DE GRUYTER OPEN

ACTA UNIVERSITATIS SAPIENTIAE, PHILOLOGICA, **9**, 2 (2017) 61–80

DOI: 10.1515/ausp-2017-0018

Pragmatic Perspectives on Understanding Strangers

Some Methodological Issues

Nikolett VÁRHEGYI
University of Debrecen (Hungary)
nikolettvarhegyi@yahoo.com

Péter FURKÓ
Károli Gáspár University of the Reformed Church in Hungary (Budapest, Hungary)
furko.peter@kre.hu

Abstract. The present paper approaches the theme of "understanding strangers" through discussing some of the methodological issues in interlanguage pragmatics (ILP), with special reference to Hungarian–English Interlanguage (IL) requests. Written discourse completion tasks (WDCT) were used to collect data from 20 English major university students. The CCSARP Project's 9-scale request strategies table proposed by Blum-Kulka, House, and Kasper (1989) was incorporated into the research, the proposed categories were extended by labels relating to mixed strategies and responses where no answers were provided. The structure of the paper is as follows: after a brief overview of the literature in the field of ILP with a special focus on WDCT, the validity of the methodology is highlighted through the discussion of issues relating to labelling/coding categories as well as inter-annotator (dis)agreements. By analysing and comparing utterances on the basis of our annotation output and validating the results with the aid of ReCal, we have confirmed that WDCT is a reliable and valid tool for testing ILP competence in speech acts performance.

Keywords: interlanguage pragmatics, research methodology, written discourse completion task

1. Introduction

The present paper approaches the topic of "strangers", the theme of this year's (2017) imagological conference, through discussing some of the methodological issues in interlanguage pragmatics (ILP). "Strangers" in ILP terms can be conceptualized as speakers of a language other than one's native tongue, while

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

"understanding strangers" from an ILP perspective involves developing pragmatic competence, i.e. becoming aware of and applying the norms and strategies that are necessary for successful cross-cultural communication. The relationship between ILP, which is a relatively new field of Pragmatics, and Second Language Acquisition (SLA) has become the focus of research only recently (e.g. Blum-Kulka et al. 1989, Kasper 1995, Bardovi-Harlig 1999, Schauer 2009), and within this field the amount of research carried out in terms of Hungarian–English Interlanguage (IL) requests is minimal. The reason that ILP as a new branch of pragmatics was formed is that a simple contrastive analysis of L1 and L2 pragmatics is not enough in order to take account of interlingual errors that can be associated with pragmatic competence (Selinker 1972, Kasper 1996).

This study was also informed by Bardovi-Harlig (2001: 15), who states that even learners with advanced language proficiency struggle to develop concomitant pragmatic competence, which will result in pragmatic failure (miscommunication with "strangers") and/or non-target-like performance. Even with a good command of grammatical and lexical knowledge, students face difficulty in the successful pragmatic production of speech acts, and misunderstandings can often occur as consequences of this lack of pragmatic awareness.

The purpose of the study is twofold: firstly, we would like to introduce the basic terms related to ILP and give an overview of the literature in this field of pragmatics. Secondly, after presenting the different methods of data collection with their specific advantages and disadvantages, the aim is to prove that Written Discourse Completion Test (WDCT) is a reliable and valid tool when ILP competence in speech acts performance is tested.

Analysis of collected data was done by annotation, which is making a decision for the category of requests in this instance. We want to provide the results of our research on the given dataset with some metric about how certain we are about the annotation. This is where the importance of inter-annotator agreement lies. Inter-annotator agreement is a measure of how well two (or more) annotators can provide evidence for making the same annotation decisions for certain categories. To put it simple, the validity of research can be proven on statistical grounds by using an inter-annotator tool. Two statements can be concluded from the measures – if the annotators agree on most of the cases, then the categories are clearly defined, and it also reveals the trustworthiness of the given annotation from a quantitative point of view.

Hence, the research problem is to find out how much we can trust previous results by single annotators and how the validity of research can be improved if there is more than one annotator. We will establish evidence with the help of an analysis and comparison of data results from two expert annotators. Therefore, this research has produced novel results in the field of ILP – firstly, on account of the fact that previous studies (Blum-Kulka et al. 1989, Kasper 1996, Bardovi-Harlig

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

2006, Szili 2002, Szili, Bándli & Maróti 2016) collected and analysed data on the basis of single annotators, and the reliability of the results was not tested with the help of statistical multi-user annotator tools (cf. Furkó: forthcoming); second, to our knowledge, there is no previous study in the field of ILP where ReCal (short for Reliability CALculator) was used as a statistical measure for testing inter-annotator agreement. This online inter-agreement tool calculates the following indices for nominal-level variables: percent agreement, average pairwise percent agreement if three or more annotators upload their results, Scott's Pi, Cohen's Kappa, and Krippendorff's Alpha. See further details in the research section.

As for the number of annotators, it can be observed that multiple annotators were present in many international research studies (cf. Liu 2006); however, in Hungary, this method is rare if present at all (cf. Furkó et al. forthcoming). The majority of studies in Hungary are based on the findings of a single annotator, i.e. researcher who categorizes pragmatic strategies on the basis of a pre-existing taxonomy (cf. Szili et al. 2016). But there are several reasons for having more than one researcher annotating the same data set. First of all, it can easily reveal marginal cases highlighted by inter-annotator disagreements as well as point out fuzzy boundaries between categories. Secondly, when requests are studied, it is important to differentiate between the head act and the supporting move because we cannot be sure with previous cases of the CCSARP project: for example, whether the researcher identified and categorised the head act consistently across the different items in their data sets. In our present study, we have revealed that multiple annotators could also disagree on which element to categorise as head act and which one as the supportive move in some cases (see further details in the results). The CCSARP project (Blum-Kulka et al. 1989) involved observations of two speech acts, namely requests and apologies in eight language varieties with data collected from both native and non-native speakers analysed by single annotators. Situational variability, cross-cultural variability, individual, native versus non-native variability were observed back in the days. As opposed, our research focuses on requests in EFL produced by Hungarian–English major university students.

