Using conversation tasks and retrospective methodology to investigate L2 pragmatics development: The case of EFL criticism and responses to criticism

Abstract

Data collection methods constitute a major area of concern in interlanguage pragmatics (ILP) research, since there is no easy way to collect the type of data that is relatively ‘naturalistic’ while at the same time allowing for researcher control. Further, current ILP research also suffers from the shortage of studies making use of self-report data to understand how L2 learners plan and execute their speech act behavior. This paper discusses the advantages of an innovative method devised for collecting naturalistic data on ‘institutional’ criticisms and responses to criticisms in the L2 in a controlled manner. The paper also addresses the potential of retrospective interviews in researching IL pragmatic knowledge and issues involved in pragmatic decision-making, especially when used alongside with the performance data. This paper is grounded on data from 4 hours of recordings of 72 peer-feedback conversations by 36 EFL learners and 8 hours of recordings of 36 retrospective interviews with these learners.

Key words: conversation elicitation task, L2 pragmatic competence, retrospective interview
Introduction

Data collection is a “powerful determinant of the final product” (Kasper and Dahl 1991: 216); thus, it constitutes a major area of concern in any field of research. This is even more of the case when it comes to interlanguage pragmatics (ILP) research, where there is no easy way to collect the type of data that is both relatively ‘naturalistic’ while at the same time allowing for researcher control. In Author’s (xxxx) I reviewed the data collection instruments of 80 published ILP studies and found that only 21 were based on naturally occurring data (collected via field notes, observation, or telephone conversations). This is a small number, given that naturalistic data are desirable for the study of speech acts. The main reason for this may be that naturalistic data do not always allow for researcher control of relevant social and contextual variables, thus making the findings less comparable. Additionally, it is also not easy to gather a large enough corpus of data for comparison in this way (Yuan 2001). Given these difficulties, many ILP researchers tend to draw on elicited data as an alternative. For example, in the above review 35 studies employed Discourse Completion Tasks (DCT), both oral and written, 27 studies employed Role Play (RP), both open and closed, and another 15 employed questionnaire (including Multiple Choice Questionnaire –MCQ). However, recent studies of various elicitation methods found that they, too, may be problematic (see Golato 2003, Kasper and Dahl 1991, Kasper 1999, 2008 and Yuan 2001 for a review).

The main difficulty associated with DCT and questionnaire is that although these instruments tend to be effective in gathering a large amount of data in a short time (Beebe and Cummings 1985) and do not involve the hassle of transcribing data (Johnston, Kasper and Ross
1998), they are much criticized for their shortcomings in representing authentic speech in terms of response lengths, turn-taking, chance for opting out, and actual wordings (Aston 1995, Beebe and Cummins 1985, 1996, Bodman and Eisenstein 1988, Hartford and Bardovi-Harlig 1992, Holmes 1991, Kasper 2008, Rintell and Mitchell 1989, Turnbull 1997. Also see Bonikowska 1988 for a discussion of opting out as a pragmatic strategy). Even if such instruments are designed to allow for multiple-turn exchanges and opting out (e.g. Billmyer and Varghese 2000), they may still not be able to exactly capture what the participants would say in real life contexts. This is because, as highlighted by Wolfson (1989), what participants think they would say may be totally different from what they actually say under communicative pressure. DCTs have also been found problematic in eliciting appropriate data from speakers of non-western languages (Rose 1994).

An RP, on the other hand, allows for online L2 production in conversational sequences, and thus shares more similarities with natural speech production than a DCT or a questionnaire (Houck and Gass 1996, Kasper and Dahl 1991, Kasper 2008), but is still of doubtful authenticity. One difficulty is that if interlocutors are not good actors, or if the RP situations are not realistic enough, they might find it hard to perform naturally (see Bonikowska 1988, Kasper and Dahl 1991). An RP also has no consequences for the role players, and so it does not matter if they present themselves rude, which is completely different from real-life communication (Golato 2003). Finally, like any other type of elicitation method, an RP is predominantly motivated by the researcher’s goals rather than by the interactants’ goals (Kasper and Dahl 1991, Houck and Gass 1996). Thus, it might invite a certain degree of conscious decision-making, which is
usually absent in natural communication (Ellis, personal consultation), and consequently might affect the naturalness of the data generated.

The fact that elicitation methods such as DCTs, questionnaires, and RPs cannot replace ethnographic observations in studying spoken language has given rise to the need for a method that, while allowing the researcher to control relevant variables, also generates relatively natural speech. To date, quite a few research attempts have been made to improve the authenticity of data gathered via elicitation methods. For example, in her study of L2 English and Japanese compliment responses, Baba (1999) recruited conversational leaders, who were requested to bring their real-life friends and to give compliments on the photographs that the latter had brought along with them. The purpose of this procedure was to allow the conversation leaders to elicit spontaneous compliment responses from their friends. Similarly, Author (xxxx) attempted to elicit spontaneous criticisms and responses to criticisms in the target language (TL) by getting the informants to write essays on a given topic and arranging them into pairs to critique one another’s work. Both of the above methods allow the researchers to ‘naturalize’ their data while at the same time manipulating the interlocutor role relationships as well as the topic of the speech acts in question, both of which factors might impact significantly on the speech produced.

Another area of concern in current ILP research is the undue shortage of studies making use of self-report data to understand how L2 learners plan and execute their speech act behavior. Self-report data include learners’ “descriptions of what they do, characterized by generalized statements, for instance, about their language learning strategies” (Cohen 2013: 1). To date, very few ILP studies have employed retrospection methods (e.g. Arent 1996, Barron 2000, Cohen and
Olshtain 1993, DuFon 1999, Eisenstein and Bodman 1986, Han 1992, Hoffman-Hicks 1999, Nakashama 1999, Author xxxx, Robinson 1992 and Widjaja 1997). This is in stark contrast to other lines of SLA research, for example those examining learner strategies in reading, lexical inferencing, test-taking, and so on (see Gass and Mackey 2000 for a comprehensive review). Earlier ILP studies are often restricted to the use of judgment rating-scales or assessment questionnaires for examining learners’ social and contextual assessment (e.g. Fraser, Rintell, and Walters 1980, Hassall 1997, House 1988, Maeshiba, Yoshinaga, Kasper and Ross 1996, Rintell 1981). However, these instruments, while appropriate for understanding learners’ pragmatic knowledge, may be inadequate for investigating other thought processes involved in learners’ decision-making (see Robinson 1992).

