inquiry into the relationship between grammar and discourse” (Lambrecht 1994: 2). More specifically, I will be concerned with lexico-grammatical and syntactic means of information highlighting located at the interface of lexico-grammar, syntax and pragmatics. This interface is often referred to as information structure or information packaging, viz. the structuring of sentences by syntactic, prosodic, or morphological means that arises from the need to meet certain communicative demands, e.g. emphasizing a certain point, correcting a misunderstanding, or repairing a communicative breakdown.⁸ Information highlighting is clearly pragmatically motivated because, more generally speaking, it serves to express certain pragmatic functions in discourse, e.g. intensification or contrast. Compared to their frequency of occurrence and difficulty of acquisition there are still remarkably few (corpus-based) studies that have examined the linguistic means of information highlighting in learner language from a pragmalinguistic perspective (see e.g. Boström Aronsson 2003; Herriman and Boström Aronsson 2009; Callies 2008a, b, 2009a, b). L2 learners’ knowledge (that includes awareness, comprehension, and production) of discourse organization and the (contextual) use of linguistic means of information highlighting is thus still an underexplored area in SLA research, as is the interplay of pragmalinguistic knowledge and discourse organization in general. Interface relations, opaque form-meaning mappings, optionality and discourse-motivated preferences are assumed to be the main areas of difficulty in advanced SLA (DeKeyser 2005). Recent findings

⁸Deppermann (2011) provides a recent overview of the role and relevance of pragmatics for grammar, in particular as to the structuring and packaging of information and the framing of discursive action by means of grammatical constructions such as clefts.

t1.1 **Table 1** Learner corpora used in the case studies

t1.2	Name	Writers' L1	Professional status	No. of interviews	No. of turns (only interviewees)
t1.4	LINDSEI-F	French	University students	50	5,504
t1.5	LINDSEI-G	German	University students	50	6,051
t1.6	LOCNEC	British English	University students	50	8,436

t1.7 In view of the manifold problems to operationalize the concept of sentence in transcribed spoken
 t1.8 language and thus, to count the amount of sentences in the corpora, I chose to apply the number
 t1.9 of speech turns as a basis of comparison

347 suggest that information structure management is problematic even for advanced L2
 348 learners and that such learners have only a limited awareness of the appropriate use
 349 of lexical and syntactic focusing devices in formal and informal registers (Callies
 350 2009a).

351 The following sections report on two learner-corpus studies that investigates L2
 352 learners' use of specific lexico-grammatical means of information highlighting in
 353 English: emphatic *do* and a special type of cleft construction introduced by the deictic
 354 demonstratives *that* or *this* (demonstrative clefts). Three research questions will
 355 be examined:

- 356 1. Are there differences in the frequencies of use of emphatic *do* and demonstrative
 357 clefts in the speech of native speakers of English and learners of English as a
 358 foreign language?
- 359 2. Are there differences in how native speakers and learners use these devices con-
 360 textually, i.e. as to their discourse functions and characteristic lexical co-
 361 occurrence patterns?
- 362 3. Are there differences between learners from different L1 backgrounds, and if so,
 363 how can these be explained?

364 3.1 Data and Methodology

365 Both case studies are contrastive interlanguage analyses (CIA) based on corpora of
 366 spoken interlanguage. In a CIA, two types of comparisons are combined. First, the
 367 interlanguage of a certain learner group, e.g. German learners of English, is com-
 368 pared with the language of English native speakers in order to pinpoint possible
 369 differences between the two groups. This comparison is then subsequently com-
 370 bined with a corresponding analysis of the interlanguage produced by a second
 371 group of learners, e.g. French learners of English. For the present case studies, the
 372 learner data are drawn from the German and French components of the LINDSEI
 373 (Gilquin et al. 2010). For comparable native speaker data the *Louvain Corpus of*
 374 *Native English Conversations* (LOCNEC) was used. The LOCNEC contains tran-
 375 scribed interviews with native speakers of British English (university students at
 376 Lancaster university in the UK) aged between 18 and 30 years. The interviews
 377 involved the same tasks, topics and stimuli that were used for the interviews in the
 378 LINDSEI. Table 1 provides an overview of the corpora.

The target structures were extracted semi-automatically⁹ using *WordSmith Tools 5* (Scott 2008), followed by manual inspection and filtering of false positives. The analysis of the data consisted in a quantitative analysis of frequencies of occurrence and a qualitative study of lexical co-occurrence patterns (e.g. verbs, connectives, pragmatic markers, intensifying adverbs) and discourse functions.

3.2 Emphatic *Do*

Emphatic *do* is a lexico-grammatical means of information highlighting that commonly serves to emphasize the meaning of a following predicate (underlined in example 1).

- (1) <A> So you want to become a teacher now. <\A>
 I **do** want to become a teacher yeah I always thought I wanted to teach English. But now I want to teach French. <\B> (LOCNEC)¹⁰

Emphatic *do* is discussed only briefly in the standard reference grammars of English (Quirk et al. 1985; Biber et al. 1999; Huddleston 2002) and there are only very few corpus-based studies that have examined this feature in detail (Nevalainen and Rissanen 1986; Luzón Marco 1998/99). Emphatic *do* usually carries nuclear stress and is one of the few options to explicitly highlight its following predicate. Syntactic options like predicate fronting or *wh*-clefting are available to highlight a verb phrase, but are contextually much more restricted.

Table 2 shows that the frequential distribution of emphatic *do* varies across spoken and written registers.

Emphatic *do* is clearly most frequently used in spoken language. In addition, a breakdown of the individual genre sections for the spoken register in the BNC shows that it is particularly frequent in highly argumentative contexts such as (parliamentary) debates, meetings, lectures, interviews, and discussions, where its frequency even rises to more than a thousand occurrences per million words.

There are two views as to whether emphatic *do* expresses both contrastive and non-contrastive emphasis or whether it exclusively has a contrastive function. Quirk et al. (1985) argue that it focuses on the operator [i.e. the predicate, MC] either for contrastive or emotive emphasis. Huddleston (2002: 97f.) states that it expresses emphatic polarity, emphasizing the positive or negative polarity of a clause. As an

⁹To retrieve instances of emphatic *do* I ran a search for the forms *do*, *does* and *did* followed by an infinitive, excluding instances of grammatically conditioned inversion after negatives as in *Not only did they...*, *Even slower did...*, and elliptical sentence forms, e.g. *Yes we do* or *They never did so*. For demonstrative clefts the search involved all instances of *that* and *this* followed by a form of *be* ('s, is, was) and a *wh*-word (*what*, *when*, *why*, *where*, *how*).

¹⁰In the LOCNEC and the LINDSEI, turns marked with <A> indicate the interviewers' turns, while turns marked with mark the interviewees' turns. The transcription guidelines for the LINDSEI can be retrieved from the following webpage: <http://www.uclouvain.be/en-307849.html>. Unfortunately, some of the transcription conventions used for the LOCNEC have not been updated to follow those of the LINDSEI. For example, overlapping speech in the LOCNEC is still indicated by means of square brackets instead of the explicit tag <overlap />.

t2.1 **Table 2** Frequencies of occurrence of emphatic *do* across registers in four corpora (per million
t2.2 words)

t2.3					Academic
t2.4	Register corpus	Speaking	Fiction	News	writing
t2.5	<i>Longman Spoken and Written English</i>	400	300	150	150
t2.6	(LSWE) <i>Corpus</i> (Biber et al. 1999: 433)				
t2.7	<i>Bank of English</i> (Luzon-Marco 1998/99: 91)	~545	~218	~125	–
t2.8	<i>Corpus of Contemporary American English</i>	576	212	172	169
t2.9	(COCA, Davies 2008)				
t2.10	<i>British National Corpus</i> (BYU-BNC,	734	320	173	223
t2.11	Davies 2004)				

t2.12 Note that the frequency counts for these registers are not completely comparable across the four
t2.13 corpora. The count for the spoken register on the basis of the LSWE corpus is given for “conversa-
t2.14 tion”, and the count for fiction provided by Luzon-Marco on the basis of the Bank of English
t2.15 corpus is given for “books”. The counts for the Bank of English corpus are approximations, thus
t2.16 marked by a tilde

