



## Acquisition of English Discourse Markers by Chinese L1 Speakers Learning English in the US: Frequency and Social impact

Delin Deng

University of Florida, The United States

*Received* 02 April 2021

*Accepted* 19 June 2021

### Abstract

Sankoff et al. (1997) indicated in their research on discourse markers (DMs) used by anglophones in Montreal that the mastery of DMs is a good indicator of the non-native speakers' integration into the linguistic community. As DMs, especially the informal ones, are not taught explicitly in language classes, their acquisition could only be fulfilled by extracurricular contact with native speakers. Despite fruitful works done on DMs in both native and non-native speech, to our knowledge, most of them deal with only one or several DMs at a time without providing a comparable complete list of frequency of DMs in non-native speech. In this article, by exploring the data from 29 semi-guided sociolinguistic interviews conducted in English with non-native speakers in the US, we established a list of frequency of 72 DMs employed by Chinese L1 speakers learning English. By conducting statistical tests, we examined the impact of some extralinguistic factors relevant to non-native speakers in their use of DMs. Our results showed that gender and social network are the two most influential factors for informal DMs, while the age factor seems to be the weakest for all DMs.

**Keywords:** *Discourse markers, Non-native speakers, English, Social impact, Frequency*

### Introduction

Discourse markers (DMs) are the small unity of words that help the speakers in the organisation of their speech. Over the past decades, numerous researchers have dedicated their work to the study of DMs in both native and non-native speech. DMs first caught the attention of linguists in the late 70s when Labov and Fanshel (1977) discussed the use of the DM *well* in the work of Rhoda. Following that, other researchers tried to define DMs and contributed significantly to the literature. Until recent years, SLA researchers started to look at the acquisition of DMs by non-

native speakers. Nevertheless, despite the fruitful work done so far, none of the work has offered a complete list of the frequency of DMs used by native or non-native speakers of English to the best of our knowledge. Therefore, by adopting a corpus-driven methodology, we investigated, on the one hand, the frequency of DMs used by Chinese speakers of English in the United States, on the other hand, the impact of some social factors that are relevant to non-native speakers.

Hence, the article will be laid out as follows: First, the literature of discourse markers and previous work done on English discourse markers in both native and non-native speech will be reviewed. Second, the methodology of the current work, including the corpus and dataset used in the current study, the personal information of participants, the social factors examined, and the statistical analysis, will be presented. Third, the results will be discussed. Lastly, the conclusion, as well as the implication of the current work, will be drawn.

### **Literature Review**

Hansen (1998) defined DMs, which she referred to as discourse particles, “as linguistic items which fulfill a non-propositional, meta discursive (primarily connective) function, and whose scope is inherently variable, such that it may comprise both sub-sentential and supra-sentential units”. She argued that “semantically, markers are best seen as processing instructions intended to aid the hearer in integrating the unit hosting the marker into a coherent mental representation of the unfolding discourse”. Schiffrin (1987) initially defined DMs as “sequentially dependent elements that bracket units of talk”. Then in her later work (Maschler & Schiffrin, 2015), she proposed that DMs “could be considered as a set of linguistic expressions comprised of members of word classes as varied as conjunctions, interjections, adverbs, and lexicalised phrases”. According to Blakemore (2001), “within coherence-based approaches to discourse, expressions like utterance-initial *so*, *well*, *still*, *after all* are classified as discourse markers, a term which is intended to reflect the role that these expressions play in marking, signaling, or indicating how one unit of discourse is connected to another”.

Over the past decades, numerous works have been dedicated to DMs. Nevertheless, no unanimity has been reached regarding its definition, discursive functions, or name. Researchers refer to DMs using different terminologies. This diversity in terminology is mainly due to the divergent points of view taken by the researchers. A few terminologies far from being exhaustive to be cited here as examples: discursive markers (see, for example, Schiffrin, 1987; Jucker & Ziv, 1998; Fraser, 1999; Dostie, 2004 and Andersen, 2007), pragmatic markers (Erman, 2001), enunciative particles (Fernandez, 1994), discourse particles (Schourup, 1985 and Hansen, 1998), pragmatic particles (Beeching, 2002), discursive operators (Redeker, 1990, 1991) or discourse connectives (Blakemore, 1987). Fraser (1998) pointed out that the research on discourse markers is indeed “a growth market in linguistics”.