The shortage of self-report data also makes it difficult to investigate how learners at different stages of L2 development differ in their pragmatic decision-making. This insight is crucial, especially when learners of different proficiency levels are found not to differ in their performance (e.g. Takahashi 1996 found no proficiency effects on the amount of transfer; Takahashi and Beebe 1987, Omar 1991 found no differences in speech act realization strategies by proficiency levels). Presumably, while high proficiency and low proficiency groups may show no significant quantitative difference in their preferences for a particular pragmatic strategy, their reasons underlying these choices and judgments may not be the same. Thus, a combination of both quantitative data (on pragmatic performance) and qualitative data (on pragmatic decision-making) is preferable. Such qualitative data can be elicited via self-report methods such as retrospective interviews (as suggested by Cohen 1996).
The present study aims to address the above gaps by examining an innovative method for collecting L2 speech acts, hereafter referred to as a conversation elicitation task, and highlighting its advantages in researching complex and highly offensive speech acts such as criticizing and responding to criticisms. The present study also discusses the potential of using retrospective interviews for researching IL pragmatic knowledge as well as issues involved in pragmatic decision-making. Both of the conversation task and the retrospective interviews have been employed by Author (xxxx) to investigate L2 pragmatic development in L2 use of criticisms and responses to criticisms. In this paper the two methods will be discussed from a methodological perspective, with a view to evaluating their contributions to the field of ILP especially when used alongside each other.

The study

Participants

In order to achieve the aims of the current study, the conversation elicitation task was carefully devised, validated in a pilot study (see Author xxxx) and employed for gathering data on the speech act of criticizing and responding to criticisms from a group of 36 randomly selected Vietnamese learners of English as a Foreign Language in Hanoi, Vietnam. These learners were attending an English for Academic Purposes course in preparation for their higher education in Australia at the time of data collection. They comprised 12 high beginners, 12 intermediates, and 12 advanced learners. The learners’ proficiency levels were determined by their pre-entry International English Testing System (IELTS) scores using the classification by the Program Administrators. According to this classification, an overall score of 5.0 and below represented
the high beginning level. An overall score between 5.5 and 6.0 represented the intermediate level. An overall score of 6.5 and above represented the advanced level. (In this study, the actual mean IELTS score was 5.0 for the beginner group, 5.5 for the intermediate group, and 7.3 for the advanced group). This grouping was in agreement with the learners’ self-assessment of their English proficiency levels. Among them, twenty four were female and twelve were male, with the mean age being 22.9 for the high beginners, 23.8 for the intermediate group, and 25.8 for the advanced group. Also, twenty-four of them were prospective postgraduates and twelve were prospective undergraduates for their study programs in Australia, with various majors. The learners originated from different parts of Vietnam and spoke different dialects of Vietnamese.

**Instruments**

For the conversation elicitation task, the learners were firstly required to write a 250-word argumentative essay on the benefits of public as opposed to private transportation. This writing topic was taken from a commercially available IELTS practice book and was chosen as it did not require specialist knowledge and would not be too difficult for learners of high beginning level. Then the learners were randomly arranged into dyads within each group (i.e. beginner-beginner, intermediate-intermediate, and advanced-advanced dyads) to give critical feedback on each other’s essays. The learners were provided three main prompting assessment criteria, of which they could take advantage when commenting on their peers’ written work. These criteria comprised essay organization, essay content and language use. To make the task more natural, it was emphasized that besides commenting on points with which they were dissatisfied, the learners were also free to make positive remarks about their friends’ essays. The instruction for
the task was written in the learners’ mother tongue, i.e. Vietnamese, to ensure their full understanding of the task. The language used in the instruction and the prompting questions was made as simple as possible so that it would be comprehensible to laypeople. Additionally, the instruction also explicitly encouraged the learners to ask questions before they started if they found something confusing or unclear.

The peer-feedback conversations were then audio-taped for later analysis. It should be noted that giving peer-feedback was a quite common learning task in the program where the learners were undertaking their English studies. Therefore, it was expected that they were all quite familiar with this type of task, thus making the data more natural. Also, this task allowed for researcher control of relevant social and contextual variables, such as the relative social power (equal) and distance (neutral) between the participants (peer-to-peer), and the topic of criticisms (an English essay), thus making the data more comparable.

In order to account for the learners’ pragmatic decision making when completing the conversation task, the researcher also interviewed the learners in audio-playback sessions. Like the conversation task, the interview questions were also carefully validated before being employed in the present study. As discussed above, the advantage of the restrospective interview lies in its provision of access to the process-oriented data which is unobtainable via other instruments (Cohen and Olshtain 1993). It might also help overcome such potential pitfalls as the researcher’s incorrect inferences about the causes of the observed behaviors (Gerloff cited in Robinson 1991). Furthermore, when collated with performance data, the interview data might reveal more information about the thought processes by which the learners finally arrive at their
pragmatic choices, information which cannot be discovered from the performance data alone (see Cohen 2013, Robinson 1992). The interview also helps the researcher clarify any unclear details in the peer-feedback conversation data. It should be noted that despite the above advantages, retrospection is not free from limitations. Some researchers (Schmidt cited in Robinson 1991, Seliger 1983) caution that some processes, especially those acquired implicitly, may not be entirely available for conscious reporting. Or, some mental processes may be too complex to be accurately verbalized (Cohen 1987, 1991). Importantly, interviewees may also not be able to recall all their thoughts at the moment of retrospection, and thus offer post hoc rationalization instead (Basturkmen, Loewen and Ellis 2004, Cohen 2013, Faerch and Kasper 1987).