409 emphatic positive it contrasts a positive with a corresponding negative proposition
410 that has been expressed or implicated in the preceding discourse. As an emphatic
411 positive it may also occur to indicate the strength of one’s beliefs or feelings.
412 Lambrecht (1994) analyses emphatic *do* as a conventionalized, grammaticalized
413 way of expressing emphasis that involved a gradual loss of the presupposition in
414 three steps: (1) the construction originally required the presupposition that the truth
415 of a proposition was questioned in the immediately preceding discourse (fully contra-
416 stastive contradiction), (2) the presupposition weakened so that a contradiction was
417 merely suggested and left implicit (implicit contradiction), and finally, (3) the pre-
418 supposition disappeared completely with *do* functioning as an intensifier like *really*
419 (non-contrastive emphasis). Nevalainen and Rissanen’s (1986) analysis compared
420 358 instances of emphatic *do* in the London-Lund Corpus (spoken British English)
421 and the Lancaster-Oslo-Bergen Corpus (written British English). Their findings
422 lend support to the view that emphatic *do* can indeed express non-contrastive
423 emphasis. While 63 (18 %) and 101 instances (28 %) in the two corpora signaled
424 either explicit opposition or implicit contrast respectively, a majority of 194
425 instances (54 %) expressed neither opposition nor contrast.

426 Biber et al. (1999: 433) note that “emphatic *do* usually marks a state of affairs in
427 contrast to some other expected state of affairs which is by implication denied”.
428 This contrast can then be explicitly marked by contrastive connectives such as *but*,
429 *however*, *nevertheless* or (*al*)*though*). Similarly, Luzón Marco (1998/99) argues that
430 contrastive and emotive emphasis are not two different functions of emphatic *do*.
431 She suggests that it always implies contrast, concession or correction with regard to
432 something that has been previously said or is supposed to be known, expected or
433 assumed. Moreover, it expresses simultaneously contrastive emphasis and involve-
434 ment (i.e. carries an emotive effect).

435 Emphatic *do* is also characterized by distinct lexical co-occurrence patterns that
436 partially reflect its discourse functions. Contrastive uses are often explicitly marked
437 by contrastive connectives (*but*, *however*, *nevertheless*, [*al*]*though*) as in example (2)

[AU1]

and can also occur in conditional sentences introduced by (*even*) *if*. Contrastive and non-contrastive instances frequently co-occur with intensifying adverbs (*really, certainly, indeed*) and pragmatic markers (*well, yes/yeah, actually, you know, I mean*) as in (3). The types of predicates that are highlighted often include cognition verbs (e.g. *think, know, believe*) and emotive verbs (e.g. *like, hope, feel, need, want*).

(2) er ... you know I I'm I'm not a real big fan of the cinema **but I do think it's a good night out** and I'd much prefer to go to the cinema than to watch er a video <\B> (LOCNEC)

(3) <A> must be quite hard after you you've played something [to to to find yourself back <\A>

 [oh ... it d= **well yeah it it definitely does take a while to come back down** <\B> (LOCNEC)

In the present chapter, the manual qualitative analysis of the discourse functions of emphatic *do* is based on its contextual use and distinguishes between three functions: (1) an intensifying, non-contrastive use (e.g. to indicate the strength of one's beliefs or feelings), and two types of contrastive uses, i.e. (2) explicit contrast/opposition (both referents are explicitly mentioned and contrasted) and (3) implicit contrast (the contrasted referent is not explicitly mentioned but contextually implied, i.e. presupposed, expected or assumed). These three functions are illustrated in example (4).

(4) <A> I mean you're independent here you can do whatever you want to and then [you go back home. <\A>

 [Yes ... mhm. <\B>

<A> How do you feel about that. is it sometimes difficult I mean. you have to to I guess to tell your parents where you're going to if you leave and that kind of thing. <\A>

 Erm ... yeah it it is it is quite. difficult to I suppose it's something I've got used to a lot more **I do I do like going home** it has it has advan= some advantages over being here and being here <\B>

<A> You don't have to cook <laughs> <\A>

 <begin_laughter> **Well I do have to do some cooking** <end_laughter> but <\B>

<A> Yeah I mean but <\A>

 Yeah not so much yeah [so <\B>

<A> [not so much <\A>

 Er ... yeah I I like going home <X> **I do get on with my parents** and they're not they're not very . strict but erm **Yes I d= I do . feel yeah I do have to . tell them . where I'm going** and <\B> (LOCNEC)

The first and the third instance can be classified as cases of implicit contrast. The interviewer (A) does not explicitly deny that the interviewee (B) does not like going home to his/her parents place or does not get on well with them, but this is implicitly questioned ("How do you feel about that. is it sometimes difficult") and subsequently

t3.1 **Table 3** Frequencies of occurrence of emphatic *do* in the three corpora

t3.2	Corpus	Absolute frequency	Normalized frequency per thousand turns
t3.3	LINDSEI-F	8	1.45
t3.4	LINDSEI-G	22	3.64
t3.5	LOCNEC	99	11.74

480 clarified by B (“I do like going home”, “I do get on with my parents”). The second
 481 instance is a case of explicit contrast. A mistakenly presupposes that B does not have
 482 to do any cooking when spending time with his/her parents (“You don’t have to
 483 cook”) which B explicitly corrects (“Well I do have to do some cooking”). Finally,
 484 the fourth instance exemplifies the intensifying, non-contrastive use. B responds to
 485 A’s earlier turn (“you have to to I guess to tell your parents where you’re going to if
 486 you leave and that kind of thing”) and emphasizes the truth of this statement by
 487 confirming it (“I do . feel yeah I do have to . tell them . where I’m going”).

488 They only previous corpus study of emphatic *do* in learner language (Callies
 489 2009a), was based on a subset of the German component of the *International*
 490 *Corpus of Learner English* (ICLE, Granger et al. 2009), a corpus of L2 learners’
 491 argumentative writing. This study found a significant underrepresentation of
 492 emphatic *do* when compared to similar NS writing, differences in contextual use
 493 and lexical co-occurrence patterns and several apparently unmotivated uses. The
 494 much higher frequency of occurrence in speaking and the strong intonational
 495 component of emphatic *do* makes it necessary to replicate this study on the basis of
 496 spoken learner data. On account of the previous research findings and the fact that
 497 French and German lack a clear one-to-one equivalent that expresses the functions
 498 of emphatic *do* in English, emphatic *do* is hypothesized to be underrepresented in
 499 both spoken learner corpora when compared to native speaker usage. In French and
 500 German the functions of emphatic *do* are often fulfilled by modal particles like *doch*
 501 or *schon* (in German) and *si* (in French) (König et al. 1990; Lambrecht 1994: 72),
 502 both of which can be translated as ‘but’.

503 The quantitative analysis of the frequency of occurrence of emphatic *do* in the
 504 three corpora (Table 3) confirms the hypothesis and shows that *do* as a marker of
 505 emphasis is significantly underrepresented in the two learner corpora when com-
 506 pared to the native speaker corpus (LOCNEC vs. LINDSEI-F: Log Likelihood
 507 (LL)= -57.4***; LOCNEC vs. LINDSEI-G: LL= -30.7***). In particular, with
 508 only eight occurrences in total, it is largely absent in the LINDSEI-F.

509 When analyzing the use of emphatic *do* by individual learners (Figs. 1 and 2) it
 510 is striking that it is only very few learners who use it. In particular, in the LINDSEI-G
 511 there is a fairly uneven distribution with two learners (ge024 and ge034) producing
 512 40 % of all instances (9 out of 22) whereas the majority of learners do not use
 513 emphatic *do* at all.