The research on DMs initiated in the study of the English language. Following Labov and Fanshel (1977) pioneering work, Levinson (1983), in his book entitled *Pragmatics*, suggests that DMs deserve to be studied as one independent category. Nevertheless, he does not give any specific name to DMs, and he did not continue the studies on DMs. Two years later, Zwicky

(1985) expresses his interest in DMs. He calls them “independent words”. In the following decade, a growing interest could be seen in the theoretical status of DMs. These researches (see, for example, Schiffrin, 1987; Fraser, 1988, 1996; Blakemore, 1996; Mann & Thompson, 1988; Hobbs, 1985; Knott & Dale, 1994) mainly focus on the definition, the signification, as well as the discursive functions of DMs.

In second language acquisition, the study of DMs has also attracted much attention in recent decades. For example, Trillo (2002) investigated the role of pragmatic fossilisation of *look, listen, you know, you see, I mean* and *well* by comparing both native and non-native data and adult and child data. Müller (2005) compared the use of DMs in native and non-native speech. Nevertheless, her study mainly focused on four DMs, *so, well, you know* and *like* used by German English speakers at the college level. Huang (2011) explored the use of *like, oh, well, you know, I mean, you see, I think*, and *now* by Chinese speakers of English and found that the use of these discourse markers correlates with the genre, context, type of activity and identity of the speakers. However, as indicated earlier, these works only studied a small subset of DMs at a time without being able to provide a complete list of the frequency of DMs used by non-native speakers. Therefore, in this article, we will provide a complete list of the frequency of DMs and explore the impact of social factors, such as age, gender, and social status of the speakers, on their use of DMs.

## **Methodology**

### **Corpus and Data**

The dataset used for the current study was collected between 2018 and 2019 in New York. The corpus consists of 16 hours’ recording of interviews conducted in English with 29 Chinese speakers of English. The interviews are all semi-directed conversations between the interviewer and the interviewee while following the protocols of traditional Labovian sociolinguistic interviews. Each conversation varies from 30 minutes to 1 hour 30 minutes. For this task, a list of questions was prepared before the interview. However, these interviews were not a simple question-and-answer exchange between the interviewer and the interviewee. These questions were merely a medium for the interviewee to tell his life experiences freely. All interviews were recorded and transcribed entirely on Praat (Boersma and Weenink, 2018). Since the speakers in our corpus are all non-native speakers, there might be grammatical errors in their oral production. For the transcript, we did not make any grammatical corrections. All interviews were transcribed as they were produced. All speakers in our study were anonymised and were referred only by their initials.

### **Participants**

All 29 speakers were born and raised in mainland China. Therefore, they all speak Chinese as L1 and English L2. These speakers were recruited via the snowball technique. Their participation in the interview is entirely voluntary. In terms of their profile, first, these speakers are all between 22 and 36 years old. Instead, they represent a young generation. We could not find older

speakers as expected during our fieldwork who wanted to interview with us. This difficulty is often due to two main reasons: the language barrier and the time cost. It is mainly speakers between the age of 20 and 35 who are more interested in our study and have time to do these interviews. As a result, we will not divide our speakers into conventional age groups, but rather by decade according to their date of birth, namely: post-80s and post-90s.

Secondly, speakers also differ according to their length of stay in the host country. Some of them just arrived in the States for less than a year, while some have already spent more than ten years in the target country. This might impact their use of DMs in that it decides to some degree how much input of the native speech they can get. The longer they are exposed to the native environment, the more input they can get, hence, a better chance of acquiring the informal DMs, which are never taught explicitly in a classroom setting. At the same time, those who lived in New York for more than a decade are better integrated into the local community and the social environment in general. This would explain the difference in the length of the interviews. There is often a big difference in the length of their answer for the same set of questions asked. This could be indicated by the total time of the interview. Often, more advanced speakers prefer to provide longer answers to justify their history. In contrast, beginner speakers always try to answer our questions with a simple yes or no with very little development or justification.