Fortunately, these limitations can be minimized in a number of ways. To facilitate interviewees’ recalls, the time lapse between task completion and retrospective sessions should be minimized and aids to task performance such as video or audio-recordings should be employed to remind participants of their thoughts (see Gass and Mackey 2000). In the current study, the learners were interviewed immediately after they completed their peer-feedback conversation tasks (see Cohen 2013). During the interview, these conversations were also played back to refresh their memories. The researcher also probed the learners’ thoughts in connection with specific instances of their language use rather than as non-task related memory (see Ericsson and Simon 1984, 1987). Furthermore, the researcher also avoided asking questions that may exert additional cognitive load on interviewees, thus interfering with their recalls (e.g. “Why-questions”) (Gass and Mackey 2000). The interview was also conducted in the learners’ mother tongues to avoid the possibility that making them to translate their thoughts from one
language to another may affect their short-term memory (see Cohen 2013, Faerch and Kasper 1987). To avoid the possibility of post hoc rationalization, the researcher only asked about the information that learners attended to and did not “lead” them or try to elicit generalizations. Also, where probes were needed, they were used in such a way as to clarify unclear answers but not to ask about information that the learners did not give (see Cohen 1987, 1991). Finally, where the answer is “I don’t know” or “I don’t remember” the researcher made sure to stop probing the learners further (see Gass and Mackey 2000).

Results and discussion

The conversation elicitation task

The conversation task generated approximately 4 hours of recordings of 72 peer-feedback conversations. These conversations yielded 163 criticisms and 126 responses to criticisms, which were realized by means of 459 and 427 strategies, respectively. On average, each criticism was realized by means of 2.8 strategies and each response to criticism was realized by means of 3.4 strategies. Each peer-feedback conversation contained 2.3 criticisms and 1.8 responses to criticisms. Table 1 and Table 2, adapted from Author (xxxx), illustrate the different pragmatic strategies that Vietnamese EFL learners employed for expressing their criticisms and responses to criticisms.

PUT TABLE 1 AROUND HERE

PUT TABLE 2 AROUND HERE
In Author (xxxx) I found that the three different proficiency groups varied significantly from one another only in the frequency with which they mitigated their criticisms. More specifically, the more proficient the learners became, the more frequently they mitigated their speech acts. They did not, however, differ in their patterns of preference for a particular pragmatic strategy. When compared to the Australian native speaker group, the learners, regardless of their proficiency levels, tended to lag far behind in terms of their frequencies of using mitigators. The learners also tended to produce considerably more “demands” and “advice” than the NSs when criticizing their peers. Especially, they preferred the two structures “You should” and “You must” when indicating where improvements should be made in their peers’ essays. The learners also resisted the criticisms given to them nearly as often as they accepted these criticisms, while the NSs tended to accept criticisms most of the time and rarely resisted these. The strategies the learners used most often when accepting a criticism included “agreement”, “seeking help”, and “offer of repair”. On the other hand, the strategies they used most often when resisting a criticism included “dispute” and “justification”. In many instances, they failed to hedge these oppositional talks. In the following section, learner criticisms and responses to criticisms will be discussed with specific regard to the question to what extent their interactional aspects can be captured by the conversation task.

First, a prominent feature of using the conversation task is that it produced not single but multiple turns. On average, a criticism and response to criticism were negotiated via 6.4 turns. This resulted in an average of 29.3 turns per peer-feedback conversation, representing a highly dynamic interaction with extensive negotiation. Let’s consider the following example, in which the speakers required a total number of 37 turns to reach a compromise.
Exchange 1 (Two advanced learners)

01 L1: and one more idea in this paragraph it also means LESS traffic accident I AGREE that less traffic accident happen when we use more public transport

02 L2: hmm

03 L1: but let's consider when (..) we have one case of accident with public transport so there may be MORE people involved in death and injury than the private use do you think so?

04 L2: uhh

05 L1: maybe ah TEN times (.2) maybe 10 case of the [private accident =

06 L2: [yeah I see

07 L1: = cause 10 deaths

08 L2: yes [I see ah =

09 L1: BUT

10 L2: = your point I see that

11 L1: just one

12 L2: I see that point but ah frequency of occurrence of ah (. ) of public transport is MUCH MUCH less than the frequency of ah of ah of private transport

13 L1: ^I don't think so. yeah because maybe the air plane accident, the train accident, yeah something like that

14 L2: yeah you may afraid that in plane accident and train accident because of the the number of people die and the number of ah I mean injuries and casualty but ah (. ) the frequency of it I mean the number of people die and the number of people die in such ah -in when they have when they using their private vehicles is I mean that far more ah equal than the number of people die in the public transport

15 L1: yeah just frequency

16 L2: yeah I mean the frequency, yeah

17 L1: uhh, I agree but it means that because [here you said =

18 L2: [and because

19 L1: = it is the outcome. it means you consider maybe the number of people die or the number of people injured

20 L2: Yeah

21 L1: but here it's not I -DON'T think it's a strong argument

22 L2: Yeah

23 L1: for the use of public transport in case of bad traffic accidents

24 L2: I mean less traffic accident it means that the more people are get involved in ah in ah -the more people rush in streets and using their OWN vehicles there will be more ah more