514 The comparative analysis of the discourse functions of emphatic *do* does not
 515 reveal any major differences between the corpora: it is mostly used to express con-
 516 trast by all three groups. Native speakers and German learners show a fairly balanced
 517 distribution of the three functions (see Fig. 3). More interesting, however, is the
 518 qualitative analysis of the most frequent collocates and verbs that co-occur with

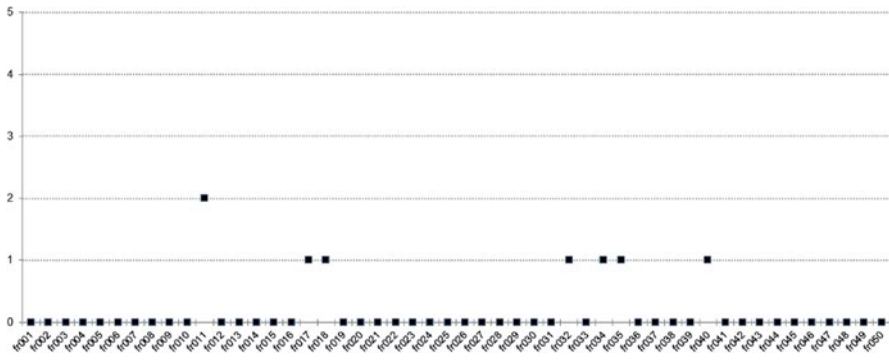


Fig. 1 Distribution of emphatic *do* in the LINDSEI-F

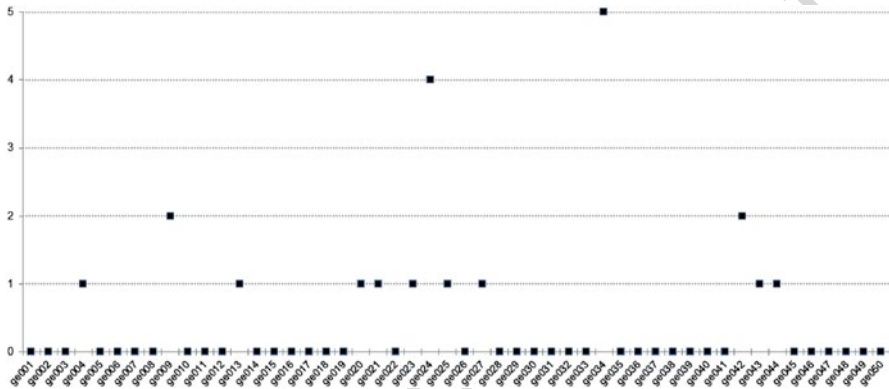


Fig. 2 Distribution of emphatic *do* in the LINDSEI-G

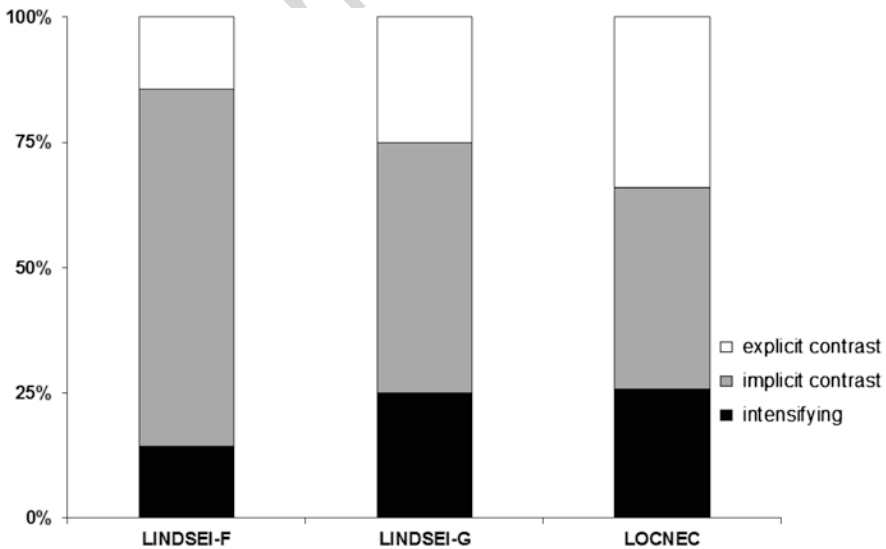


Fig. 3 Discourse functions of emphatic *do* in the three corpora

this figure will be printed in b/w

this figure will be printed in b/w

t4.1 **Table 4** Most frequent collocates and verbs occurring with emphatic *do* in the three corpora

t4.2			All verbs	All verbs		Most freq. verbs		
t4.3	Corpus	Collocate	N	(tokens)	(types)	TTR	(N ≥ 3)	N
t4.4	LINDSEI-F	<i>but</i>	4	8	6	0.75	–	–
t4.5	LINDSEI-G	<i>but</i>	6	22	16	0.72	<i>have</i>	5
t4.6		<i>yes, yeah</i>	4				<i>like</i>	3
t4.7	LOCNEC	<i>but</i>	24	99	48	0.48	<i>have (to)</i>	13
t4.8		<i>yes, yeah</i>	19				<i>like</i>	11
t4.9		<i>I mean</i>	8				<i>look</i>	8
t4.10		<i>so</i>	8				<i>get</i>	5
t4.11		<i>actually</i>	5				<i>think (about), work</i>	4 each
t4.12		<i>well</i>	4				<i>feel, go, know, miss</i>	3 each
t4.13		<i>if</i>	4					

519 emphatic *do*. It is striking that emphatic *do* is not only significantly underrepresented
 520 in the two learner corpora, but also that the few instances that can be found do not
 521 occur in their typical lexical co-occurrence patterns (contrastive connectives, inten-
 522 sifying adverbs, pragmatic markers, cognition verbs and emotive verbs, see Table 4).

523 How can the differences between native speakers and learners, and the differ-
 524 ences between the two learner groups be explained? Considering recent findings
 525 that even advanced L2 learners have only a limited awareness of the appropriate use
 526 of lexical and syntactic focusing devices in formal and informal registers (Callies
 527 2009a), the results are not surprising. Moreover, linguistic structures that are
 528 optional and subject to discourse-motivated preferences are assumed to be among
 529 the most difficult to acquire in advanced SLA (DeKeyser 2005). One explanation to
 530 account for the differences between the German and the French EFL learners could
 531 be that the German learners are benefitting from positive L1-transfer. In Standard
 532 German, the insertion of the semantically empty verb *tun* ('do') is obligatory in
 533 contexts where a lexical verb is topicalized and no other verb (auxiliary or modal)
 534 is present (Duden 1997: 726), see example (5a).

535 (5a) Tanzen **tut** Katja immer noch häufig.
 536 Dance does Katja always still often.
 537 'Katja does still dance often.'

538 *Do*-insertion is also frequently used in colloquial German and some German dialects
 539 to mark progressive aspect, see example (5b).

540 (5b) Sie **tut** gerade schreiben.
 541 She does just now write
 542 'She is writing just now.'

543 While another reason for why the Germans differ from the French learners may
 544 simply be differences in their general level of proficiency (see Sect. 3.3 for more
 545 explanation), further evidence for the influence of the learners' native language,
 546 possibly even in terms of a typological parameter, is suggested by the results of

preliminary analyses of other LINDSEI subcorpora: learners whose L1 is a (Germanic) language that has *do*-support seem to use emphatic *do* more often than learners from other L1 backgrounds (Callies in preparation).

The significantly lower frequency counts in the learner data may, however, also be an effect of the task and/or the interlocutor. It is a well-known fact that interlanguage variation is influenced by a number of external sociolinguistic factors that have to do with the situational context of language use, e.g. task, topic and interlocutor (see e.g. Ellis 2008: 141ff.). It is thus possible that L2 learners may be less inclined to disagree or object (hence experience much less need to make use of the linguistic means that convey contrastive emphasis) when they are interviewed by a native speaker who is of the opposite sex and not familiar to them rather than when interviewed by a same-sex non-native speaker who they know. Although variables such as the interviewer's mother tongue, gender and distance/closeness to the interviewee have been recorded in the LINDSEI, their influence cannot (yet) be assessed on a broad basis because of the small corpus size: strict control of all the relevant variables results in a very small database of sometimes only a handful of interviews.