However, it must be admitted that this is not necessarily due to the difference in the language competence of the speakers. Many other factors, such as personality, could also intervene. Also, this difference could be caused by the nervousness of the speakers during the interview. Especially for newcomers, it would be less comfortable for them to maintain the conversation in English for an hour than those who have been in the host country for more than ten years. As for their professional status, on the one hand, some of them have already started to work either in schools or in companies. In this case, they necessarily have more contact with native speakers from a more diverse background. They may have more chances to have access to a more informal register. On the other hand, we also have university students. For this type of speaker, since they mainly stay in the academic environment, they are more exposed to formal register. As we pointed out earlier, DMs are best found, in less careful speech, hence the informal register. We might expect to see the professionals use more DMs than the student group. The following table presents the detailed information of all 29 speakers in our corpus, including the gender, the length of the interview, total words produced by each speaker, their duration of stay in the host country, age group, and professions.

Table 1.

*Detailed information of 29 speakers in our corpus*

| Participant (Initials) | Gender | Interview Length | Total words produced | Duration of stay | Age Group | Profession  |
|------------------------|--------|------------------|----------------------|------------------|-----------|-------------|
| HWL                    | M      | 47mn49s          | 4411                 | 7-9              | post-80   | researcher  |
| LS                     | M      | 34mn45s          | 3085                 | 4-6              | post-80   | student     |
| ZG                     | M      | 43mn11s          | 3940                 | 7-9              | post-80   | employee    |
| MF                     | M      | 26mn26s          | 1335                 | 7-9              | post-80   | employee    |
| SL                     | M      | 26mn19s          | 1477                 | 7-9              | post-80   | researcher  |
| WD                     | M      | 31mn55s          | 2940                 | >10              | post-80   | employee    |
| FYL                    | F      | 35mn35s          | 3091                 | 4-6              | post-80   | teacher     |
| ZJS                    | F      | 39mn53s          | 3685                 | >10              | post-80   | employee    |
| ZX                     | F      | 26mn35s          | 1788                 | 7-9              | post-80   | student     |
| YJ                     | F      | 31mn12s          | 2245                 | 7-9              | post-80   | student     |
| ZL                     | F      | 33mn58s          | 3030                 | >10              | post-80   | free lancer |
| ZLL                    | F      | 31mn22s          | 2735                 | 4-6              | post-80   | teacher     |
| YB                     | F      | 36mn43s          | 3076                 | 7-9              | post-80   | student     |
| QY                     | F      | 35mn44s          | 2983                 | 7-9              | post-80   | housewife   |
| S                      | F      | 24mn13s          | 1497                 | 7-9              | post-80   | employee    |
| WM                     | F      | 36mn33s          | 3407                 | 7-9              | post-80   | intern      |
| XJC                    | F      | 25mn30s          | 1897                 | 4-6              | post-80   | student     |
| DYJ                    | M      | 35mn11s          | 3282                 | 7-9              | post-90   | student     |
| MCL                    | M      | 33mn10s          | 2274                 | 4-6              | post-90   | intern      |
| WDF                    | M      | 30mn56s          | 1917                 | 1-3              | post-90   | student     |
| WHQ                    | M      | 25mn27s          | 2293                 | 7-9              | post-90   | student     |
| HQ                     | F      | 28mn12s          | 2503                 | 1-3              | post-90   | student     |
| JP                     | F      | 27mn22s          | 1776                 | 1-3              | post-90   | employee    |
| LLY                    | F      | 45mn49s          | 5209                 | 7-9              | post-90   | student     |
| LJX                    | F      | 28mn08s          | 1874                 | 4-6              | post-90   | student     |
| ZZ                     | F      | 34mn45s          | 3753                 | 4-6              | post-90   | student     |
| SS                     | F      | 34mn45s          | 3525                 | 1-3              | post-90   | student     |
| WSY                    | F      | 31mn38s          | 2255                 | 4-6              | post-90   | employee    |
| ZJ                     | M      | 47mn49s          | 3383                 | 7-9              | post-80   | employee    |

### Extralinguistic Factors

We will first examine some traditional extralinguistic factors for extralinguistic factors, such as gender and the speaker's age. Since the speakers in our corpus do not differ much according to their social class, we do not look at the effect of social class on their DM employment. Next, we will also look at extralinguistic factors that are more relevant to non-native speakers: their social status, the length of stay in the States, the social network of speakers, and extracurricular contact with native speakers.

