A CORPUS-BASED ANALYSIS OF THE PRAGMATIC MARKER LIKE IN NON-NATIVE CONVERSATION

Marija Kusevska
Department of English, Faculty of Philology, Goce Delchev University, Shtip, Republic of North Macedonia
marija.kusevska@ugd.edu.mk

Abstract: The study presented in this paper is part of the research project “Developing cross-cultural and interlanguage pragmatics research and its practical implications” currently implemented at Goce Delchev University in Shtip, Republic of North Macedonia. It investigates the use of the pragmatic marker ‘like’ by Macedonian learners of English. For this purpose, we compiled a research corpus of conversations produced by 76 students of English at Goce Delchev University. The participants had been asked to choose and discuss three topics out of the following five: problems with stray dogs, living and working abroad, body piercing and tattoos, the healthy amount of time to spend with the person you’re dating, and talking on the phone while sharing time with friends. The time of the conversations mounted to 9.9 hours, or 66,696 words. The conversations were then transcribed and analysed. Additionally, attitudinal data were collected from 40 of the participants about their perception of ‘like’ with respect to the age and gender of the speakers, formality of the situation, grammaticality, acceptability, distractibility, and politeness of the utterances. They also rated users for fluency and their level of English. The findings of this investigation show that the pragmatic marker ‘like’ is salient for the learners and that they use it similarly as native speakers. It also shows that its frequency correlates with language proficiency levels. However, other factors also influence its usage, such as learners’ perception of the marker, length of turns and speakers’ personal features. The present study makes an important contribution to interlanguage pragmatics. First, it investigates spoken language and reveals some aspects of learner communication that cannot be observed in class. Second, it shows that learning a foreign language is a complex process that involves not only instruction but all other resources that learners have access to through the Internet.

Keywords: interlanguage; pragmatic markers; functions; language proficiency; language corpus.

1. Introduction
All researchers who have investigated pragmatic markers (PM) agree that these elements facilitate spontaneous speech production and interaction and prevent the speaker from being seen as impolite or awkward (Crystal, 1988). The aim of this study, therefore, is to investigate how foreign language learners, with little exposure to real-life communication in the foreign language that they are learning, apply pragmatic markers to manage communication. In particular, the paper analyses the use of the PM like by Macedonian learners of English (MLE).Our
attention to it was attracted by the surprisingly high frequency of *like* in the conversations of some learners, as in the example below:

> Just *like* two months ago I was … you know we were *like* having coffee at a friend’s place and we stayed up to *like* two or three in the morning and *like* everybody left and I was *like*, ‘I live nearby’ and I’m *like*, ‘OK I’m gonna walk home you know I’m not gonna call a cab’ and there are *like* I’m going home and five dogs are, *like*…

There are eight instances of *like* as a pragmatic marker in this turn in which a learner is describing his experience with stray dogs at night. In addition to its frequency, *like* is used with a variety of different functions. Motivated by this phenomenon, this paper sets to investigate how MLE use the PM *like*. This general question breaks down into three sub-questions:

- Which functions of the PM *like* occur in native and which in non-native conversation?
- Do proficiency levels of English influence learners’ use of *like* as a pragmatic marker?
- What other factors influence the use of *like* by foreign learners of English?

### 2. Review of Literature

The PM *like* is the most common innovative feature of modern English (Andersen, 2001; Blyth, Recktenwald, and Wang, 1990; Buchstaller and D’Arcy, 2009; Tagliamonte and D’Arcy, 2004; Tagliamonte and Hudson, 1999). As it tends to be associated with southern Californian teenage talk (Valley speak; Valley girls), most of the studies of *like* deal with American data (Blyth, Recktenwald, and Wang, 1990; Fox Tree, 2006; Fuller, 2003; Jucker and Smith, 1998; Meehan, 1991; Miller and Weinert, 1995; Schourup, 1985; Tannen, 1986; Underhill, 1988). It is believed that from the US it spread to other English speaking countries as a result of extensive cross-cultural contact. Some in-depth research on *like* has also been carried out in Britain (Andersen, 2001; Cheshire, 2007; Dailey-O’Cain, 2000; Tagliamonte and Hudson, 1999), Canada (Tagliamonte and D’Arcy, 2004), as well as Australia (Winter, 2002). An often-cited study is Buchstaller and D’Arcy (2009) who compared the usage of quotative *like* in the United States, England, and New Zealand.