2. Literature review

2.1. Interlanguage pragmatics

Selinker (1972) was the very first to use the term "interlanguage" (IL), and it can be given account of as the L2 learners knowledge of the target language, which features characteristics of L1, L2, and other languages the person speaks; furthermore, autonomous properties can be observed in IL, too. Kasper (1995)

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

was amongst the first researchers who used and defined interlanguage pragmatics as the convergence of pragmatics and the study of SLA. ILP can be defined as the performance, acquisition, and production of L2 speech acts (Ellis 2003, Kasper and Dahl 1991). Leech (1983) and Thomas (1983) established that pragmatic comprehension is of a double nature: it can be divided into pragmalinguistics and sociopragmatics. On the one hand, pragmalinguistics is concerned with how the linguistic meaning of conveying illocutionary force and politeness is formed. On the other hand, sociopragmatics deals with culture by observing socially adequate and appropriate linguistic behaviour. L2 language users draw conclusions of utterance meaning with the help of more linguistic cues than contextual information despite the fact that they have access to non-literal pragmatic meaning of utterances. In addition to this, it was observed that different social degrees and approaches are used by native and non-native speakers on a target language (cf. e.g. Blum-Kulka et al. 1989, Szili et al. 2016), which also shows a variety of cross-cultural differences from a sociopragmatic view.

ILP can give account of the culture-specific variations when observing and analysing interaction during speech acts. The differences amongst cultures can be identified both within the speaker's production and the hearer's reaction. It is wrong then to presuppose that based on traditional pragmatic methodological studies behavioural norms could be linking, the study of cultures and cultural background are essential (Wierzbicka 1991, 2003).

To summarise, ILP is the study of *strangers* "comprehension, production, and acquisition of linguistic action in L2, or, put briefly, ILP investigates how to do things with words in a second language" (Kasper 1995: 184). With its bilateral nature, ILP can be seen as part of IL studies deriving from SLA research and also as a discipline under pragmatics; "ILP is a relatively young area in linguistics that originated from pragmatics theory and developments in L2 pedagogy and research in the 1970s", and, as Schauer (2009: 15) continues, "it uses pragmatic theories, principles and frameworks to examine how foreign/second language learners encode and decode meaning in their L2". ILP, when it is seen as part of the study of L2 use, is concerned with the comprehension and production of speech act in a given second language. ILP, when it is seen as the study of L2 learning, is dealing with "how L2 learners develop the ability to understand and perform action in a target language" (Kasper & Rose 2002: 5).

2.2. Pragmatic awareness

Pragmatic awareness can be defined as awareness of target-language sociolinguistic and sociocultural features or simply as the knowledge about pragmatics underlying appropriate language use. It plays a crucial role when acquiring a foreign language (cf. Takahashi 2005). Takahashi (2005) revealed some relationships between the

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

pragmalinguistic awareness of Japanese EFL learners and their motivation and proficiency level. His ILP research has proved that implicit pragmatic instruction helped some but not all learners to notice the target pragmalinguistic features. He also tried to identify the individual difference (ID) variables, which may enhance the noticing of pragmalinguistic features. To sum it up, Takahashi's major findings are: 1. there was a great difference among learners in terms of their noticing the TL pragmaliguistic features, 2. motivation is an important factor in this procedure but proficiency is not, 3. intrinsic motivation is in close relation with pragmatic awareness.

Now, let us see the relation between awareness and attention. Furthermore, Tomlin and Villa (1994) maintain that attention involves: alertness (which is general and related to motivation), orientation (which directs input of alertness to the sensory system), and detection (which is the cognitive side of input stimuli). These three are separable, and it can be concluded that awareness is not part of attention. Simard and Wong 2001 (in Takahashi 2005) argue that this model is not relevant to SLA. All in all, direct or indirect awareness is a key element in SLA.

As we could see from the above mentioned studies, motivation, which is one of the ID variables, is a factor highly affecting L2 pragmatic attention and awareness. Furthermore, proficiency as an affective factor can be seen as an independent operator, although it is true that highly motivated learners at a higher proficiency level may be more aware of pragmalinguistic features than learners with lower motivation at a lower proficiency level.

2.3. Pragmatic competence

ILP is concerned with strangers' communication in a second language. Since Canale and Swain (1980) and Canale (1983) redefined the components of Hymes's (1972) communicative competence as opposed to competence being exclusively grammatical as defined by Chomsky (1965), the notion of communicative competence comprises grammatical, pragmatic, discourse, and strategic competence. Studies either examine each particle on its own (in terms of its development), or they examine how these components are related to each other and whether any of them is a prerequisite for the development of the other or others (cf. Hymes 1972, Kasper 1992).