3.3 *Demonstrative Clefts* 563

Cleft sentences are information packaging constructions that involve the splitting of a sentence into two clauses. They are pragmatically motivated and differ from their basic counterparts in that they serve to highlight a certain phrase or clause, the cleft constituent. The most common types are *it*-clefts and *wh*-clefts (also known as pseudo-clefts). There are also other types of cleft constructions one of which is the reverse *wh*-cleft, in which the order of *wh*- and cleft-clause is inverted. The vast majority of reverse *wh*-clefts feature the non-contrastive, non-focal deictic demonstratives *that* or *this* as the cleft constituent, see examples (6) and (7),¹¹ and therefore this type is also referred to as demonstrative cleft in the literature (Biber et al. 1999: 961; Calude 2008, 2009).

- (6) <A> so you you did English and ling= and linguistics to: <\A> 574
 - I did English and linguistics just because **that was what I was** 575
 - interested in** the the interest in going into film industry has only devel- 576
 - oped since I've been at university <\B> (LOCNEC) 577
- (7) <A> so you had to cope with those kids <\A> 578
 - I had to cope with those kids completely on my own with no back-up she 579
 - said you know she w= she thought it was great having someone to help she 580
 - said right you're gonna take half the kids ...the worst half and you're going 581
 - to teach them the same lesson as I'm teaching them here's the book **this is** 582
 - what I want you to teach them** go off and do it for a year <\B> (LOCNEC) 583

¹¹Demonstrative clefts are given in bold print.

584 When compared to other types of cleft constructions, demonstrative clefts only
 585 rarely occur in written language but are clearly the most frequent variant in the spo-
 586 ken mode (Collins 1991: 178ff.; Oberlander and Delin 1996: 186; Weinert and
 587 Miller 1996: 176), occurring especially often in spontaneous spoken language, i.e.
 588 conversation (Biber et al. 1999: 961; Calude 2008: 86). Of the two demonstratives,
 589 *that* is much more frequent than *this* (Oberlander and Delin 1996: 189; Weinert and
 590 Miller 1996: 188; Biber et al. 1999: 962; Calude 2008: 79). Therefore, the majority
 591 of demonstrative clefts convey anaphoric deixis as in example (8),¹² but they can
 592 also express cataphoric deixis as in (9), function anaphorically and cataphorically
 593 simultaneously as in (10), or carry exophoric deixis, i.e. non-textual, extra-linguistic
 594 reference either in the form of shared world knowledge or physical/visual presence
 595 at the time of utterance, see example (11) (Calude 2008: 87ff.).

- 596 (8) <A> so what are you doing now as a major is it linguistics or is it <A>
 597 <X> ... I I thought I'd been accepted for Chinese and linguistics com-
 598 bined
 599 <A> [mm <A>
 600 [and **that's what they told me when I first . came here** but now they
 601 seem to think it's only linguistics (LOCNEC)
- 602 (9) that we're living I mean I had my had my own flat and it's very difficult
 603 to: go from having your own flat and [<X> privacy to
 604 <A> [and share a kitchen <A>
 605 living in somewhere much smaller
 606 <A> mhm <A>
 607 but erm
 608 <A> but I mean Graduate College is quite okay <A>
 609 yeah I know **that's why I decided to pay a bit more** cos I thought
 610 sharing a kitchen and a bathroom with ten people
 611 <A> yeah <A>
 612 [I just couldn't
 613 <A> [especially the bathroom <A>
 614 yeah no I I really couldn't have faced that (LOCNEC)
- 615 (10) <A> and you don't live there and you you've never seen something like that
 616 before ... but you you live in Sheffield <A>
 617 yeah
 618 <A> it's quite a big city isn't it <A>
 619 it is quite big yeah that's why I came here cos I wanted to come
 620 to somewhere smaller (LOCNEC)
- 621 (11) and she doesn't . it's not really a glamorous picture
 622 <A> mhm <A>
 623 or anything like that ... erm the third one it looks like he's painted
 624 it again ... erm ... new hairstyle ... smiling sat up ... it makes her look
 625 more beautiful than she is

¹²The discourse segment(s) that the demonstrative *that* refers to are underlined.

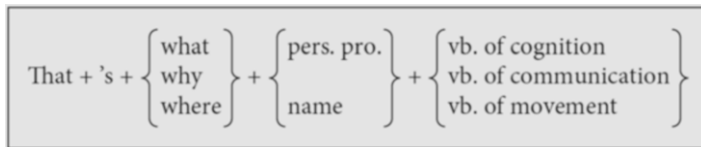


Fig. 4 The formulaic nature of demonstrative clefts (Reproduced from Calude 2009: 69)

- <A> mhm <A> 626
- <laughs> and in the fourth one she's telling all her friends of 627
that's me **that's how I look** ... things like that (LOCNEC) 628

In view of their relatively fixed structure, Calude (2009) argues that demonstra- 629
tive clefts show characteristics of formulaic expressions, allowing only a narrow 630
range of elements to occur in its structural “slots” (see Fig. 4). Prototypically, the 631
demonstrative *that* occurs as the initial element. The copula *be* only occurs in simple 632
present and simple past tense and is most commonly used in its contracted 633
form *'s*. The copula is then most frequently followed by *what*, less frequently by 634
why, *where*, *when* and *how* as *wh*-words in the cleft clause (Collins 1991: 28; 635
Oberlander and Delin 1996: 187; Weinert and Miller 1996: 188). Moreover, demonstra- 636
tive clefts have a distinct function in discourse as organizational and discourse- 637
managing markers, and are typical of a specific register, i.e. conversation.¹³ 638

Demonstrative clefts have multiple functions as to discourse organization and 639
management. In particular, what sets them apart from other cleft types is their point- 640
ing function by means of the initial demonstrative pronoun (Weinert and Miller 641
1996: 188; Oberlander and Delin 1996: 189). They typically have extended text 642
reference that spans over three or more turns prior to the cleft (Calude 2008: 79f.). 643
With *that* as the initial element, demonstrative clefts have a strong anaphoric and 644
attention-marking function (Weinert and Miller 1996: 192f.) and are typically used 645
to underline or sum up previous discourse or to make reference to what has been 646
said before (Collins 1991: 145f.; Weinert and Miller 1996: 192f.; Biber et al. 1999: 647
961ff.), while those introduced by *this* have a forward-pointing function and are also 648
used as an attention marker (Weinert 1995). 649

Calude (2008: 99ff.; 108) suggests four discourse functions of demonstrative 650
clefts. For the qualitative analysis of the discourse functions in the present case 651
study, her taxonomy was adopted with slight modifications and two more functions 652
(summarizing and projecting) were added. The six functions are exemplified in turn 653
in (12)–(17). 654

- (12) **quoting**: signaling direct speech, indirect speech or self-reported thought 655
- erm and I I wanted to come to university and do literature <XXX> 656
interested<?> in that ... and it was only really when I was looking 657
through the prospectus sort of thinking well I don't just want to do lit- 658
erature what can I put [with it 659

¹³One may add here that another feature that adds to their formulaicity is that in contrast to other types of clefts, demonstrative clefts are not reversible (Biber et al. 1999: 961).