25 L1: more case, more chance

26 L2: more chance, yeah

27 L1: I see

28 L2: more chance of accidents yeah

29 L1: yeah but the outcome of accident it means the NUMBER of people die or the
In the above exchange, Learner 1 was trying to point out that Learner 2 made use of a weak argument (i.e. public transport causes fewer accidents and number of casualty than private transport does). Learner 2 disagreed with this criticism and was trying to justify her claim (i.e. accidents caused by public transport attract more of public attention because they are often in the news, which leads to the wrong impression that public transport causes more damages than private transport does). What can be seen from this exchange is the criticism and response to criticisms were not achieved in one turn but over a number of extended turns, during which time the speakers challenged the other party (e.g. turns 6, 10, 12, 13, 21, 32, 33), clarified, modified and elaborated their arguments (e.g. turns 5, 7, 11, 12, 14, 16, 18, 19, 23, 24, 25, 29, 32, 34, 36), and “generally became involved in negotiating semantic, pragmatic, and social meaning” (Houck and Gass 1996, p. 52) until they reached the final resolution (turn 37). In other words, the speakers required a great deal of efforts to achieve their intentions due to the complexity and high degree of face-threat involved in giving and responding to criticisms.

The extent to which the learners interacted to resolve the disharmony seemed to depend on the topic of the criticism and response to criticism, or the level of imposition involved in
producing these speech acts. They seemed to need more negotiation when discussing an opposing idea (e.g. exchange 1) or dealing with an organizational issue than when they were dealing with a grammatical or spelling error. Perhaps this is because criticizing someone’s ideas or way of organizing his or her essay is considered more obtrusive than indicating his or her grammatical errors and thus requires more extensive face-work. It could also be that the former type of criticisms is based more on one’s subjective assessment, and thus requires more clarification and elaboration than the latter type, which is based merely on facts. Nevertheless, when the two speakers did not agree on a grammatical issue, they also required relatively lengthy interactions to reach a compromise. For example, Exchange 2 below involved a total of 31 turns over which these learners argued why they believed or did not believe that the three phrases “reduced density of traffic”, “better health” and “cost saving” constituted parallel structures.

Exchange 2 (Two advanced learners)

01 L3: *maybe the very small idea I think it's better to change some ah -in terms of “reduced density of traffic”, “better health” and “cost saving” so I think it ah they are a bit not very balanced. for example here you use verb-ing, here is noun and noun with the ah adjectives before it, but not very clear*

02 L4: *but <they are all phrase> I mean that they are phrase, “cost”, “health”*

03 L3: *yes =*

04 L4: *=and I means it is the verb, it is the noun and adjective but they are all phrase*

05 L3: *yeah they are phrase*

06 L4: *yes*

07 L3: *but they should be maybe should be balanced in terms of structure. it means here if you use “cost saving”, “cost” is noun, “saving” is a adjective for example*

08 L4: *hmm?*

09 L3: *so () some other phrases like here you should be () -it's better to have noun and then verb-ing*

10 L4: *yeah I see*

11 L3: *like adjective I think it should be [more balanced*

12 L4: *[I think it is B:Alanced because they are all 3 phrases put together, 3 =*

13 L3: *yeah I agree they are 3 phrase noun phrases BUT I mean that the phrase should*
be structured similarly

14 L4: is it ah compulsory or because if I say that it's "to reduce density" and then "saving cost" or "cost saving" it is the ah infinitive verb and then a verb –ing
15 L3: no if you use “to reduce density” “to have better health”
16 L4: Yeah
17 L3: and “to save cost”
18 L4: yeah it is (.) so it is equal right
19 L3: yes, it's more balanced
20 L4: but they're all PHRASES
21 L3: phrases but here ah here it is the head noun and before the head noun it's a kind of comparison (.) but here it's a noun and after noun it's verb-ing, here it's a noun
but before noun
22 L4: hmm
23 L3: it's adjective in the passive voice, so it means I mean the structure, the structure of the phrase ONLY (.) they are not balanced
24 L4: I see I see it's ah
25 L3: I think it's BETTER if it's ah (.) if they are balanced
26 L4: (.3) uhh (.4) I have never thought of ah use the balances between the phrases (.)I just think that it is not here it's not a verb, it's not ah so they are should be all verbs
27 L3: Yeah
28 L4: but if they are phrases, so sounds OK, in my opinion
29 L3: ((laugh))
30 L4: yeah thank you for your comment
31 L3: yeah ((laugh))

From the developmental perspective, the extent to which learners negotiated their criticisms and responses to criticisms seemed to depend on the learners’ proficiency levels in the TL. On average, an advanced learner produced a criticism in 10 turns and a response to criticism in 9.7 turns. The respective corresponding figures were 5.5 each for the intermediate learners and 4.6 and 4.7 for the high beginners. Note that the three groups did not statistically differ in the frequencies with which they discussed an opposing idea (F=1.161, n.s. at p =.326), organizational issue (F=.467, n.s. at p =.631) or a grammatical issue (F=.820, n.s. at p=.449); thus it was unlikely that the differences in their lengths of negotiation resulted from their varying topics of criticisms. On the contrary, these differences might have resulted from the higher level
linguistic competence, which allowed the advanced and intermediate learners to produce a greater number of supportive moves to preface and compensate for their offensive acts (cf. mean number of supportive moves per criticizing strategy: $M=.38$ for high beginners, $M=1.37$ for intermediate learners, and $M=1.74$ for the advanced learners; mean number of supportive moves per responding strategy: $M=.07$ for high beginners, $M=.11$ for intermediate learners, and $M=.10$ for advanced learners). The higher level of linguistic competence also allowed the advanced learners to produce a greater number of strategies for realizing their speech acts, which might have explained for their greater number of turns (cf. mean number of strategies per criticism: $M=3.2$ for high beginners, $M=3.0$ for intermediate learners, and $M=4.4$ for advanced learners; mean number of strategies per response to criticism: $M=2.8$ for high beginners, $M=2.6$ for intermediate learners, and $M=4.4$ for advanced learners) (see Author xxxx).