Furthermore, Shauer points out another issue: whether or not there is interdependence between grammatical and pragmatic (also referred to as sociocultural) competence. The findings observed (Schauer 2006) suggest that these two types of competences are independent; furthermore, ILP studies prove that a higher proficiency level does not correlate with better, more native-like pragmatic ability. Kasper (1992) thoroughly examines cases when grammar supersedes pragmatics, and then she gives scenarios where pragmatic knowledge

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

precedes the grammatical one. As shown by the examples above, pragmatic development can be studied either from the point of view of information processing or from a socio-cognitive theoretical view.

2.4. Assessing IL pragmatic awareness and pragmatic competence

Based on the findings of Schauer's (2006) study, both pragmatic awareness and productive pragmatic competence advance during the time strangers spend in the target-language environment. EFL learners are outperformed by ESL learners in terms of their ability to notice pragmatic violations, and the ESL group scored as high as the native speaker control group. Internal and external modifiers can be differentiated in speech acts, too. Internal modifiers can be given account of "as linguistic and syntactic devices that modify the illocutionary force of the request, such as politeness marker 'please' or the downtoner 'maybe'" (Schauer 2009: 28). If the internal modifier increases the illocutionary force of the request, we can speak about an upgrader, whilst if it decreases the illocutionary force, then it can be referred to as a downgrader. External modifiers can be referenced as supportive moves, which support the request with additional statements. As an example, a grounder is the reason for a request for it to be carried out. The research also confirmed that internal modifiers are acquired earlier than external modifiers, which were the findings of previous studies, too. ILP development is influenced by the duration of stay in a L2 environment and by individual learner differences. Further studies are needed to explore how and to what degree factors such as learners' motivation, time spent with native speakers, educational background, and proficiency level influence pragmatic development. Learner diaries could reveal qualitative data about learners' usage of acquired knowledge. Future studies with focus on the investigation of teaching methods, approaches, activities, and learning goals could also reveal what enhances better pragmatic awareness.

2.5. Pragmatic transfer

"Transfer" as a term refers to influences from languages other than the L2. "The study of transfer involves the study of errors (negative transfer), facilitation (positive transfer), avoidance of target language forms, and their over-use" (Ellis 2003: 341). Pragmatic transfer, as defined by Kasper (1995), comprises all the influences on L2 comprehension, production and acquisition of pragmatic information that are foregrounded by a particular learner's pragmatic knowledge of languages and cultures other than L2 (in this very specific case: English). Moreover, it is proved that contrastive analysis of L1 and L2 pragmatics is not enough to take account of pragmatic transfer, and interlanguage pragmatics is proposed to figure out the possible reasons for this phenomenon (cf. Blum-Kulka 1989, Schauer 2006, Szili et

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

al. 2016). It is possible to differentiate positive and negative transfer of pragmatic nature. Positive transfer (facilitation) in connection with request occurs when a non-native speaker uses a request strategy which is not a universal one but a pragmalinguistic feature which can be attributed to their mother tongue. Requests in pragmatic universals can be performed directly, conventionally indirectly, and indirectly. For example, when the past tense of the "could/would" modal verbs in the request of "Could/Would you lend me your notes, please?" from L1 Danish (kunne/ville) and L1 German (könntest/würdest) and languages other than English language is transferred, then it is a great example of positive transfers (Blum-Kulka et al. 1989). Negative transfer (interference) occurs when the applied structure or expression is applicable in L1 but not perfect in L2. Kasper (1995) observed that the syntactic strategy of modal verb + interrogative + negation used in English for a request "can't you clean the kitchen" is derived from the Danish structural expression.

As far as communicative effect is concerned, negative pragmatic transfer is not equal to pragmatic failure or miscommunication, although it can lead to it. Also, pragmalinguistic divergence, which is often regarded as a linguistic problem, is not as serious as a sociopragmatic one, which can be understood as a means of conveying bad manners or being rude (Thomas 1983, in Kasper 1995). As a conclusion, L1 pragmatic preference patterns greatly influence L2 performance.

3. Data collection

3.1. Options for data collection

In an ideal world, researchers would be able to examine an oral linguistic phenomenon as it appears in the given context within the given cultural environment by taking into account the speaker's intentions and the full process of the interpretation on the part of the hearer. As it is hardly possible, there are quite a few options for different empirical research methods in speech act studies, namely ethnographic fieldwork, role-plays, discourse completion tests, and multiple choice questionnaires (cf. e.g. Brown 2001, Szili et al. 2016).

Data collection can be done in the form of ethnographic fieldwork, when "living and breathing" amongst and with the observed community is part of the everyday life of the researcher, which takes up a long time for a single study. Naturally, it takes long to assimilate into the culture of the observed individuals, if it is possible at all. Further issues related to the field notes method, for example, are recording the speech situation and noting down the exact words from memory, as fieldworkers do not use tapes or video recording in order to ensure a naturalistic setting for the research. Despite the limitations of this method, many speech act

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

studies have used it – for example, Olshtain and Cohen 1983, Trosborg 1987, and Holmes 1990, who have studied the speech act of apology.

Role-plays (cf. e.g. Usó-Juan & Martínez-Flor 2006, Kasper & Dahl 1991) are situations where participants act and behave according to their given script. The best scenario is when they can be themselves and behave as they would normally do in a pre-described situation given by the researcher. Actual spoken data can be produced and observed through the use of this method. Although longer responses are collected in these life-like situations, the participants might not act naturally as they might feel pressured to do so. According to one classification, there are two types of role plays (RPs), namely open and closed ones according to the extent of interlocutors' interactions with each other. A closed RP is usually a one-turn conversation and is set in a specific situation (Martinez-Flor 2006), whilst an open one might take several turns and probably represents a more natural discourse with the help of a full operation of the dynamic turn-taking mechanism and more authentic setting (Martinez-Flor 2006, Kasper & Dahl 1991). On the one hand, the researcher could be in control of specific variables such as gender relationships between interlocutors, situations, power status and distance of participants, and other social factors, too. On the other hand, the observer can also make the interactions rather artificial in nature, especially when the data collection is predominantly motivated by the researcher's goals rather than those of the interlocutors (Kasper 2000).