The PM *like* is one of those markers which are often stigmatized. In his corpus on narratives, Underhill (1988) calls it "ungrammatical in standard English". The use of *like* as in *we hitch a ride out of there with uh this like one crazy like music major guy* is considered by many "to be superfluously sprinkled into talk, a bad habit best avoided" (Fox Tree, 2006). Dailey-O’Cain’s study (2000) also shows that generally people have negative opinions about *like*. In fact, the use of *like* is perceived as both positive and negative at the same time. She also concludes that “the use of *like* tends to be associated with ‘solidarity’ traits, while the non-use of it with ‘status’ traits” (Dailey-O’Cain, 2000: 76). It is significantly more frequent between friends than between strangers and it steers the conversation to a more casual style (Fuller, 2003; Jucker and Smith, 1998; Redeker, 1990).

Markers are not only multifunctional in the sense that they can serve different pragmatic functions in different contexts, but they can also have several pragmatic features at the same time. Underhill (1988: 234) suggests that *like* functions as a marker of new information and focus as in *But then the first day of our skiing, you*
know we’re gettin all excited to go skiing the first day it’s like snowing ... blizzard snowing. The information that it brings in is not only new but also most significant in that situation. Several publications view like from a relevance-theoretic perspective. For example, Jucker and Smith (1998) see like as an information-centered presentation discourse marker which modifies the information it accompanies. In this sense, like marks a clause, or an expression, to indicate that it should not be taken too literally but only as a sufficient approximation. Andersen (2001) gives a comprehensive account of the usage of like in London teenage talk in terms of its syntactic-semantic, pragmatic, and sociolinguistic properties. Her analysis of the functions of like is also based on the theory of relevance and looks at it as a signal to the hearer that there is a non-identical resemblance relation between a linguistically encoded concept and the concept in the speaker’s thought and instructs the hearer to opt for a less-than-literal interpretation of the utterance. Beeching (2016: 133) adds that like is also used for hedging the potentially critical and emphatic stance of the speaker.

The functions of quotative like have also received significant attention (Andersen, 2001; Beeching, 2016; Blyth, Recktenwald and Wang, 1990; Buchstaller and D’Arcy, 2009; Meehan, 1991; Tagliamonte and Hudson, 1999). Belike operates as a fixed unit which may be close in meaning to verbs of saying/thinking (Andersen, 2001). It cannot be omitted in this position, and it is essential for the syntactic well-formedness of the utterance. Its function is to point to a thought produced by the speaker, or someone else, and which may or may not have been verbally realized in an utterance: and then, and then Kevin came up to me and said erm […] if you go and see Mark this afternoon erm he would like to speak to you, I was like, ‘He should come and speak to me’ (Andersen, 2001: 250). Although the quotative like may appear both in the present and in the past tense, it is more likely to convey the effects of the historical present, such as adding vivacity and immediacy, demarcating authority, or structuring discourse (Buchstaller and D’Arcy, 2009). Most of the studies on pragmatic markers focus on their functions in native-speaker conversations. In spite of the great interest in the development of pragmatic competence in foreign/second language learners, the role of the pragmatic markers in general, and of like in particular, has been largely neglected. The most extensive study to my knowledge in this field was done by Müller (2005). It focuses on so, well, you know and like, and compares how they are used by American native speakers and German learners of English. Hellermann and Vergun (2007) investigate the usage and functions of three discourse markers (like, you know, well) in the speech of beginning adult learners of English. Polat’s paper (2011) approaches the area of pragmatic markers by using a developmental learner corpus to examine pragmatic marker usage by one naturalistic adult language learner for one year. The focal markers of this study are you know, like and well.

3. Description of the Study

3.1. Participants
The participants of this study were 76 students of English enrolled at the Department of English language and literature at Goce Delcev University, Stip, Republic of North Macedonia. 28 of the participants were male and 48 were female. Their ages ranged from 19 to 25. All participants learned English for at
least five years in primary school and four years at high school. After high school, they enrolled in the English department. Very few of the students had been to an English speaking country and had had a chance to learn English in an informal environment: seven had been on *Work and travel* for two to six months in the USA and one of them had been on a student exchange programme to the USA for ten months. Their primary language of communication is Macedonian. All students sat the Quick Placement Test designed by Oxford University Press and University of Cambridge Local Examinations Syndicate and their proficiency level in English was determined in compliance with the Common European Framework of Reference for Languages (Council of Europe, 2001). Most of the participants were B2, C1, or C2 level. There were only five students at the B1 level. I decided to mention them in the analysis because of the language performance of one of the participants. All students signed a consent form before the beginning of the investigation.