Finally, another feature that made the conversation task data comparable to natural discourse is that they contained abundant characteristics specific to spoken language such as overlapping (e.g. turns 6-10, turns 17-19 in exchange 1, turns 11-12 in exchange 2), back channels (e.g. hmm, oh, okay), pause fillers (e.g. ah, uh), discourse markers (e.g. well, you know, I mean, I see), false starts and self-repair (e.g. turns 5, 21, 32 in exchange 1, turns 7, 9 in exchange 2), hesitation, and repetition.

*The retrospective interviews*

As mentioned earlier, the retrospective interview has two purposes. Firstly, it was designed to investigate the factors that may have affected the learners’ choice of pragmatic strategies for realizing criticisms and responses to criticisms that were found in the conversation task data. For
this purpose, the learners were asked questions regarding their L2 pragmatic knowledge, L1 influence, processing issues, and learning experience. Secondly, the interview was used for the purpose of triangulation of data, i.e. the combination of different data sources to maximize information on complex issues. This helps to reduce possible task bias and enhance confidence in the objectivity and reliability of the findings. The following section will discuss the above-mentioned four major sources of influence on learners’ pragmatic behavior: (1) L2 pragmatic knowledge, (2) L1 influence, (3) processing difficulties, and (4) learning experience. References to their performance data (elicited via the conversation task) will also be made for the triangulation purpose. The interview extracts cited in this section are English translations from their Vietnamese equivalents.

First, the meta-pragmatic information provided by the learners in the retrospective interview seems to indicate that in many instances their L2 pragmatic knowledge is incomplete. This accounts for several instances of their non-native use of criticisms and responses to criticisms as revealed in the conversation task data. For example, in Author (xxxx) I found that a number of learners overused “demand” to give feedback on their friends’ essays. The interview data reveal that in many instances, they did not realize the inappropriateness of this use. For example, 53% of them who realized “demand” by the structure “must” thought it was an appropriate way of expressing rules and principles. Likewise, there was a misconception that English NSs preferred ‘directness’, which was wrongly perceived by the learners as involving a lack of modality. Eighty two percent of the learners who demonstrated over-explicitness, when asked whether they thought the Australian NS would criticize in the same way as them in a similar situation,
responded “Yes”. Not surprisingly, due to this conception, they produced many over-explicit criticisms and responses to criticisms such as bald “disagreements”.

Others, who tried to be polite, misused “You should” due to their misconception of the pragmalinguistic form-function mapping of this structure. For instance, 90% of learners who employed “You should” thought it was a polite way of criticizing. A few learners, who demonstrated correct knowledge of the pragmalinguistic function of this structure, still misused it due to lack of knowledge of sociopragmatic constraints. For example, one learner who considered the strategy “advice” to be “obligatory” and “imposing” believed that it was still acceptable among people of equal social status: “It depends on the relationship between the interlocutors. If two friends or two people of the same social status say “You should do this” to each other, it is simply advisory. It’s not too strong or imposing. But if a teacher says “You should do this” to his or her student, it may imply obligation.”

In Author (xxxx) I also observed that in the absence of L2 pragmatic knowledge, the learners might have transferred a number of L1 pragmatic routines, such as a preference for “demands” and “advice” as criticizing strategies, the tendency to challenge one another in replying to criticisms, and the low frequency of mitigating criticisms and responses to criticisms. During the interview, many learners commented that they were influenced by their L1 when selecting these routines. For example, they reported preferring to give “advice” because this is an acceptable practice within Vietnamese culture, encouraging intimacy rather than personal space. On the other hand, they also reported the belief that once one engages in “true argumentation”, one needs to put forward strong opinions in order to sound convincing with the result that they
produced many over-explicit utterances: “*When we argue, we usually want to defend our opinion. Sometimes we use strong words just because we want to emphasize our opinion, defend it, and make the listener understand it.*”

The learners’ retrospective reports about their perception of the L1 and L2 pragmatics support Kellerman’s (1983) transferability hypothesis. Put briefly, Kellerman identifies three factors that affect transferability. These include psycholinguistic markedness (i.e. learners’ awareness of the specificity or universality of an L1 feature), the reasonable entity condition (i.e. TL reasonableness assumption by learners in the absence of TL knowledge), and psychotypology (i.e. learners’ belief regarding the distance between their L1 and the TL). In the present study, since the learners presumed that “advice” was a ‘friendly’ way of criticizing in the L2 just as in their L1 and that strong argumentation is appropriate when people need to defend their opinion, they made extensive use of “advice”, “demands”, and “disagreements” (i.e. the reasonable entity condition). Since they also perceived proximity between the English structure “*should*” and the Vietnamese structure “*nèn*”, and similarity between the English structure “*must*” and the Vietnamese structure “*phải*”, they reported transferring these two L1 structures when respectively realizing “advice” and “demand” in the L2 (i.e. psychotypology). It should be noted that in the present study non-transfer occurred in cases where the learners did not perceive any correspondence between an L1 structure and that of the L2. For example, those learners who did not perceive any correspondence between “*nèn*” and “*should*” did not engage in the transfer of this structure.
The interview data also revealed two types of communication transfer (i.e. the use of the L1 for comprehension and production of the L2 in a specific communicative situation – see Ellis 1994: 336-338) or what the learners referred to as “translation”. One is strategic transfer, which involves learners’ reliance on the L1 to solve a specific communication problem despite their awareness of the non-transferability of the structure. For example, four high beginner learners admitted deliberately resorting to their L1 resources when they were lacking the L2 structures needed to convey their intended meaning, although they were aware that this translation might not have been accurate: “Actually I did not know what could be the best structure to use in criticizing. So I normally used Vietnamese structures but I also know that this use might have been inaccurate in English.” The other type of communication transfer is automatic transfer, which takes place when highly automatized L1 routines override the awareness of non-transferability. For example, an advanced learner reported that her automatic use of the structure “You should” to give “advice” was due to L1 influence, even though she was aware that this might not be an appropriate use in English: “Since I started learning English at high school, I have got used to this structure (i.e. “should”). It is like a habit. And although now I am aware that it is not an appropriate way to give advice and have learnt other alternative structures, I still keep coming back to use it when I do not pay enough attention.” (also see Olshtain, 1983 for further such findings). Importantly, these types of transfer could not have been identified from the performance data alone.