Tran (2006.) developed his own type of this method and named it the Naturalized Role-Play (NRP), in which the participants are firstly involved in a natural conversation, but in the middle of the conversation the researcher guides the participants to a specific act in which she or he is to produce the required speech expressions – and the respondents are not aware that they are being observed. In this way, the observed speech act can be elicited and observed in a natural-like setting, while the method also involves careful planning and research design.

Another method is using multiple choice questionnaires, which could provide a large amount of data with quick annotation, as the assessed individuals only need to choose their preferred response from a given set of answers. Obvious disadvantages include the fact that spoken discourse is represented in a written form and that the small number of choices mean limited insights for the researcher. As Maróti (2016) concludes, "intralingual research into language behaviour may aim to explore […] the diverse language usage […] as well as to conduct empirical studies […] to analyse […] communicative failure among native speakers".

Finally, written discourse completion tests (DCTs) explain the given situations to the research subjects followed by some space to write their answers. It is an appropriate method for the collection of a large amount of data in a short period of time and the comparison of native and non-native speakers in a highly controlled research design. On the other hand, however, it is questionable if "oral linguistic

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

reactions could be represented in writing; also, whether the validity of data is … influenced…by…space" (Maróti 2016: 31). Mey (2004), amongst others, criticised the use of WDCTs due to construct validity. However, in the present paper, Kasper and Rose's perspective (2002: 95–96) will be adopted, who state that "when carefully designed, WDCTs provide useful information about speakers' pragmalinguistic knowledge of the strategies and linguistic forms by which communicative acts can be implemented and about their socio-pragmatic knowledge of the context factors under which particular strategic and linguistic choices are appropriate". In sum, WDCT is a pragmatic tool which requires in this case strangers to read a written description of a variety of situations with highly controlled contextual parameters, such as participant roles and the degree of imposition, in order to write down what would be said in that given situation to represent an oral-like SA.

Therefore, in our research, we treat the choices of the speakers being observed as pragmatically appropriate linguistic forms and not as actual discourse episodes. Many researchers have applied this form of data collection [cf. e.g. Blum-Kulka et al. 1989, Takahashi & Beebe 1993, Kasper & Rose (1999, 2002), and Liu (2006a, b) amongst others]. Kim (2007) also studied DCTs, and based on Beebe and Cummings' (1985) work he concluded that they are highly effective means of collecting large amounts of data, where semantic formulas and strategies could be observed in natural-like speech production and certain social and psychological factors affecting speech act performance can be observed. However, Kim suggests that other methods of data collection, such as roleplays, could reveal more about the pragmatic performance in communicative context in a more near-authentic setting. Also, elicited conversations could be used to observe the targeted speech acts in higher frequency over a limited time. Xu and Wannaruk (2015) analysed the reliability and validity of WDCT based on the Many Facets Rasch Model. Their research results proved that WDCT scored high both on reliability and validity when interlanguage pragmatic competence in speech act performance was tested. Now, let us move onto our research agenda.

4. The research

4.1. Methodology

This paper presents a cross-sectional study (cf. Cook 1993, Rose 2000, Kasper & Rose 2002) providing information on the pragmalinguistic features of advanced English as a Second Language learners' pragmatic performance in 3 different scenarios with 3 different status situations. As far as the timeframe of data collection is concerned, students were given an hour to finalize their answers, which was more than enough time, and they did not feel rushed at all.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

4.2. Participants

Data was elicited from 20 respondents in altogether 177 cases in order to investigate L2 linguistic proficiency and the acquisition of pragmatic competence among advanced Hungarian learners of English with the help of WDCT tasks, with a more detailed description to follow. As previously mentioned, WDCT can be recognized as a reliable and very effective, time-efficient method of pragmatic data collection, which will also be proven by results later on (cf. e.g. Blum-Kulka et al. 1989, Rose & Kasper 2001, Ellis 2003). In this way, the researcher had full control over the language, the context, and the social distance in the individual items of the data set. Participants were between the ages of 18 and 26 with the vast majority of them aged 19–21, who had been studying English as a second language mostly for 10–12 years but had not been exposed to direct pragmatic instruction or courses before the study took place. They are advanced students of English, many of them attained A-Level GCSE and/or C1 accredited language exams, and also all passed the Karoli Gaspar University's advanced yearly yardstick exam in the previous academic year of 2015/2016. The results showed that mood derivable, hedged performative, and preparatory strategies were predominantly used by students in their native Hungarian language, while other strategies were rarely present, which correlates with previous research findings (Szili 2002). However, the negative transfer of these strategies results in more directness in requests during EFL communication than native speakers would have. Szili's (2002) findings provided a good starting point in order to elicit further request strategies used by Hungarians – especially in their native language –, while the frame of reference for the native English data was provided by Blum-Kulka et al. (1989).