3.2. Data collection
The research in this study was carried out on a small corpus of conversations produced by the participants, henceforth referred to as Macedonian Learner Corpus (MLC). The conversations were organized in dyads or triads. The conversations are semi-spontaneous because they were collected in an experimental environment, but the respondents were not aware of what was being observed in their speech. The participants were asked to choose three of the following five topics and discuss them: problems with stray dogs, living and working abroad, body piercing and tattoos, the healthy amount of time to spend with the person you’re dating, and talking on the phone while sharing time with friends. The conversations were then transcribed and analysed. The time of the conversations mounted to 591.35 minutes (9.9 hours), with the total number of 66,696 words. In addition to this, attitudinal data were collected from 40 of the participants about their perception of *like* with respect to the age and gender of the speakers, formality of the situation, grammaticality, acceptability, distractibility, and politeness of the utterances. They also rated users for fluency and their level of English. The participants were asked to explain if they use *like* in a similar way and for what purposes as well as how they learned it. The attitudinal survey was adapted from Dailey-O’Cain (2000). The examples, however, were from my data. The results of the attitudinal survey give clues about the usage or avoidance of *like* by the learners.

4. Results
As Müller (2005, p. 204) notes, *like* appears in three functional groups: as a pragmatic marker; as a quotative; and as a verb, noun, adjective, adverb, preposition, conjunction, and in fixed expressions. The total number of tokens of *like* in the MLC was 957: 637 of these were PM or quotative *like* and 320 were the other usages. This paper is devoted to the first two functions.

4.1. The functions of the PM *like* in the English interlanguage of Macedonian learners
The theoretical framework of the analysis was based on the functions of *like* described by Andersen (2001), Müller (2005) and Beeching (2016). The taxonomies in all three studies include approximation, exemplification, and
quotative *like*. Andersen’s and Müller’s taxonomies add the function of *like* as a hesitational/ linking device while Müller’s and Beeching’s taxonomies include the function of *like* to mark lexical focus. Additionally, Andersen includes the function of metalinguistic focus, Müller includes the function of introducing an explanation and Beeching includes the function of hedging. The following functions of *like* were found in the MLC:

- Approximation of a measurable unit (*It’s like five hundred euros I believe*)
- Loose use of lexical expression (*They... were like speaking English I’m... I really... I was confused. That was the English department, but they didn’t speak good English*)
- Exemplification (*When you’re out of the comfort zone like you go running, you see a movie*)
- Introducing an explanation (*No, I... I’m not judgmental about that, like it’s your own body, you do...*)
- Marking lexical focus (*I don’t agree with like poison, but at the end of the day ...; I have a cousin and he like moved abroad; But, that’s like crazy like all the time*)
- *Like* as a hesitational/ linking device (*I thin... Yes, I think it’s different like ... and so many of hmm... our students and the youth, in general, go because of the... To have a bet... Hmm... Like... To live hmm... To earn money because the standards here are low and...*)
- Quotative *like* (*You know I’m like, ‘First I’ll finish college and settle hmm...’*)

4.2. Distribution of *like* across MLC data

The quantitative results of the distribution of *like* as a pragmatic marker across the MLC are shown in Figure 1. A certain number of pragmatic functions of *like* were difficult to classify due to its multifunctionality. In those cases, I opted for the function that seemed the most salient in the given context. The three top functions among MLE were introducing an explanation, using *like* as a hesitational/ linking device and marking lexical focus, while the least used as an approximation of a measurable unit.

![Figure 1 Distribution of all functions of *like* across the MLC](image-url)
4.3. Distribution of like across proficiency levels

Most of the participants of this study were learners at B2, C1, and C2 level according to the Common European Framework of Reference for Languages (CEFR). Table 1 presents the descriptive statistics for the proficiency groups, along with the mean values and standard deviations (SD) of the number of words and uses of like, and the mean rate of like for each level. The frequency rate of the usage of like was calculated by dividing the total number of likes produced by the speakers of each level by the total number of words produced by them, times 100. Although the B1 level was represented by only 5 participants it was also included in the table. However, their mean and frequency rates were not calculated because all 34 tokens of like were produced by a single speaker.