- 660 <A> [mhm mhm <\A>
 661 I sort of discovered the linguistics department and thought ... ah
 662 yeah **that's what I've always wanted to do** <\B> (LOCNEC)
- 663 (13) **explaining**: giving a reason for a point previously made; explaining how two
 664 prior utterances relate to each other (linking function)
 665 yeah I think geography is interesting **that's why I study it**
 666 <laughs> (LINDSEI-G)
- 667 (14) **evaluating**: giving opinions, evaluations or assessments; expressing agree-
 668 ment, disagreement or a neutral opinion with a previous comment
 669 yeah it wasn't much of a holiday really <\B>
 670 <A> oh no <laughs> <\A>
 671 <laughs> <\B>
 672 <A> it was just a a working holiday <X> <\A>
 673 a working holiday yeah <\B>
 674 <A> just work <\A>
 675 well that's that's <X> **that's exactly what what our bosses were**
 676 **saying** exactly the same phrase said er you're here for no holiday you
 677 work you're here to work <\B> (LOCNEC)
- 678 (15) **highlighting**: singling out a preceding discourse element, thereby foreground-
 679 ing it and giving it special prominence
 680 <A> since you like the cinema so much <\A>
 681 [mhm <\B>
 682 <A> [would you like to: to do: ... later to work . in relation . to <\A>
 683 <X> what I'd like to do well I mean my degree is a primary school teach-
 684 ing degree **that's what I'm aiming to do at the[i:] end** <\B> (LOCNEC)
- 685 (16) **summarizing**: summing up a longer stretch of previous discourse
 686 he's changed the picture so that she's erm she looks considerably
 687 younger ... erm obviously the hair's changed the face has changed <\B>
 688 <A> [mhm <\A>
 689 [she's she's got a slight smile erm ... and then now she's sort of
 690 erm just telling all her all of her friends sort of oh this is a picture of me
 691 isn't it lovely and doesn't it look so much like me but er \B>
 692 <A> <laughs> <\A>
 693 **that's that's how I would say the story is going** she's er ... she's
 694 she's eh this woman is actually quite vain <\B> (LOCNEC)

¹⁴This function is in line with Weinert's (1995) analysis of demonstrative clefts introduced by *this* as forward-pointing and attention marking devices. It is usually demonstrative clefts that have cataphoric deixis that can be said to have a projecting function. In general, the development of cleft constructions in spoken English is strongly related to their discourse-pragmatic functions (see e.g. Callies 2012a for a study of the pragmatization of *wh*-clefts). For example, *wh*-clefts have been analysed as projector constructions that foreshadow upcoming discourse (e.g. Hopper and Thompson 2008) in which the *wh*-clause opens a projection span that draws the recipient's attention to the following highlighted constituent.

Table 5 Frequencies of occurrence of demonstrative clefts in the three corpora t5.1

Corpus	Absolute frequency	Normalized frequency per thousand turns	t5.2
LINDSEI-F	27	4.72	t5.3
LINDSEI-G	57	9.42	t5.4
LOCNEC	73	8.65	t5.5

(17) **projecting**: drawing attention to a following stretch of discourse (only with cataphoric deixis)¹⁴ 695
 so . it was a really nice (erm) .experience . I had and . what I found most 696
 (erm) impressive and I think **that's what everybody says when . he has seen** 697
Australia is that . (erm) the distances are so huge . it's (er) that's really amazing so 698
 one day we drove for twelve hours and there was nothing . li<?> (eh) it's only dust 699
 . around us and so . but . it was really . yes impressive (LINDSEI-G) 700
 701

Previous corpus-based studies of reversed *wh*-clefts in learner language are based on 702
 subsets of the ICLE. While Herriman and Boström Aronsson (2009) found an over- 703
 representation of reversed *wh*-clefts in the writing of Swedish EFL learners when 704
 compared to native speaker writing (93 vs. 62 instances), Callies (2009a) noted that 705
 native speakers used demonstrative clefts slightly more often when compared to the 706
 writing of German EFL learners (27 vs. 19 instances, but not statistically significant 707
 difference). Moreover, Callies observed that the learners showed little variation in 708
 how they used this construction: *what* was by far the most commonly used *wh*-word 709
 in reversed *wh*-clefts by both groups of writers, but the native speakers employed a 710
 broader range of *wh*-elements, while *how*, *where*, and *when* were completely absent 711
 from the learner data. They also strongly preferred *that* as a deictic marker and used 712
 the copula almost exclusively in its contracted form 's, which may indicate that the 713
 learners saw this as a formulaic expression. Non-deictic elements in reversed *wh*- 714
 clefts (e.g. *Music is what I like most*) were exclusively used by native speakers. 715

In view of these previous research findings and a contrastive analysis of such 716
 cleft types in French, German and English (see further below), the following two 717
 working hypotheses can be put forward for the case study: (1) demonstrative clefts 718
 are underrepresented in both learner corpora when compared to native speaker 719
 usage, and (2) advanced learner language is characterized by a narrower range of 720
 the formal and functional uses of this construction. 721

In fact, the quantitative analysis of the frequency of occurrence of demonstrative 722
 clefts in the three corpora (Table 5) shows that demonstrative clefts are significantly 723
 underrepresented in the LINDSEI-F when compared to the LOCNEC (LL= -7.7**), 724
 but that there is no statistically significant difference between the LINDSEI-G and 725
 the LOCNEC (LL= +0.23). Similar to emphatic *do*, the distribution of demonstra- 726
 tive clefts in the two learner corpora shows a high degree of inter-learner variability. 727
 In both corpora, it is merely a handful of learners who provide for almost 50 % of 728
 all tokens whereas half (or more) of the learners do not use this construction at all 729
 (see Figs. 5 and 6). 730

It is interesting to compare the two learner groups and the native speakers as to 731
 the relatively fixed structure of demonstrative clefts. Similar to the findings reported 732
 in the research literature, the deictic *that* and the *wh*-words *what* and *why* are the 733

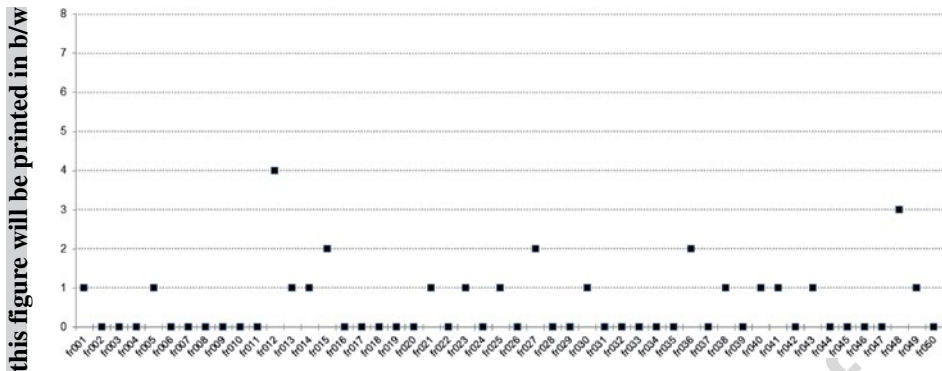


Fig. 5 Distribution of demonstrative clefts in the LINDSEI-F

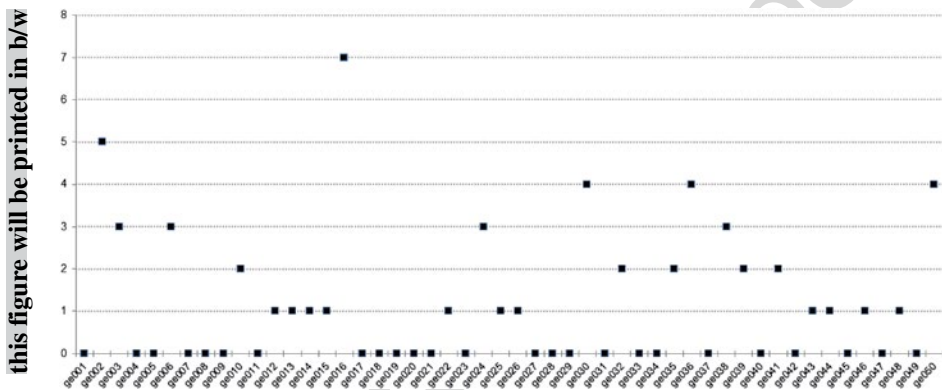


Fig. 6 Distribution of demonstrative clefts in the LINDSEI-G

734 most frequently occurring elements (Table 6). Demonstrative clefts primarily convey
 735 anaphoric deixis in all three corpora. While it is not surprising that the native speakers
 736 employ the full range of options that this construction allows in terms of the use of
 737 initial demonstratives, *wh*-words and deictic reference, it is indeed striking to see
 738 major differences between the two learner groups. The way how the German learners
 739 use this construction very much resembles native speaker usage in terms of struc-
 740 tural variation. By contrast, demonstrative clefts are not only significantly under-
 741 represented in the spoken language of French learners, but the degree of formulaicity
 742 (or invariability) is also highest in the LINDSEI-F.