4.3. Annotators

The collected data was analysed by two expert annotators, i.e. scholars who each had at least four years of experience with discourse-pragmatic annotation. The male annotator is also a university associate professor, while the female annotator is a final-year PhD student of English linguistics. As for their education around the request categories, they both carried out an extensive research on previous works written by Blum-Kulka et al. (1989), Kasper (2001), and Szili (2002) and heavily used the CCSARP coding manual, the examples of which served as points of reference. The annotators then uploaded their results onto the ReCal online system to assess inter-annotator agreement (further details to follow below). Knowledge of ReCal has been attained using online resources as well as relying on the annotators' previous experience.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

4.4. Materials – the data-collecting tool (based on the CCSARP Project and Szili's previous works)

As we have mentioned previously, the CCSARP project involved observations of two speech acts, namely requests and apologies in eight language varieties with data collected from both native and non-native speakers. Situational variability, cross-cultural variability, individual, native versus non-native variability were observed back in the days. For this present study, the data collecting instrument was an adaptation of Blum-Kulka et al.'s (1989) request strategies WDCT situations, also used in Szili's works (2002, 2004). The questionnaire situations were slightly changed and adapted to the modern age and tailored for the audience of English major students. When assessing learners' knowledge, different social variables such as power status, distance, and degree of imposition were taken into account and were varied in each of the six different situations. Models of analysis designed by Blum-Kulka et al. (1989) and Szili (2002) were combined, upgraded, and complemented with modern generational needs.

Brown and Levinson (1987) differentiated among Face-Threatening Acts (FTAs) as being off record or on record. With regards to on record, it can be done without redressive action, baldly or with redressive action either as a positive or negative politeness strategy. Hence, during the data collection, we have included questions about the informants' sex, age, length of English learning history, etc. The discourse completion task comprised 3 different scenarios: requesting for pen and paper, borrowing a mobile phone, lecture notes requests. The form was designed to elicit on requests at 3 different levels: students were to interact with their best friend, another (less familiar) student, or their teacher/professor; hence different levels of power status were observed (on an equality-inequality scale). After a short description of the situation involving setting the scene, participants were asked to formulate requests in English in 3 different ways based on familiarity and social power relations, respectively. The questionnaire was designed to assess students' pragmatic competence and shed light on differences between native-like and non-native-like speech acts of requests.

The table below shows the coding categories and provides examples.

Table 1. *Request descriptive categories based on Blum-Kulka et al. 1989*

Descriptive category	**Examples**
1. Mood Derivable	Clean up the kitchen.
2. Performative	I'm asking you to move your car.
3. Hedged Performative	I would like to ask you to move your car.
4. Obligation Statement	You'll have to move your car.
5. Want Statement	I would like you to clean the kitchen.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

Descriptive category	Examples
6. Suggestory Formulae	How about cleaning up?
7. Query Preparatory	Could you clean up the mess in the kitchen?
8. Strong Hints (A)	
9. Mild Hints (B)	We don't want any crowding (as a request to move the car).
10. Mixed	
11. No answer	

4.4.1. ReCal

ReCal (short for Reliability CALculator) is an online inter-agreement calculator. The greatest advantages of this system-independent operating program are that its data specification is general and that it offers more reliability coefficients than any other competing products. The software manual is also available online and is complemented by research articles and scholarly documentations by the authors on their website, which greatly helps annotators to understand the software and how to interpret results. Intercoder, interrater, or inter-annotator reliability can be given account of as "the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion" (Lombard, Synder-Duch, and Bracken 2002). ReCal calculates the following indices for nominal-level variables: percent agreement, average pairwise percent agreement if three or more annotators upload their results, Scott's Pi, Cohen's Kappa, and Krippendorff's Alpha.

In content analysis, reliability is usually estimated using the above-mentioned four estimation tools (Lombard et al. 2002). Percentage agreement is easily calculated as a percentage of agreements (1 – perfect agreement code) and non-agreements (0 – non-agreement code). Chances of agreement (due to its dual variability of 0 and 1 values) were challenged by Krippendorff (2004). To measure the chance of occurrences, Cohen (1960: 96) developed kappa, which "derives the expected agreement by chance using multiplicative terms of the marginal distributions of coders, assuming differences in the distribution of values across the categories for different coders". Scott's Pi is also a statistic to measure inter-rater reliability, which gives an overestimated calculation for cases in which the raters actually had an agreement (Lombard et al. 2002). "While Scott's Pi can calculate the index with two different groups of coders and different numbers of coding items, Cohen's Kappa assumes that two raters each code all content items" (based on Craig 1981, in Lombard et al. 2002). Finally, Krippendorff's Alpha could be seen a more flexible statistic, which "can be used regardless of the number of coders, type of measurement, amount of data, and presence of missing data" (Krippendorff 2004: 13). As we have seen from the involvement of

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

several statistical measures above, ReCal can be used in research to validate data based on pure statistics.

Another part of the results calculated by ReCal lists number of agreements, disagreements, cases, and decisions. As for the definition of these terms, "N Agreements" (number of agreements) is the nominal for agreements out of all cases, while N Disagreements shows the number of mismatches in the total number of observed cases with the heading of N Cases, and the number of decisions by all annotators are shown under the heading of N Decisions. In the present study, 177 cases (i.e. 177 items coded by both annotators) were observed during the study; thus, the 2 coders made a total of 354 decisions (shown under "N Decisions"). 149 agreements and 28 disagreements were found. Further details will follow later.

5. Results

As we mentioned in the previous section, ReCal was used to calculate inter-annotator agreement, which yielded the values presented in *Table* 2 below. The software allows multiple annotators to work on the same file, and therefore it is able to measure inter-annotator agreement. As the table shows, even though there is no perfect agreement between the two expert annotators (Scott's Pi & Cohen's Kappa < 0.8), the coding scheme appears to be reliable on the basis of the random sample.