Further, a combination of the interview and performance data provides important information on processing issues, especially the cognitive difficulties which faced the learners in spontaneous interaction and which they deemed partly responsible for many of the over-explicit
utterances produced in this condition. As discussed elsewhere (e.g. Author: xxxx), the learners demonstrated inappropriate explicitness in realizing criticisms and responses to criticisms due to a lack of modality markings in these two speech acts. Evidence of this modality reduction was also apparent from the interview data when a comparison was made between the learners’ intended meaning as expressed in the L1 and its realization in the L2. The following examples reflect what Kasper (1982: 107) refers to as “L1 transfer with reduction” or the psycholinguistic process whereby learners “preserve the illocution and propositional content” of what they intend to say based on their L1 knowledge, but reduce the modal component:

(1) L1 intended meaning: “Some of your ideas sound interesting but there are some I don’t really agree with” vs. L2 realization: “I don’t agree with you.”

(2) L1 intended meaning: “It might be necessary to give a definition of public transport” vs. L2 realization: “Hmm when you say using public transport, is it necessary to give a definition what the public transport is.”

When interviewed, many of the learners commented that under the pressure of online speech production, they concentrated first and foremost on the basic speech act. For example a learner said “I concentrated on what to criticize and how to express it precisely but not what word to use to sound polite. Whatever word came along, I used it”. To put it another way, they focused on ensuring that their intentions were precisely understood before making themselves sound polite. Others reported choosing fluency over modality: they would rather not pause to select appropriate linguistic realization structures at the expense of speech flow. As a result, these learners tended to simplify their language by resorting to simple structures that were most
accessible to them at the moment of speech production and by reducing modality in order to cope
with the competing processing demands. In many instances, the learners commented that had
they had more time to plan their utterances, they could have made their utterances more polite.
For example, this learner commented: “Sometimes I had to be spontaneous and thinking back
about it I wondered “Oh gee why did I use that structure, why did I not use a better one?”.

Indeed, when completing a written DCT, which were designed to elicit criticisms and
responses to criticisms in eight hypothetical peer-feedback situations (see Author xxxx), the
learners generally produced a higher mean of “advice” and “suggestions” – the two strategies
that they considered “polite” - than did they in the conversation task. This seems to indicate that
in the un-pressured written condition, the learners could pay more equal attention to what to say
and how to say it than in the spontaneous speech condition. Interestingly enough, however, the
learners did not produce a greater number of modifiers when performing criticisms and
responses to criticisms in the written DCT. These results suggest that although the written
condition allowed them to pay more attention to politeness, the learners still did not mitigate
their criticisms and responses to criticisms to a greater extent. Presumably, this was the result of
a lack of L2 pragmatic knowledge combined with the transfer of L1 pragmatic routines as
discussed in the previous sections.

Regarding the third source of influence, i.e. learning experience, research has shown that the
misleading information that teachers (both via their instruction and classroom management
discourse) and textbooks sometimes inadvertently introduce to learners may hinder rather than
promote their learning of L2 pragmatics (Bardovi-Harlig 2001, Boxer and Pickering 1995,
Kasper 1997, Author xxxx). Although the present study did not conduct any classroom observation of the use of the speech acts of criticizing and responding to criticism within textbooks, it found from the learners’ retrospective reports several instances of teacher-induced errors, which in many cases interplayed with L1 transfer. For example, 65% of the learners who employed the structure “You should” commented that they were explicitly taught that this was a polite way of giving “advice” or feedback to their interlocutors. They also claimed that their teachers often used this structure when correcting their essays and that dialogues in textbooks displayed similar usage.

Likewise, in the case of “demand”, many learners held the view that it was a pragmatically appropriate criticizing strategy to use when principles and rules were involved, perhaps because they had not been taught important pragmalinguistic information about this strategy. As these learners pointed out, the structure “must” and “have to” were only taught in terms of grammatical properties and semantic meanings by both textbooks and teachers. Apparently, in those cases, it was the misleading input given via learning materials, teacher-talk, and insufficient pragmatic instruction by teachers that encouraged the learners to draw incorrect conclusions.

Further, when cross-checked with the performance data, the retrospective data reveal some insights into grammatical constraints on L2 pragmatic development. One significant difference between the three proficiency groups of learners was the higher frequency of use of criticism modifiers by the intermediate and advanced learners as compared to their high beginner peers (see Author xxxx). This finding was supported by the retrospective data which showed that as
the learners became more proficient in the L2, they were more likely to pay attention to both the propositional content of their speech acts and modality marking. On the other hand, lower proficiency learners had had more difficulties in expressing their intended meanings due to limited L2 competence, which seemed to inhibit them from activating politeness realization devices in spontaneous speech conditions. For example, while 10 out of 12 high beginners reported focusing more on being understood correctly rather than trying to be polite when criticizing, only 8 out of 12 intermediate learners and 6 out of 12 advanced learners did so. The high beginner group mentioned “focus on message clarity” 24 times in total, while the intermediate group did so only 11 times and the advanced group 9 times. This finding also suggests a better control over processing, or a more developed procedural knowledge by the higher proficiency learners (see Bialystok 1993).