743 A similar picture emerges when analyzing the discourse functions: the native
 744 speakers and the German learners use all six functions, but only four different ones
 745 occur in the LINDSEI-F (Fig. 7).

746 In this case, it is unlikely that the observed differences between native speakers
 747 and learners as well as the differences between the two learner groups are due to

Table 6 Use of demonstratives, *wh*-words and deictic reference in the three corpora

	LINDSEI-F	LINDSEI-G	LOCNEC
demonstrative			
<i>that</i>	26 (96 %)	44 (77 %)	67 (92 %)
<i>this</i>	1 (4 %)	13 (23 %)	6 (8 %)
<i>wh</i>-word			
<i>what</i>	12 (44 %)	27 (47 %)	30 (41 %)
<i>why</i>	14 (52 %)	17 (30 %)	15 (21 %)
<i>where</i>	0	1 (2 %)	11 (15 %)
<i>when</i>	0	4 (7 %)	6 (8 %)
<i>how</i>	1 (4 %)	8 (14 %)	11 (15 %)
deixis			
anaphoric	26 (96 %)	42 (74 %)	57 (78 %)
cataphoric	0	5 (9 %)	4 (5 %)
both	1 (4 %)	4 (7 %)	6 (8 %)
exophoric	0	6 (11 %)	6 (8 %)

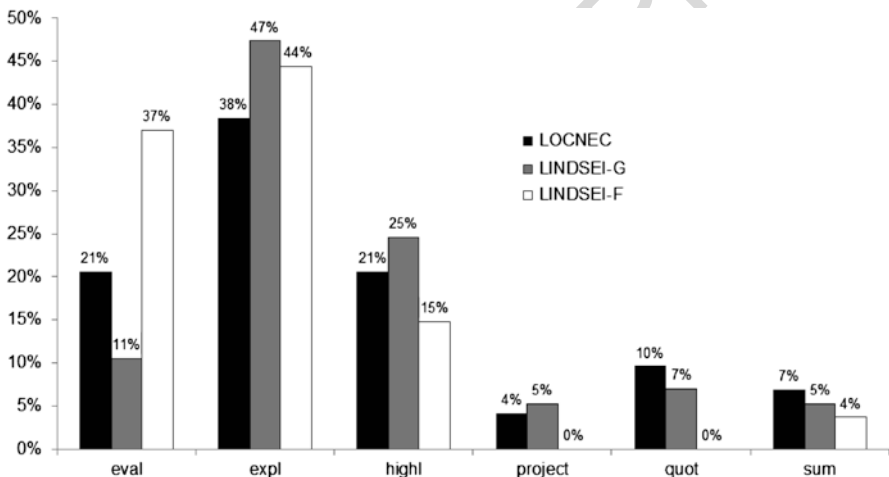


Fig. 7 Functions of demonstrative clefts in the three corpora

cross-linguistic influence, at least as far as the German learners are concerned. 748
 Although German does have cleft constructions, they are dispreferred options to 749
 convey focus and have only peripheral status because of the less restricted use of 750
 topicalization (see e.g. Weinert 1995 and Callies 2009a for discussion). Weinert 751
 (1995) compared *wh*- and reversed *wh*-clefts in English and German, contrasting 752
 their discourse functions with those of preposing/topicalization based on corpora of 753
 structured dialogue and conversation. Her findings showed that in contrast to speakers 754
 of English, Germans used only very few reversed *wh*-clefts because reversed clefts 755
 are extremely rare in German, structurally and functionally more restricted, and 756

757 often combine with focus or modal particles to supplement their focus, and thus
758 create an even stronger focus than their English counterparts (Weinert 1995: 355).
759 Moreover, topicalization in German is less restricted and not as strongly associated
760 with contrastiveness as preposing in English. On account of this, demonstrative
761 clefts should be expected to be underrepresented in LINDSEI-G, but this is clearly
762 not the case.

763 Transfer in the form of underproduction may be an explanatory factor in the case
764 of the French learners. French does have two types of clefts, the *c'est*-cleft, which
765 often carries a contrastive and even exclusive value, and the *il y a*-cleft, which has
766 presentational character, but in contrast to German and English, French does not
767 have reversed *wh*-clefts because it does not allow pre-verbal focus (Lambrecht
768 2001: 492; Miller 2006: 185). The absence of this cleft type in the L1 may thus at
769 least partially explain the observed underrepresentation.

770 It seems more likely that differences in general language proficiency may help
771 explain the differences between the two learner groups. The assessment of language
772 proficiency is a notoriously difficult (and also frequently neglected and underesti-
773 mated) challenge in SLA and Learner Corpus Research (LCR).¹⁵ In LCR, learners'
774 proficiency level has been a fuzzy variable in that it has often been assessed globally
775 by means of external criteria, most typically learner-centered criteria (e.g. Carlsen
776 2012). There are several problems connected with this practice (Thomas 1994,
777 2006). As a consequence, in some corpora learners' proficiency level varies consid-
778 erably, both across and within subcorpora. This is also true for the LINDSEI, in the
779 compilation of which proficiency was assessed globally on account of institutional
780 status with learners being described as "university undergraduates in English (usu-
781 ally in their third or fourth year)" (Gilquin et al. 2010: 10). The proficiency level of
782 learners who are represented in the LINDSEI in fact ranges from higher intermedi-
783 ate to advanced. While some LINDSEI subcorpora predominantly seem to include
784 learners from either the C1 or C2 proficiency levels of the *Common European*
785 *Framework of Reference for Languages*, e.g. Dutch, Swedish or German learners,
786 others rather seem to include learners from higher intermediate (or lower) profi-
787 ciency levels, e.g. those whose L1 is Italian, Spanish or French (Gilquin et al. 2010:
788 10f.). The LINDSEI handbook also provides information about two variables that
789 have often been used to help operationalize proficiency: the amount of formal class-
790 room instruction in the foreign language and time spent in a country where the tar-
791 get language is spoken. Comparing these two variables, it turns out that the number
792 of years spent learning English in school and university is 4.6 and 3.8 on average in
793 LINDSEI-F, while the German learners spent 8.6 and 3.6 years learning English.
794 Thus, the Germans spent significantly more time learning English in school (they
795 are also on average 2 years older than the French: 24.6 vs. 22.1 years). More impor-
796 tant, though, is the difference in the time spent abroad: on average, speakers in
797 LINDSEI-F spent only 1.9 months in an English-speaking country, while those in
798 LINDSEI-G spent 9.3 months abroad (Gilquin et al. 2010: 40f.).

¹⁵It is not possible to go into detail here, but see Callies, Zaytseva & Present-Thomas (to appear) for further discussion as to the operationalization and assessment of (advanced) proficiency in LCR.

4 Conclusion

799

This chapter has provided a critical assessment of research on pragmatics in the context of SLA showing that in mainstream ILP, the significance of L2 pragmatic knowledge beyond the domain of speech acts has been neglected to date. I have argued that the field of inquiry in ILP needs to be extended because pragmatic knowledge in an L2 includes more than sociopragmatic and pragmalinguistic abilities for understanding and performing speech acts. I have proposed a wider definition of L2 pragmatic knowledge and have highlighted the crucial role of learner corpora in the expansion of the narrow research agenda of ILP. Two case studies of EFL learners' use of emphatic *do* and demonstrative clefts have exemplified how spoken learner corpora enable researchers to study a much broader range of different pragmatic phenomena and can help overcome several problems and limitations posed by the dominance of data elicitation techniques in ILP to date.