Table 2. *Inter-annotator agreement on request strategies*

	Percent Agreement	**Scott's Pi**	**Cohen's Kappa**	**Krippendorff's Alpha (nominal)**	N Agreements	N Disagreements	N Cases	N Decisions
Variable 1 (cols 1 & 2)	84.2%	0.627	0.635	0.628	149	28	177	354

Our observations concluded that most of the cases where inter-annotator disagreements occurred fall into the following categories:

(1) type 1
Annotator 1: strategy 7, Annotator 2: strategy 8
Example:
(a) Excuse me, I happened to hear that you will be late from a meeting. Actually, I'm in the same situation as well, but my battery is dead, so would it be a problem if I made a quick call from your phone?
(b) Hi! We go to the same lecture, but I couldn't attend yesterday. Could you please, lend me the notes you made? (If you don't need them right now.)

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

(c) Good morning! I'm sorry for being absent yesterday, but something urgent came up. Would it be a problem if I asked for the lecture slides from that lesson? I'd rather learn the material from those.

In these cases, Annotator 2 observed a pre- or post-sequence as the head act or felt that the head act serves as a strong hint rather than a query preparation. As a supposition, in real life, the second part would not be necessary to mention, and the request would be fulfilled after presenting the strong hint part of the speech act.

(2) type 2

Annotator 1: strategy 7, Annotator 2: strategy 8

Example:

I missed a lecture yesterday, so I don't have the notes. Could I borrow them for a few minutes?

Excuse me, Mr Teacher. My phone is dead and I need to make a call. May I use yours?

In these instances, Annotator 2 observed the pre-sequence as the head act and noted that the second part is only a duplicate of the request and not being essential for a request to be completed.

(3) type 3

Annotator 1: strategy 1 or 7, Annotator 2: strategy 8

Example:

(a) Do you have an extra pen and paper?

(b) Do you have a pen and paper for me?

In these examples Annotator 1 either chose mood derivable or query preparatory, while Annotator 2 chose strong hints in both cases. The category chosen by the second annotator agrees with the choice made by Wang in his research (2011), where he categorised "Do you have any spare screw like this one?" as a strong hint.

(4) type 4

Annotator 1: strategy 3, Annotator 2: strategy 8

Example:

(a) Good morning, Professor! I'm in an inconvenient situation. I cannot use my phone, but I really need to make an important phone call. Can I kindly ask you to borrow me your phone, please?

(b) Hi, I'd just like to ask you whether you have notes from yesterday's class? If so, could you give them to me? Thank you.

These cases show that Annotator 1 identified the final part of the utterance as head act and put it into the category of hedged performative, whilst Annotator

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

2 used the first part of the speech act and categorised it as a strong hint which continued with an additional hedged performative sequence.

(5) type 5
Annotator 1: strategy 1, 3, 7, Annotator 2: strategy 8
Example:
(a) Sorry, madam, the battery of my phone is dead, and I couldn't stop but overheard that you're being late and that's the same situation with me, so could you be so kind to lend me your phone so that I can make a call, please?
(b) Excuse me, sir. The battery of my phone died, and I have to call my schoolmates that I'm being late. Could I borrow your phone, please?
(c) Excuse me! I'm a student at the university you're teaching at, and I'm late from a very important school meeting. I would call someone to inform them about my situation, but my phone is dead. Would you be so kind to let me borrow your phone and make a phone call?
(d) Hi, I'd just like to ask you whether you have notes from yesterday's class. If so, could you give them to me? Thank you.

These examples show that Annotator 1 only used strong hint as a category once, while Annotator 2 identified it 26 times, which is due to the annotators' different points of view on the head act in the instances or the presence of multiple categories at the same time.

The following charts present the request strategies identified by the two annotators, and can serve as points of reference for future research. The categories are based on the CCSARP categories which are listed and described in table 1 above.

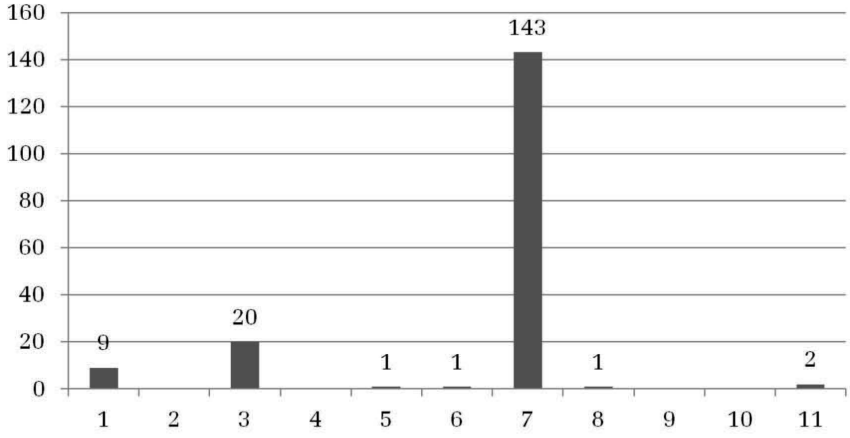


Chart 1. *Output of Annotator 1 (the request descriptive categories are listed and exemplified in table 1 above)*