For the gender effect, we will examine the difference between women and men in their choice of DMs. We try to see if some DMs are associated with female speakers while others are associated with male speakers. For example, for native speakers, it is found that the French DM

*quoi* is associated with young male speakers. (Beeching, 2007) For our Chinese speakers, we also want to test whether the same remains true. As for the age of speakers, we do not classify them according to the traditional age groups in other sociolinguistic studies, which generally distinguish three age groups: young speakers, middle-aged speakers, and older speakers. Since our speakers are all between 22 and 36, all of whom will be grouped as young speakers traditionally, we have divided them into two main groups: post-80 and post-90, according to the decade they were born. Then, as for the social network, we divided them into two groups: those with an open social network and those with a restricted social network. An open social network indicates that the speaker has regular contact with native speakers out of class, while a restricted social network indicates that speakers prefer to stay with their Chinese friends and do not have much/at all personal contact with native speakers after class. For the open social network, we have two cases: they have many native friends, go out a lot with them, or are married to a native speaker. These two types of social networks decide access to the informal registry used more between friends and outside the classroom/after work. Our hypothesis on this factor is that those with an open network use DMs more than those with a restricted social network.

Regarding the professional status, we mainly distinguish between two statuses: students and professionals. Because different occupational status provides them different access to the informal register of spoken English and the informal register favours the use of DMs, this could also influence their use of DMs. As for the length of stay: 1-3 years, 4-6 years, 7-9 years, and more than ten years. As has been seen in many other studies, the length of stay abroad would positively affect the acquisition of a second language since they have more exposure to the authentic use of that language in an authentic environment. Here, we are trying to see if this will be a significant factor in acquiring DMs. Regarding the extracurricular contact with native speakers, based on their self-reported responses, we can establish three main groups on the frequency of contact: rare, occasional, and frequent. Since DMs are not explicitly taught inside the classroom, extracurricular contact with native speakers would, according to our hypothesis, be an essential and primary means for the acquisition of DMs. Therefore, we could expect a difference between these three groups.

### **Statistical Analysis**

To examine the statistical significance of each extralinguistic factor proposed above, we mainly employ two methods: unpaired t-test (Snedecor & Cochran, 1989) and one-way ANOVA (Field, 2007). In both methods, we look for the p-value, which indicates whether the tested factor is statistically significant between groups. The p-value is usually between 0 and 1. The smaller the p-value, the more substantial the evidence that one should reject the null hypothesis (denoted as  $H_0$ ) is, noting that a null hypothesis proposes no association among groups. For the tested factor to be statistically significant, the p-value should be less than 0.05 (typically  $\leq 0.05$ ).

On the one hand, the unpaired t-test mainly compares the averages/means of two independent or unrelated groups. There are two possible hypotheses in the unpaired t-test: the null hypothesis ( $H_0$ ) and the alternative hypothesis (denoted as  $H_1$ ).  $H_1$  typically states a significant difference

between the two population means and that this difference is unlikely to be caused by sampling error or chance.

On the other hand, the one-way ANOVA is used to determine whether there are any statistically significant differences between the means of three or more independent or unrelated groups. Like the t-test, one-way ANOVA also tests the null hypothesis. However, the one-way ANOVA is an omnibus test statistic and cannot tell us which specific groups are statistically significantly different from each other, only that at least two groups are. Therefore, to determine which specific groups differ, we need to run a post hoc test when the ANOVA test indicates significant results.