The case studies have demonstrated the usefulness of corpora to abstract away from individual learners to identify a corpus-based description of a specific learner group while also providing insights into inter-learner variability. The individual differences found for both the French and the German EFL learners have important implications for learner corpus analysis and compilation in that they confirm that global proficiency measures based on external criteria alone are not reliable indicators of proficiency. However, in a substantial part of LCR to date individual differences often go unnoticed or tend to be disregarded and are thus not reported in favour of (possibly skewed) average frequency counts. Mukherjee (2009) is one study where the issue of inter-learner variability is explicitly addressed. Observing an extremely uneven distribution of the pragmatic marker *you know* in the LINDSEI-G, Mukherjee concludes that "the fiction of homogeneity that is often associated with the compilation of a learner corpus according to well-defined standards and design criteria may run counter to the wide range of differing individual levels of competence in the corpus" (2009: 216).

[AU2] References

827

- Aijmer, Karin. 2004. Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies* 3(1): 173–190.
- Aijmer, Karin. 2009. 'So er I just sort I dunno I think it's just because...': A corpus study of *I don't know* and *dunno* in learners' spoken English. In *Corpora: Pragmatics and discourse*, ed. Andreas H. Jucker, Daniel Schreier, and Marrienne Hundt, 151–168. Amsterdam: Rodopi.
- Aijmer, Karin. 2011. *Well I'm not sure I think...* The use of *well* by non-native speakers. *International Journal of Corpus Linguistics* 16(2): 231–254.
- [AU3] Aijmer, Karin, and Christoph Rühlemann. forthcoming. *Corpus pragmatics. Exploring speaker meaning in computerized corpora*. Cambridge: Cambridge University Press.
- Andersen, Gisle. 2011. Corpus-based pragmatics I: Qualitative studies. In *Foundations of pragmatics*, Handbooks of pragmatics, vol. 1, ed. Wolfram Bublitz and Neil R. Norrick, 587–627. Berlin: DeGruyter Mouton.
- Archibald, Alisdair N. 1994. *The acquisition of discourse proficiency*. Frankfurt/Main: Peter Lang.

- 841 Bachman, Lyle. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University
842 Press.
- 843 Bardovi-Harlig, Kathleen. 2010. Exploring the pragmatics of interlanguage pragmatics: Definition
844 by design. In *Pragmatics across languages and cultures*, Handbooks of pragmatics, vol. 7, ed.
845 Anna Trosborg, 219–259. Berlin: DeGruyter Mouton.
- 846 Barron, Anne. 2003. *Acquisition in interlanguage pragmatics*. Amsterdam: John Benjamins.
- 847 Belz, Julie A., and Nina Vyatkin. 2005. Learner corpus analysis and the development of L2 prag-
848 matic competence in networked intercultural language study: The case of German modal partic-
849 les. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*
850 62(1): 17–48.
- 851 Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999.
852 *Longman grammar of spoken and written English*. Harlow: Longman.
- 853 Boström Aronsson, Mia. 2003. On clefts and information structure in Swedish EFL writing.
854 In *Extending the scope of corpus-based research. New applications, new challenges*, ed.
855 Granger Sylviane and Petch-Tyson Stephanie, 197–210. Amsterdam: Rodopi.
- 856 Brand, Christiane, and Sandra Götz. 2011. Fluency versus accuracy in advanced spoken learner lan-
857 guage: A multi-method approach. *International Journal of Corpus Linguistics* 16(2): 255–275.
- 858 Bußmann, Hadumod. 1996. *Routledge dictionary of language and linguistics*. London: Routledge.
- 859 Callies, Marcus. 2008a. Argument realization and information packaging in tough-movement con-
860 structions – A learner-corpus-based investigation. In *Morphosyntactic issues in second lan-
861 guage acquisition studies*, ed. Danuta Gabrys-Barker, 29–46. Clevedon: Multilingual Matters.
- 862 Callies, Marcus. 2008b. Easy to understand but difficult to use? Raising constructions and information
863 packaging in the advanced learner variety. In *Linking contrastive and learner corpus research*,
864 ed. Gaëtanelle Gilquin, Maria Belen Diez Bedmar, and Szilvia Papp, 201–226. Amsterdam:
865 Rodopi.
- 866 Callies, Marcus. 2009a. *Information highlighting in advanced learner English. The syntax-
867 pragmatics interface in second language acquisition*. Amsterdam: John Benjamins.
- 868 Callies, Marcus. 2009b. ‘What is even more alarming is...’ – A contrastive learner-corpus study of
869 what-clefts in advanced German and Polish L2 writing. In *On language structure, acquisition
870 and teaching. Studies in honour of Janusz Arabski on the occasion of his 70th Birthday*, ed.
871 Maria Wysocka, 283–292. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- 872 Callies, Marcus. 2012. The grammaticalization and pragmaticalization of cleft constructions in
873 present-day English. In *English corpus linguistics: Looking back, moving forward*, ed.
874 Sebastian Hoffmann, Paul Rayson, and Geoffrey Leech, 5–21. Amsterdam: Rodopi.
- 875 Callies, Marcus. 2012b. Discourse completion task. In *Theories and methods in linguistics*,
876 Wörterbücher zur Sprach- und Kommunikationswissenschaft [WSK] online, vol. 11, ed. Bernd
877 Kortmann. Berlin: De Gruyter Mouton. doi: [10.1515/wsk.35.0.discoursecompletiontask](https://doi.org/10.1515/wsk.35.0.discoursecompletiontask)
- 878 Callies, Marcus. in preparation. Emphatic *do* in advanced learner English. A contrastive interlan-
879 guage analysis of spoken and written corpora.
- 880 Callies, Marcus, Ekaterina Zaytseva, and Rebecca Present-Thomas. forthcoming. Writing assess-
881 ment in higher education: Making the framework work. *Dutch Journal of Applied Linguistics*.
- 882 Calude, Andreea. 2008. Demonstrative clefts and double cleft constructions in spoken English.
883 *Studia Linguistica* 62(1): 78–118.
- 884 Calude, Andreea. 2009. Formulaic tendencies of demonstrative clefts in spoken English. In
885 *Formulaic language, Vol. 1: Distribution and historical change*, ed. Roberta Corrigan, Edith
886 A. Moravcsik, Hamid Quali, and Kathleen M. Wheatley, 55–76. Amsterdam: John Benjamins.
- 887 Canale, Michael. 1983. From communicative competence to communicative language pedagogy.
888 In *Language and communication*, ed. Jack C. Richards and Richard W. Schmidt, 2–27. London/
889 New York: Longman.
- 890 Canale, Michael, and Merrill Swain. 1980. Theoretical bases of communicative approaches to sec-
891 ond language teaching and testing. *Applied Linguistics* 1(1): 1–47.
- 892 Carlsen, Cecilie. 2012. Proficiency level – A fuzzy variable in computer learner corpora. *Applied
893 Linguistics* 33(2): 161–183.
- 894 Collins, Peter C. 1991. *Cleft and pseudo-cleft constructions in English*. London: Routledge.
- 895 Crystal, David. 2003. *A dictionary of linguistics and phonetics*, 5th ed. Malden: Blackwell.