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

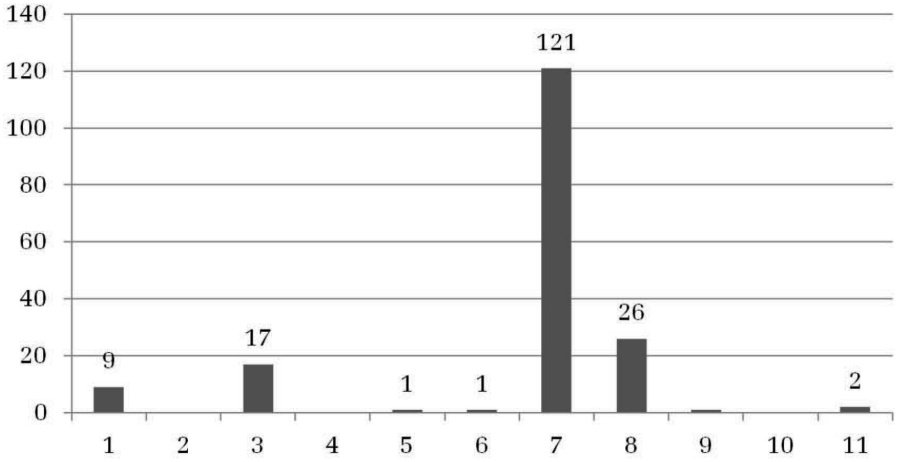


Chart 2. *Output of Annotator 2 (the request descriptive categories are listed and exemplified in table 1 above)*

6. Conclusions

As seen from the annotation and the examples above, the categories used in the CCSARP project (1989) can still be seen as valid; however, they did not contain the sociological variables of social distance, power, and absolute ranking at that time (Brown & Levinson 1987). For the present paper, the CCSARP coding manual (Blum-Kulka et al. 1989) was used to classify research data, and ReCal was used to validate results. Taking a closer look at the effects of sociological variables is a topic of future research.

Another key element observed from the point of methodology is that it is essential to differentiate between the head act and the supporting move, and it needs to be observed whether the annotator identifies the two correctly or whether there is a disagreement between the annotators with respect to the order of the head act and the supportive move. In many cases, a complex annotation is suggested to categorise the request based on the characteristics identified in both pre- and post-head acts. Also, the final decision on the identification of the head act and the supportive move would be based on a larger context (and co-text), which would provide more understanding of the intentions of the speaker in a real-life situation.

To sum it up with, in the paper, we have first looked at the literature of ILP, and then the different methods available for research in interlanguage pragmatics were presented. Moreover, by way of analysing and comparing utterances on the basis of two expert annotators' output and validating the results with the aid

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

of ReCal, we have confirmed that WDCT is a reliable and valid tool for testing ILP competence in speech acts performance. The next step for future research would be to compare performance and realization of requests as speech acts with data originating from native speakers. The present paper has also highlighted the importance of the relationship between interlanguage pragmatics and second language acquisition, which has been a recently researched area only. Furthermore, the different data collection methods used in ILP were given account of by exemplifying research carried out in the field and elaborating on the advantages and disadvantages of each method. The overview also highlights the reason why ILP was born: contrastive analysis of L1 and L2 pragmatics is not enough to take account of IL errors in the field of pragmatics.

References

Bardovi-Harlig, Kathleen. 1999. Exploring the Interlanguage of Interlanguage Pragmatics: A Research Agenda for Acquisitional Pragmatics. *Language Learning* 49: 677–713.

—— 2001. Evaluating the Empirical Evidence: Grounds for Instruction in Pragmatics. In Gabriele Kasper–Kenneth Rose (eds), *Pragmatics and Language Teaching*, 13–32. Cambridge: Cambridge University Press.

Beebe, Leslie M.–Martha C. Cummings. 1996. Natural Speech Act Data Versus Written Questionnaire Data: How Data Collection Method Affects Speech Act Performance. In Susan M. Gass– Joyce Neu (eds), *Speech Acts across Cultures: Challenges to Communication in a Second Language*, 65–86. Berlin: Mouton de Gruyter.

Blum-Kulka, Soshana–Juliane House–Gabriele Kasper, eds. 1989. *Cross-Cultural Pragmatics.* Norwood, New Jersey: Ablex.

Brown, James. 2001. Pragmatics Tests: Different Purposes, Different Tests. In Kenneth Rose–Gabriele Kasper (eds), *Pragmatics in Language Teaching,* 301–326. Cambridge: Cambridge University Press.

Brown, James–Stephen Levinson. 1987. *Politeness. Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Canale, Michael. 1983. From Communicative Competence to Communicative Language Pedagogy. In Jack C. Richards–Richard W. Schmidt (eds), *Language and Communication*, 2–27. London: Longman.

Canale, Michael–Swain, Merrill. 1981. A Theoretical Framework for Communicative Competence. In Palmer, A.–Groot, P.–Trosper, G. (eds), *The Construct Validation of Test of Communicative Competence*, 31–36. London: Longman.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The M.I.T. Press.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

Cohen, James. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20(1): 37–46.

Cook, Vivian. 1993: *Linguistics and Second Language Acquisition.* London: Macmillan.

Craig, Richard T. 1981. Generalization of Scott's Index of Intercoder Agreement. *Public Opinion Quarterly* 45(2): 260–264.

Furkó, Péter–András Kertész–Ágnes Abuczki [forthcoming]. Discourse Markers in Different Types of Reporting. In Alessandro Capone (ed.), *Indirect Reports – Perspectives in Pragmatics, Philosophy and Psychology.* Heidelberg: Springer Verlag.

Ellis, Rod. 2003. *The Study of Second Language Acquisition.* Oxford: OUP.