<table>
<thead>
<tr>
<th>Learners per level</th>
<th>Number of words per level</th>
<th>Mean number of words (SD)</th>
<th>Number of PM per level</th>
<th>Mean No of PM (SD)</th>
<th>Frequency rate (per 100 words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>12</td>
<td>9979</td>
<td>832 (562)</td>
<td>77</td>
<td>6.4 (9)</td>
</tr>
<tr>
<td>C1</td>
<td>26</td>
<td>30394</td>
<td>1169 (711)</td>
<td>377</td>
<td>14.5 (24)</td>
</tr>
<tr>
<td>B2</td>
<td>33</td>
<td>23832</td>
<td>722 (453)</td>
<td>137</td>
<td>4.2 (8)</td>
</tr>
<tr>
<td>B1</td>
<td>5</td>
<td>2491</td>
<td>34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*34 tokens of like were used by a single learner

The SD for all groups is large which means that the data is spread out in the distribution. The data are widely scattered from the central tendency and the learners of each level showed very different results with respect to using the PM like. C1 learners performed best: they produced the largest number of like tokens and the number of participants who didn’t use like at all was the lowest (23% in comparison with the 50% C2 learners and 42% B2 learners). The chi-square test with two degrees of freedom was run to calculate the dependence between the production of like and the proficiency levels. The chi-square value equals 255.838 and the two-tailed p-value is less than 0.0001. By conventional criteria, this difference is considered to be extremely statistically significant. This allows for the rejection of the null hypothesis that the levels do not influence the production of the PM like. The p-value indicates a strong relationship between proficiency levels and the use of like. The results of the analysis of the usage rate of like also positively support the correlation between the use of like and the proficiency levels. It again indicates that C1 learners used like at a much higher rate than B2 learners. It is surprising, however, that the C1 learners performed better than the C2 learners. This suggests that the relationship between the use of like and the proficiency levels is not straightforward. I will refer to this in the discussion section.

5. Discussion

5.1. Use of like by native speakers and MLE

The results of this investigation show that the PM like is salient for MLE and that they use it similarly as native speakers (see 3.1). One function that MLE do not use like for is hedging quoted by Beeching (2016) who thought that it was quite prominent in British speakers. Like as a PM is not taught in EFL classes and it is
not illustrated in EFL listening and reading materials. In the attitudinal survey, students reported that they had learned it through media exposure especially movies, reality shows, interviews, YouTube videos, etc.

5.2. Frequency of like across proficiency levels
The second question that this study aimed to answer was if there is a correlation between the use of like and language proficiency levels. I assumed that learners of higher proficiency levels are more likely to use the PM like than learners of lower proficiency levels. Most likes were produced by the students at the C1 level. The big difference between C1 and B2 level is not surprising, 14.5 and 4.2 per person respectively. The C1 learners talked more, produced longer turns, gave more arguments and supported their arguments more efficiently. The use of like enabled them to talk more fluently and sound more native-like. What is surprising, however, is the much lower production of like by the C2 learners in comparison with C1 learners, 6.4 and 14.5, respectively. Generally, the gap between C2 and C1 level is not as drastic as between C1 and B2 and I would have expected the C2 participants to have produced at least closely as many likes as C1 participants. According to previous research, it is expected that PM use would rise with proficiency level (Hasselgreen, 2004; Hellerman and Vergun, 2007; Neary-Sundquist, 2014). The B1 level student who lavishly used like is an exception. The other students at this level did not use it at all.

5.3. Other factors that influence the use of like
Another phenomenon that is easily observed is that the frequency of like across individual speakers varies tremendously. Some learners have produced a high number of tokens of like in their speech, others have produced none. The varied use of like by foreign language learners has been noticed in previous research as well (Müller, 2005; Dailey-O'Cain, 2000; Hellermann and Vergun, 2007). Our lead user produced 101 tokens of like in his speech and the next one produced 55. The production of 16 of the learners ranged between 21 and 54, 32 produced from 1 to 10 tokens while 26 of the learners completely avoided using the PM like. Such discrepancy indicates that other factors also influence the use of like. One of the possible factors mentioned in other studies is a longer stay in an English speaking country. The few examples that I had seem to suggest that this does not have a crucial influence on the usage of like. Besides, the participant who had spent ten months in the USA on an exchange programme did not produce more likes than his interlocutor who had never been to an English speaking country, 12 and 16 respectively. In fact, his production of like was lower than of some other participants who produced between 21 and 54 tokens of like.