As for the statistical report format, the result of the t-test is reported as:

$t_{\text{statistic}} (df) = t\text{-value}, p=p\text{-value}$

The result of one-way ANOVA is reported as:

$f_{\text{statistic}} (df_{\text{between}}, df_{\text{within}}) = f_{\text{ratio}}, p=p\text{-value}$

## Results

### General Tendency

In general, we can distinguish between two types of DMs: the formal DMs and the informal DMs. The formal DMs are the ones that could be used in both spoken and written contexts, such as *for example, especially, fortunately*, etc., while informal DMs are those that are only used in spoken context when used as DMs, such as *like, you know, just*, etc. After a careful data cleaning of our corpus, we identified 72 English DMs most used by our non-native speakers. The following table presents the forms, total occurrences in our corpus, and occurrences per 1000 words.

Table 2.

*Frequency of 72 English DMs used by non-native speakers in our corpus*

| No. | Forms    | Total Occurrences | Occ./1000 words | No. | Forms            | Total Occurrences | Occ./1000 words |
|-----|----------|-------------------|-----------------|-----|------------------|-------------------|-----------------|
| 1   | and      | 1600              | 19.8349         | 37  | all the time     | 13                | 0.1612          |
| 2   | like     | 1578              | 19.5621         | 38  | currently        | 12                | 0.1488          |
| 3   | so       | 1477              | 18.3101         | 39  | generally        | 9                 | 0.1116          |
| 4   | yeah     | 1035              | 12.8307         | 40  | directly         | 8                 | 0.0992          |
| 5   | but      | 882               | 10.9340         | 41  | eventually       | 8                 | 0.0992          |
| 6   | just     | 696               | 8.6282          | 42  | in general       | 8                 | 0.0992          |
| 7   | I think  | 571               | 7.0786          | 43  | seldom           | 7                 | 0.0868          |
| 8   | because  | 405               | 5.0207          | 44  | similar to       | 6                 | 0.0744          |
| 9   | you know | 399               | 4.9463          | 45  | in the future    | 6                 | 0.0744          |
| 10  | kind of  | 299               | 3.7066          | 46  | at the same time | 6                 | 0.0744          |
| 11  | If       | 291               | 3.6075          | 47  | either...or      | 6                 | 0.0744          |
| 12  | Yes      | 277               | 3.4339          | 48  | finally          | 6                 | 0.0744          |
| 13  | then     | 217               | 2.6901          | 49  | sort of          | 5                 | 0.0620          |
| 14  | also     | 204               | 2.5289          | 50  | you see          | 5                 | 0.0620          |
| 15  | I mean   | 196               | 2.4298          | 51  | otherwise        | 5                 | 0.0620          |
| 16  | as       | 195               | 2.4174          | 52  | at the beginning | 5                 | 0.0620          |
| 17  | oh       | 190               | 2.3554          | 53  | despite          | 5                 | 0.0620          |
| 18  | maybe    | 185               | 2.2934          | 54  | as well as       | 4                 | 0.0496          |

|    |              |     |        |    |               |   |        |
|----|--------------|-----|--------|----|---------------|---|--------|
| 19 | right        | 181 | 2.2438 | 55 | fortunately   | 4 | 0.0496 |
| 20 | actually     | 165 | 2.0455 | 56 | however       | 4 | 0.0496 |
| 21 | I don't know | 143 | 1.7727 | 57 | initially     | 4 | 0.0496 |
| 22 | still        | 143 | 1.7727 | 58 | such as       | 3 | 0.0372 |
| 23 | ok           | 139 | 1.7232 | 59 | similarly     | 3 | 0.0372 |
| 24 | after        | 112 | 1.3884 | 60 | at the end    | 3 | 0.0372 |
| 25 | I guess      | 110 | 1.3636 | 61 | instead       | 3 | 0.0372 |
| 26 | well         | 104 | 1.2893 | 62 | yet           | 2 | 0.0248 |
| 27 | cuz          | 98  | 1.2149 | 63 | same as       | 2 | 0.0248 |
| 28 | before       | 78  | 0.9670 | 64 | naturally     | 2 | 0.0248 |
| 29 | never        | 77  | 0.9546 | 65 | more and more | 2 | 0.0248 |
| 30 | especially   | 44  | 0.5455 | 66 | due to        | 2 | 0.0248 |
| 31 | again        | 39  | 0.4835 | 67 | in addition   | 2 | 0.0248 |
| 32 | for example  | 38  | 0.4711 | 68 | after all     | 1 | 0.0124 |
| 33 | since        | 33  | 0.4091 | 69 | anyway        | 1 | 0.0124 |
| 34 | besides      | 28  | 0.3471 | 70 | indeed        | 1 | 0.0124 |
| 35 | during       | 28  | 0.3471 | 71 | in fact       | 1 | 0.0124 |
| 36 | of course    | 23  | 0.2851 | 72 | in the end    | 1 | 0.0124 |