- Davies, Mark. 2004. BYU-BNC (based on the British National Corpus from Oxford University Press). <http://corpus.byu.edu/bnc/>. Accessed 18 Dec 2012. 896
897
- Davies, Mark. 2008. The corpus of contemporary American English: 450 million words, 1990–present. <http://corpus.byu.edu/coca/>. Accessed 18 Dec 2012. 898
899
- DeKeyser, Robert M. 2005. What makes learning second language grammar difficult? A review of issues. *Language Learning* 55: 1–25. 900
901
- Deppermann, Arnulf. 2011. Pragmatics and grammar. In *Foundations of pragmatics*, Handbooks of pragmatics, vol. 1, ed. Wolfram Bublitz and Neil R. Norrick, 425–460. Berlin: DeGruyter Mouton. 902
903
904
- Dippold, Doris. 2009. Face and self-presentation in spoken L2 discourse: Renewing the research agenda in interlanguage pragmatics. *Intercultural Pragmatics* 6(1): 1–28. 905
906
- Duden. 1997. Richtiges und gutes Deutsch: Wörterbuch der sprachlichen Zweifelsfälle, (Duden vol. 9). Mannheim: Dudenverlag. 907
908
- Ellis, Rod. 2008. *The study of second language acquisition*, 2nd ed. Oxford: Oxford University Press. 909
910
- Felder, Ekkehard, Marcus Müller, and Friedemann Vogel (eds.). 2011. *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin: DeGruyter. 911
912
- Gass, Susan M., and Larry Selinker. 2008. *Second language acquisition. An introductory course*, 2nd ed. Mahwah: Erlbaum. 913
914
- Gilquin, Gaëtanelle. 2008. Hesitation markers among EFL learners: Pragmatic deficiency or difference? In *Pragmatics and corpus linguistics. A mutualistic entente*, ed. Jesus Romero Trillo, 119–149. Berlin: Mouton de Gruyter. 915
916
917
- Gilquin, Gaëtanelle, De.Cock, Sylvie, and Granger Sylviane. 2010. *The Louvain international database of spoken English interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain. 918
919
920
- Golato, Andrea. 2003. Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics* 24(1): 90–121. 921
922
- Götz, Sandra. 2007. Performanzphänomene in gesprochenem Lernerenglisch: eine korpusbasierte Pilotstudie. *Zeitschrift für Fremdsprachenforschung* 18(1): 67–84. 923
924
- Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins. 925
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. 2009. The international corpus of learner English. Version 2. Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain. 926
927
928
- Herriman, Jennifer, and Mia Boström Aronsson. 2009. Themes in Swedish advanced learners' writing in English. In *Corpora and language teaching*, ed. Karin Aijmer, 101–120. Amsterdam: John Benjamins. 929
930
931
- Hopper, Paul, and Sandra Thompson. 2008. Projectability and clause combining in interaction. In *Crosslinguistic studies of clause combining. The multifunctionality of conjunctions*, ed. Ritva Laury, 99–123. Amsterdam: John Benjamins. 932
933
934
- Huddleston, Rodney. 2002. The verb. In *The Cambridge grammar of the English language*, ed. Rodney Huddleston and Geoffrey K. Pullum, 71–212. Cambridge: Cambridge University Press. 935
936
937
- Hymes, Dell H. 1972. On communicative competence. In *Sociolinguistics. Selected readings*, ed. John B. Pride and Janet Holmes, 269–293. Harmondsworth: Penguin. 938
939
- Kasper, Gabriele. 2008. Data collection in pragmatics research. In *Culture, communication and politeness theory*, 2nd ed, ed. Helen Spencer-Oatey, 279–303. London/New York: Continuum. 940
941
- Kasper, Gabriele. 2010. Interlanguage pragmatics. In *Variation and change. Pragmatic perspectives*, Handbook of pragmatics highlights, vol. 6, ed. Mirjam Fried, Jan-Ola Östman, and Jef Verschueren, 141–154. Amsterdam: John Benjamins. 942
943
944
- Kasper, Gabriele, and Merete Dahl. 1991. Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition* 13: 215–247. 945
946
- Kasper, Gabriele, and Kenneth R. Rose. 1999. Pragmatics and SLA. *Annual Review of Applied Linguistics* 19: 81–104. 947
948
- Kasper, Gabriele, and Kenneth R. Rose. 2002. *Pragmatic development in a second language*. Oxford: Blackwell. 949
950

- 951 König, Ekkehard, Detlef Stark, and Susanne Requardt. 1990. *Adverbien und Partikeln: ein*
 952 *deutsch-englisches Wörterbuch*. Heidelberg: Groos.
- 953 Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge
 954 University Press.
- 955 Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3):
 956 463–516.
- 957 Leech, Geoffrey. 1983. *Principles of pragmatics*. London: Longman.
- 958 LoCastro, Virginia. 2011. Second language pragmatics. In *Handbook of research in second lan-*
 959 *guage teaching and learning*, vol. 2, ed. Eli Hinkel, 319–344. New York: Routledge.
- 960 Luzón Marco, Maria Jose. 1998/99. The discursive function of *do*-support in positive clauses.
 961 *Revista Española de Lingüística Aplicada* 13: 87–102.
- 962 Miller, Jim. 2006. Focus in the languages of Europe. In *Pragmatic organization of discourse in the*
 963 *languages of Europe*, ed. Bernini Giuliano and Marcia L. Schwartz, 121–214. Berlin/New
 964 York: de Gruyter.
- 965 Mukherjee, Joybrato. 2009. The grammar of conversation in advanced spoken learner English:
 966 Learner corpus data and language-pedagogical implications. In *Corpora and language teaching*,
 967 ed. Karin Aijmer, 203–230. Amsterdam: John Benjamins.
- 968 Müller, Simone. 2004. 'Well you know that type of person': Functions of *well* in the speech of
 969 American and German students. *Journal of Pragmatics* 36(6): 1157–1182.
- 970 Müller, Simone. 2005. *Discourse markers in native and non-native English discourse*. Amsterdam:
 971 John Benjamins.
- 972 Nevalainen, Terttu, and Matti Rissanen. 1986. Do you support the *do*-support? Emphatic and non-
 973 emphatic *do* in affirmative statements in present-day spoken English. In *Papers from the third*
 974 *Scandinavian symposium on syntactic variation*, ed. Sven Jacobson, 35–50. Stockholm:
 975 Almqvist and Wiksell.
- 976 Oberlander, Jon, and Judy Delin. 1996. The function and interpretation of reverse *wh*-clefts in
 977 spoken discourse. *Language and Speech* 39(2–3): 185–227.
- 978 Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive*
 979 *grammar of the English language*. London: Longman.
- 980 Ramírez Verdugo, Dolores, and Jesus Romero Trillo. 2005. The pragmatic function of intonation
 981 in L2 discourse: English tag questions used by Spanish speakers. *Intercultural Pragmatics*
 982 2(2): 151–168.
- 983 Romero Trillo, Jesus (ed.). 2008. *Pragmatics and corpus linguistics. A mutualistic entente*. Berlin:
 984 Mouton de Gruyter.
- 985 Rühlemann, Christoph. 2011. Corpus-based pragmatics II: Quantitative studies. In *Foundations*
 986 *of pragmatics*, Handbooks of pragmatics, vol. 1, ed. Wolfram Bublitz and Neil R. Norrick, 629–656.
 987 Berlin: DeGruyter Mouton.
- 988 Sasaki, Miyuki. 1998. Investigating EFL students' production of speech acts: A comparison of
 989 production questionnaires and role plays. *Journal of Pragmatics* 30(4): 457–484.
- 990 Scott, Mike. 2008. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.
- 991 Taguchi, Naoko. 2009. Pragmatic competence in Japanese as a second language: An introduction.
 992 In *Pragmatic competence*, ed. Naoko Taguchi, 1–18. Berlin: Mouton de Gruyter.
- 993 Thomas, Margaret. 1994. Assessment of L2 proficiency in second language acquisition research.
 994 *Language Learning* 44(2): 307–336.
- 995 Thomas, Margaret. 2006. Research synthesis and historiography: The case of assessment of sec-
 996 ond language proficiency. In *Synthesizing research on language learning and teaching*, ed.
 997 John M. Norris and Lourdes Ortega, 279–298. Amsterdam: John Benjamins.
- 998 Weinert, Regina. 1995. Focusing constructions in spoken language. Clefts, Y-movement, themati-
 999 zation and deixis in English and German. *Linguistische Berichte* 159: 341–369.
- 1000 Weinert, Regina, and Jim Miller. 1996. Cleft constructions in spoken language. *Journal of*
 1001 *Pragmatics* 25: 173–206.
- 1002 Yuan, Yi. 2001. An inquiry into empirical pragmatics data-gathering methods: Written DCTs, oral
 1003 DCTs, field notes, and natural conversations. *Journal of Pragmatics* 33(2): 271–292.

Author Queries

Chapter No.: 2 0001923558

Queries	Details Required	Author's Response
AU1	Please provide opening appropriate opening parenthesis in the sentence "...as but, however, nevertheless or (al)though...".	
AU2	Please cite Callies et al. (forthcoming) in text.	
AU3	Please update Aijmer and Rühlemann (forthcoming), Callies (in preparation), Callies et al. (forthcoming).	

Uncorrected Proof