Finally, the findings from the conversation task and the interview data also suggest some proficiency effects on L1 transfer, as is often claimed in the transfer literature (see Takahashi 1995 for a review). Elsewhere (e.g. Author xxxx) I found that the high beginners were closest to the Vietnamese NSs who provided L1 baseline data in the frequency with which they modified the illocutionary force of their criticisms, thus suggesting a negative correlation between L2 proficiency and L1 transfer. This finding was supported by the interview data which showed a decreased reported rate of L1 transfer for the higher proficiency learners. For example, it was found that 8 out of 12 high beginners claimed to transfer from the L1 while only two the intermediate learners and one advanced learner reported such transfer. In a few instances, it was also found that the higher proficiency learners reported selecting L2 speech act realization structures based on the impact that these structures might have on the listener rather than on their
linguistic equivalence with the L1. For example, this advance learner said, when explaining the mismatch in terms of linguistic structure between her criticism given in the L1 (“probably you should elaborate more on it”) and that in the L2 (“yeah I think you might it MIGHT be better if you elaborate on it a bit a little bit more I think. yeah?”):

“When speaking English I use ah I care about how to maintain the idea ah my intended meaning. I think these two utterances are equivalent ah they both can convey what I mean. ah because I am focused on the listener, like when I say something, what impact does it have on the listener? And I think that in this case my English and Vietnamese criticisms were equivalent.”

Conclusion

Similarly to many previous studies (see Golato 2003, Ellis 2008, Kasper and Dahl 1991, Kasper 1999, 2008 and Yuan 2001, for a review), this study shows that being an online oral interactive procedure, the conversation task tends to share many similarities with natural spoken communication. Like natural oral interaction of any type, it contains features of online speech production, many of which may carry pragmatic force, such as repetition, hesitation, overlapping, back-channeling, false starts, self-repairs, pause fillers and discourse markers. It also tends to prompt extensive negotiation through turn exchange and elaborated responses and thus allows to capture the complexity involved in performing highly face-threatening speech acts such as criticizing and responding to criticisms, especially when it is employed in combination with the retrospective interview. Its other strengths were that it allows for researcher control and results in a large corpus of data in a relatively short time. Its limitation, on the other hand, lies in the fact that as a data collection method, it has been designed for research purposes and thus is
not truly authentic conversation in its strictest sense. However, the fact that the participants in this study are not required to take on social roles different from their own can be considered to have reduced the unnaturalness of this task. Indeed, its capacity to elicit data that, to a certain extent, reflect online speech production, as indicated above, suggests that it is an adequate instrument for collecting spontaneous speech.

Another limitation of the conversation task is that while it is helpful in investigating one special type of criticisms, which is giving critical feedback in institutional settings, it does not allow an investigation of criticisms in other social scenarios which may vary in interlocutors’ role relationships and which might involve a higher level of ‘infraction’, for example, the more ‘biting’ type of criticisms such as criticizing about one’s appearance or behavior. In other words, it might not provide insights into learners’ sensitivity to different social and contextual variables. Also, while this method might be modified to investigate other institutional talks, for example arguing, disagreeing, challenging, or giving advice, the range of speech acts it allows to investigate seems restricted just to institutional discourse (see Bardovi-Harlig and Hartford 2008, Ellis 2008:167-168). Thus, if the research had focused more on the socio-pragmatic than pragma-linguistic aspect of learner language, or on non-institutional discourse, this method might not have been as useful as other methods of data collection, for example open role-plays or enhanced DCTs (see Billmyer and Varghese 2000). By discussing the pros and cons of the conversation task, it is therefore argued that there are no optimal data collection instruments which work effectively for all studies and that methodological choices should depend largely on the particular aims of the particular research study.
The use of a retrospective interview to probe into the learners’ pragmatic decision-making in this study has contributed to evidence of the usefulness of this method in researching meta-pragmatic awareness and processing issues, especially when it is employed alongside with the performance data. In the retrospective interview the learners reported a number of overlapping and interacting influences on their choice of criticizing and responding to criticism strategies, for example, their incomplete L2 pragmatic knowledge, transfer of L1 pragmatic norms, processing capability and learning experience. Some of these sources of influences could not have been identifiable otherwise from the performance data alone, for example, learners’ perceptions of L1 and L2 pragmatics that affect the likelihood that they will or will not transfer a pragmatic routine (see Kellerman 1983), their employment of communication transfer to compensate for their incomplete L2 pragmatic knowledge, or their cognitive difficulty that prompts them to practice “L1 transfer with reduction” under communicative pressure (Kasper 1982: 107).

A word of caution, however, should be raised in regard to the reliability of the interview data. As noted by some researchers (see Gass and Mackey 2000), stimulated recalls are not free from limitations. For example, some processes, especially those acquired implicitly, may not be entirely available for conscious reporting (Schmidt cited in Robinson 1992, Seliger 1983). Some processes may also be too complex to be accurately verbalized (Cohen 1987, 1991). The respondents may also be unable to recall all their thoughts at the time of retrospection and offer post hoc rationalization instead (Cohen 2013, Faerch and Kasper 1987, Basturkmen et al. 2004). The present study made several efforts to minimize the above possibilities by (1) conducting the interview shortly after completion of the conversation task, when the learners’ memory was still fresh, (2) conducting it in the learners’ L1 so that they could freely express their thoughts, (3)
always replaying the cassette tape for the learners when asking them a question related to the instances of inquiry, (4) avoiding the questions “why” and “why not” as they were believed to overload the informants’ memory, (5) avoiding leading questions and feedback in response to the informants’ answers except back-channeling (e.g. Hmm, huh, yeah), and (6) not probing further if the answer was “I don’t remember” or “I don’t know” (see Gass and Mackey 2000). The present study also employed a triangulation of data in the hope of reducing possible task bias and enhancing the objectivity of the findings. Although these precautions may not completely eradicate the potential problems, this technique tends to provide a valid means for researching IL pragmatic knowledge as well as process thoughts involved in IL pragmatic decision-making. Given its numerous benefits, it should be employed more abundantly and unreservedly in future ILP research.