Hymes, Dell. 1972. On Communicative Competence. In James Bruce Pride–James Holmes (eds), *Sociolinguistics,* 269–293. Harmondsworth: Penguin.

Kasper, Gabriele. 1995. Interlanguage Pragmatics. In Jef Verschueren–Jan-Ola Ostman–Jan Blommaert (eds), *Handbook of Pragmatics,* 1–7. Amsterdam: John Benjamins Publishing.

—— 2000. Four Perspectives on L2 Pragmatic Development. *Annual Conference of the American Association of Applied Linguistics (AAAL)*, 1–46. Vancouver.

Kasper, Gabriele–Soshana Blum-Kulka. 1993. *Interlanguage Pragmatics.* New York: Oxford University Press.

Kasper, Gabriele–Kenneth R. Ros. 2001. *Pragmatics in Language Teaching* Cambridge: Cambridge University Press.

Kasper, Gabriele–Marta Dahl. 1991. Research Methods in Interlanguage Pragmatics. *Studies in Second Language Acquisition* 13: 215–247.

Kasper, Gabriele–Kenneth Rose R. 1999. Pragmatics and SLA. *Annual Review of Applied Linguistics* 19: 81–104.

—— 2002. *Pragmatic Development in a Second Language*. Oxford: Blackwell.

Kim, Kyung S. 2007. A Study of the Problems of Discourse Completion Tests. *English Teaching* 62(4): 241–256.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, CA: Sage.

Leech, Geoffrey N. 1983. *Principles of Pragmatics*. London: Longman.

Liu, Jianda. 2006a. *Measuring Interlanguage Pragmatic Knowledge of EFL Learners*. Frankfurt am Main: Peter Lang.

—— 2006b. Assessing EFL Learners' Interlanguage Pragmatic Knowledge: Implications for Testers and Teachers. *Reflections on English Language Teaching* 5: 1–22.

Lombard, Matthew–Jennifer Snyder-Duch–Cheryl C. Bracken. 2002. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. *Human Communication Research* 28: 587–604.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

Maróti, Orsolya. 2016. Research Methods in Speech Act Studies. In Szili, Katalin–Judit Bándli– Orsolya Maróti (eds), *Pragmatics in Practice*, 29–36. Budapest: ELTE.

Mey, Jacob. 2004. Between Culture and Pragmatics: Scylla and Charybdis? The Precarious Condition of Intercultural Pragmatics. *Intercultural Pragmatics* 1(1): 27–48.

Rose, Kenneth R. 2000. An Exploratory Cross-Sectional Study of Interlanguage Pragmatic Development. *Studies in Second Language Acquisition* 22(1): 27–67. Cambridge: Cambridge University Press.

Schauer, Gila. 2006. The Development of ESL Learners' Pragmatic Competence: A Longitudinal Investigation of Awareness and Production. In Kathleen Bardovi-Harlig et al. (eds), *Pragmatics and Language Learning*, 135–163. Hawaii: Second Language Teaching and Curriculum Center University of Hawaii.

—— 2009. *Interlanguage Pragmatic Development.* New York: Continuum International Publishing Group.

Scott, Walter. 1955. Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly* 17: 321–325.

Selinker, Larry. 2009. Interlanguage. *IRAL – International Review of Applied Linguistics in Language Teaching* 10(1–4): 209–232.

Simard, Daphnée–Wong, Wynne. 2001. Alertness, Orientation and Detection: The Conceptualization of Attentional Functions in SLA. *Studies in Second Language Acquisition* 23: 103–124.

Szili, Katalin. 2002. A kérés pragmatikája a magyar nyelvben [The Pragmatics of Requests in the Hungarian Language]. *Magyar Nyelvőr* 126(1): 12–30.

—— 2004. A bókra adott válaszok pragmatikája. [The Pragmatics of Compliment Responses.] *Magyar Nyelvőr* 128(3): 165–185.

Szili, Katalin–Judit, Bándli–Orsolya, Maróti. 2016. *Pragmatics in Practice – Empirical Studies in the Hungarian Language.* Budapest: ELTE.

Takahashi, Sugao. 1993. Transfer in Interlanguage Pragmatics: New Research Agenda. *Studies in Languages and Cultures* 11: 109–129.

Thomas, Jenny, 1983. Cross-Cultural Pragmatic Failure. *Applied Linguistics* 4(2): 91–112.

Tomlin, Russell S.–Victor Villa. 1994. Attention in Cognitive Science and Second Language Acquisition. *Studies in Second Language Acquisition* 16(2): 183–203.

Tran, Giao Quynh. 2006. The Naturalized Role-Play: An Innovative Methodology in Cross-Cultural and Interlanguage Pragmatics Research. *Reflections on English Language Teaching* 5(2): 1–24.

Usó-Juan, Esther–Alicia Martínez-Flor. 2006. Current Trends in the Development and Teaching of the Four Language Skils. Berlin: Mouton de Gruyter.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)

Xu, Lan–Anchalee Wannaruk. 2015. Reliability and Validity of WDCT in Testing Interlanguage Pragmatic Competence for EFL Learners. *Journal of Language Teaching and Research* 6(6): 1206–1215.

Wang, Vincent. 2011. *Making Requests by Chinese EFL Learners.* Amsterdam: John Benjamins.

Wierzbicka, Anna. 2003 [1991]. *Cross-Cultural Pragmatics: The Semantics of Human Interaction.* Berlin–New York: De Gruyter Mouton.

BDD-A27685 © 2017 Scientia Kiadó
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:58:59 UTC)