As shown in the table above, the speakers in our corpus use some informal markers whose use as DMs are relatively recent among native speakers, such as *like, just, you know, kind of*, with a high frequency in their oral production. If the use of formal DMs can be learned in the language class, the mastery of informal DMs does require some extracurricular contact with native speakers outside the class. Such a high frequency of informal DMs in their speech suggests that these speakers might be well integrated linguistically. However, it is worth pointing out that among these 72 DMs, only less than a third of them are relatively frequent compared to the other two-thirds if we look at the total occurrences of each form. This suggests that the variety of DMs mastered by non-native speakers is still restrained. Very often, they would rely on several available forms and might overgeneralise the use of those forms. However, till this point, we cannot provide any conclusive evidence for overgeneralisation. It is imperative to collect further data from native speakers to make the comparison. For the current work, we will only focus on the social impact of DMs' use by our non-native speakers.

### Gender

Gender is one of the traditional extralinguistic factors in sociolinguistic studies. As proposed by Labov (2001), women differ significantly from men in their language use. He indicates that “women conform more closely than men to sociolinguistic norms that are overtly prescribed but conform less than men when they are not” (Labov, 2001: 293). This is referred to as the gender paradox. At the same time, the gender factor is proven to be statistically significant for the use of many DMs in native speech. This means that some DMs are used more by male speakers than by female speakers and vice versa. The following table presents five DMs among the total 72 DMs with a P value of less than 0.05: *like, then, since, generally* and *initially*, meaning for which the gender factor is statistically significant.

Table 3.

*Gender effect*

|       | No. | <i>like</i>     | <i>then</i>     | <i>since</i>    | <i>generally</i> | <i>initially</i> |
|-------|-----|-----------------|-----------------|-----------------|------------------|------------------|
| F     | 18  | 1940            | 301             | 53              | 2                | 0                |
| M     | 11  | 773             | 67              | 6               | 13               | 7                |
| Total | 29  | 2713            | 368             | 59              | 15               | 7                |
| P     |     | <b>0.042271</b> | <b>0.033912</b> | <b>0.017192</b> | <b>0.01241</b>   | <b>0.033662</b>  |

As shown in the table above, *like*, *then*, and *since* are more used by female speakers, while *generally* and *initially* are more associated with male speakers. With our earlier classification, it should be pointed out that only the *like* is the informal DM because this marker only appears orally when used as DM.

The use of *like* as a DM is relatively recent in native speech. It is well documented that this employment is spreading rapidly in urban centres throughout the English-speaking world. (See, for example, Dailey-O’Cain, 2000; Tagliamonte, 2005; D’Arcy, 2005; Cheshire et al., 2005). Andersen (2001) showed that the use of this DM is widespread, especially among pre-teens. However, Miller and Weinert (1995) suggested that learning this DM by native speakers is not done until 10. This statement is also corroborated by other researchers such as Levey (2006). For our non-native speakers, according to table 2, *like* is the second most frequently used DM on our list with a frequency of 19.5621 occurrences per 1000 words. Since our speakers are all aged between 22 to 36, representing a young generation, even though they are no longer pre-teens or adolescents, the relatively young age of our speakers still partially explains a large number of occurrences of *like* in our corpus.

In addition, Levey (2006) reported a distributional difference between women and men regarding the use of *like*, which is also confirmed by Croucher (2004): *like* is more associated with female native speakers. According to table 2, the difference between these two gender groups remains valid for our non-native speakers. Non-native female speakers use more *like* as a DM than male speakers. The first question arises here: How could we explain this difference between men and women?