References


Table 1. Taxonomy of criticizing strategies used by Vietnamese EFL learners

<table>
<thead>
<tr>
<th>Criticizing strategies</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Direct criticisms</strong></td>
<td>S explicitly points out the problem with H’s essay</td>
</tr>
<tr>
<td>1. Statement of problems</td>
<td><em>There are some incorrect words, for example “nowadays”</em>. I find it difficult to understand your idea.</td>
</tr>
<tr>
<td>2. Expression of disapproval</td>
<td><em>I don't like the way you write that ah &quot;I'm convinced about the idea&quot; or &quot;in my opinion&quot;</em>.</td>
</tr>
<tr>
<td>3. Negative evaluation</td>
<td><em>I think ah it's not a good way to support to one's idea</em>.</td>
</tr>
<tr>
<td>4. Expression of disagreement</td>
<td><em>I don’t quite agree</em> with you with some points (.) about the conclusion.</td>
</tr>
<tr>
<td><strong>II. Indirect criticisms</strong></td>
<td>S implies the problem with H’s essay by means of various strategies</td>
</tr>
<tr>
<td>1. Correction</td>
<td><em>&quot;safer</em> not <em>“safe”</em>, comparison. Theoretically, a conclusion needs to be some sort of a summary.</td>
</tr>
<tr>
<td>2. Request for change</td>
<td><em>I still want you to consider some points</em>.</td>
</tr>
<tr>
<td>3. Demanding change</td>
<td><em>You must pay attention to grammar. You have to talk about your opinion in your summary</em>.</td>
</tr>
<tr>
<td>4. Giving advice about change</td>
<td><em>I mean conclusion should have some sort of improvement</em>.</td>
</tr>
<tr>
<td>5. Giving suggestions about change</td>
<td><em>I think if you make a full stop in here ah (.) this sentence is clear</em>.</td>
</tr>
<tr>
<td>6. Rhetorical questions</td>
<td><em>Did you read your writing again after you finish it?</em></td>
</tr>
<tr>
<td>7. Other hints</td>
<td><em>I prefer a writing style which are not too personal</em>.</td>
</tr>
</tbody>
</table>

Symbol: S = The speaker  H = The hearer
Table 2. Taxonomy of strategies for responding to criticisms used by Vietnamese EFL learners

<table>
<thead>
<tr>
<th>Responding strategies</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Total acceptance</strong></td>
<td></td>
</tr>
</tbody>
</table>
| 1. Agreement | L1: *in the body paragraph you also use repeated words like ah "the reason" "the reason is that" "the reason like that". I think you can use another way.*  
L2: *yes, that's my weakness.* |
| 2. Offer of repair | L1: *for example one in one [para]graph you give 2 sort of examples for 2 ideas, so ah that is difficult for readers to follow.*  
L2: *difficult to follow yes maybe, I think I can divide this paragraph into 2 paragraph.* |
| 3. Seeking help | L1: *"That's why buses, subways, trains ARE encouraged to use", I think ah () that's a grammar mistake because ah bus hmm bus, subway, train is not ah is not the objective.*  
L2: *can you correct this sentence for me?* |
| 4. Admission of difficulty | L1: *I think our teacher used to talk to us that we should ah shouldn't rewrite the [essay] question.*  
L2: *yes I remember I know it but ah I have no way to rewrite [paraphrase].* |
| 5. Explanation | L1: *I didn't see your conclusion, yeah I mean that you have you did not write the concluding paragraph.*  
L2: *I agree with you that I lack ah the conclusion paragraph however ah this ah this writing is just ah I have only 15 minutes to write.* |
| **II. Partial acceptance** | |
| 1. Agreeing with one part and disagreeing with another | L1: *just a quick look for example each line has 1,2, 4, 5,6 words, just only 10, 11 lines so it's just about 150 words. that's why there's not enough information here to support for ah () your idea.*  
L2: *I agree because I ah drop out the ah last part* |
but disagrees that she lacked supporting ideas.

which is conclusion. I'm sorry for that but ah I DON'T agree with you that ah the reason lack of supportive ah supporting ideas. there are supporting ideas for my thesis statement

2. Offer of an alternative to S’s suggestions
L1 suggests L2 put a full stop.
L2 suggests another way to correct the sentence.

L1: I think ah you can make a full stop here.
L2: no. maybe we don't need the word "because"
L1: but so if you ah cross "because" so this sentence has three clause.
L2: we make a new sentence.

III. Total resistance
1. Dispute
L1 criticizes L2 for not meeting the word requirement.
L2 disagrees and challenges L1
L2: <do you count it yet?>

L1: I don't think you write enough words.
L2: why don't you divide the two main points into 2 paragraphs?
L1: I think this is LONG ENOUGH ((laugh)). I think it is -it can EVEN be divided into 3 paragraphs. I think it is enough to be divided into TWO paragraphs.
L2: oh yes ah I find it interesting when I write a long paragraph in the introduction and you write a rather SHORT paragraph in the introduction, yes. you have ah 1, 2, 3, 4 paragraphs but the length of the paragraph each paragraph is NOT ah as long enough.

L1: I prefer a writing style which do not show a ah too personal idea. actually I don't like the way you write that I ah (. ) "I'm convinced about the idea" or "in my opinion". it's a bit personal because the word "I" here maybe I prefer the general (. ) general pronouns or something else.
L2: oh but I'm sorry I don't agree with you ah on the point that it's TOO personal (. ) in here because (. ) I use only one once the word "I" in the introduction to <refer to (. ) the very strong ah rectification of my personal VIEWpoint> you
4. Requesting evidence  
L1 criticizes L2 for making grammatical mistakes  
L2 requests that L1 point out where the mistakes are.

5. Opting out  
L1 criticizes L2 for not explaining his ideas clearly enough.

L2 does not signal agreement or disagreement but shifts the topic.

---

Symbol: $S =$ The speaker  $H =$ The hearer