What seems to influence the use of like, however, is learners’ perception. The findings from students’ reports in the attitudinal survey of their view of like suggest that there is a connection between the frequency of use of like in learners’ conversations and their attitudes towards this marker. The more positive they are, the more likely they are to use it more frequently. In spite of the high level of proficiency in English (C1 or C2), learners who took a very negative attitude to like and reported to find it annoying avoided using it in their conversations. Interestingly, a lot of the learners have a negative perception of the usage of like. The 40 participants who did the survey see native speakers who use it abundantly as unfriendly (25), not cheerful (24), uninteresting (24), impolite (31). As for
education, 12 learners see the speakers using like uneducated while 19 said that they wouldn’t know. They find such conversations distracting (26) and difficult to follow (29). They see like as a marker of disfluency (24) and a low level of proficiency in English (30). It is interesting that the respondents in Dailey-O’Cain’s (Dailey-O’Cain, 2000) study also indicated an abundance of very strong negative opinions. Only 6 of the students were positive about the usage of the PM like in conversation and think that it helps them organize their speech, 19 said that they don’t use it at all and 15 said that they sometimes or rarely use it. Previous studies of discourse markers showed that students at higher proficiency levels in the learned language who are more acculturated to the L2 environment are more likely to use more discourse markers (Hellermann and Vergun, 2007; Polat, 2011). The usage of like in our data supports this finding.

Personal features of learners also play a significant role. More talkative learners tend to produce longer turns. Longer turns are more demanding and require more time and effort to be organized. There is pressure for selecting the exact vocabulary and structure. Other authors have also noted that there is a correlation between frequency of disfluencies (fluencemes) and length of utterances (Shriberg, 1994; Eklund, 2004; Crible 2018). The group of disfluencies (fluencemes) includes repetitions, substitutions, insertions, deletions, filled pauses, explicit editing terms, uses of discourse markers, coordinating conjunctions, word fragments, misarticulations, contractions, and syntactic incompletions. Learner data at higher proficiency levels do appear to display better command of the items that contribute to fluency (McCarthy, 2010).

6. Conclusion

The results of the present study provide new data on the use of the PM like in the interlanguage of MLE. Our research confirms that the functions that they identify with like are similar to those used by native speakers. This is remarkable when we consider that like is not taught in the classroom. The research also implies that proficiency levels play a significant role in the use of like. However, they are not the only factor. There is a strong indication that learners’ perception of the marker, length of turns and personal features are also very influential. Gender and stay in an English speaking country, on the other hand, do not seem to be of significant importance.

The present study aims to make an important contribution to interlanguage pragmatics. First, it investigates spoken language and reveals some aspects of learner communication that cannot be observed in class. Second, it shows that learning a foreign language is a complex process that involves not only instruction but all other resources that learners have access to through the media. The study presents a strong case for encouraging students’ initiative and autonomy in language learning in order to increase learners’ awareness of many aspects of the target language.

The present study is by no means without its limitations, some of which may be addressed by future research. First, it investigated the use of like in mostly argumentative discourse in which learners expressed their opinions on different topics. Other types of discourse (narrative, expositional, or descriptive) may provide insight into different patterns. Second, this study focused on B2+ levels. Following the example of the B1 learner in this study, it would be useful to
investigate the use of *like* at lower proficiency levels. Other similar examples may also give valuable information about the developmental patterns of *like* across proficiency levels. It may also be useful to conduct further research on C2 learners. A larger population of C2 learners would certainly give more reliable results about their use of *like*. Finally, there are certain methodological issues to be considered. As more people were involved in the first stage of this project, it was difficult to ensure consistency in collecting the data. For example, the learners were more spontaneous with some of the interviewers but less with others. Also, caution should be raised about the difficulty of collecting spoken language. Namely, some of the transcribers were more thorough and noted all pauses, longer syllables, laughter, repetitions, overlappings, etc. Others were less so and missed some of these. Their transcriptions were improved by two reviewers who checked them and inserted the omissions.

References
Jucker, A. H. and Smith, S. W. (1998) "And people just you know like 'wow':