Labov (2001) observed that women often take the initiative to use new language forms compared to men. Since the appearance of *like* used as a DM began in the 1980s (see, for example, Schourup, 1985; Underhill, 1988; Andersen, 1998, 2001; Smith & Jucker, 1998; Dailey- O’Cain, 2000; Fuller, 2003), it is not surprising to see women using more DM *like* than men. As a result, it could be said that non-native speakers behave linguistically similarly to native speakers concerning the use of new linguistic forms even if the forms are newly emerging in the target language.

The second question that one may ask here is: why non-native speakers behave in the same way as native speakers? For this question, we might say that since the gender paradox proposed by Labov is universal, it does not make the distinction between native and non-native speakers. As non-native speakers are also language users, they only differ from native speakers in their linguistic competence. It is natural to see that when it comes to newly emerging forms, they

behave the same as native speakers since the gender paradox is not about proficiency but the behaviour of language users in general.

However, it would be interesting to argue if there is any L1 transfer impact on the DM *like*. In other words, do we have an equivalent form of *like* in Chinese that is frequently used as DM? According to Liu (2009), the equivalent of *like* is not among the top ten most used Chinese DMs. The fact that *like* is the second most used by non-native speakers here is less probable to be the consequence of the L1 transfer.

### Age

As shown in Table 4, the four DMs for which the age factor is statistically significant are: *still*, *never*, *currently*, and *otherwise*.

Table 4.  
Age effect

|       | No. | <i>still</i>    | <i>never</i>    | <i>currently</i> | <i>otherwise</i> |
|-------|-----|-----------------|-----------------|------------------|------------------|
| 80s   | 18  | 192             | 118             | 20               | 2                |
| 90s   | 11  | 60              | 21              | 2                | 7                |
| Total | 29  | 252             | 139             | 22               | 9                |
| P     |     | <b>0.023183</b> | <b>0.005029</b> | <b>0.024437</b>  | <b>0.044425</b>  |

Our result suggests that post-80 and post-90 do not differ significantly in their use of the aforementioned informal DMs but only in the use of certain formal DMs. It is shown that post-80s use more *still*, *never* and *currently*, while post-90s use more *otherwise*. However, we must admit that none of these four markers have abundant occurrences in our corpus.

However, this might partly be because the age difference between the two groups is not significant enough to see the inter-group difference for this factor. Hence, our future data collection must include middle-aged and older speakers to verify the impact of the age factor.

### Social Networks

Concerning the social networks, table 5 indicates that this factor is statistically significant to six DMs: *and*, *so*, *just*, *then*, *in the future*, and *yes*.

Table 5.  
Effect of social network

|            | No. | <i>and</i>      | <i>so</i>      | <i>just</i>     | <i>then</i>     | <i>in the future</i> | <i>yes</i>      |
|------------|-----|-----------------|----------------|-----------------|-----------------|----------------------|-----------------|
| Open       | 17  | 2045            | 1724           | 869             | 300             | 0                    | 206             |
| Restrained | 12  | 752             | 912            | 369             | 68              | 12                   | 289             |
| Total      | 29  | 2797            | 2636           | 1238            | 368             | 12                   | 495             |
| P          |     | <b>0.000236</b> | <b>0.03981</b> | <b>0.020169</b> | <b>0.017405</b> | <b>0.002349</b>      | <b>0.018044</b> |

According to our list, *and* is the most used DM by our non-native speakers, with a frequency of 19.8349 occurrences per 1000 words. What surprises us is that the group with an open network uses much more *and* than the other group. This could mean that those who have an open

social network and interact more with native speakers pay more attention to ensure the fluidity of their speech and consequently add more *and* to connect different sentences.

The DM *so* also deserves a discussion here. It is documented in other research that the DM *so* is considered one of the most popular DM by native English speakers. Bolden (2008) indicated that *so* is often used to initiate a conversation that plays a transient role between conversational topics. Bolden (2009) believes that the DM *so* is a resource to establish discursive coherence and, perhaps more fundamentally, to accomplish understanding. According to Tagliamonte (2005), the DM *so* is a marker associated more with the younger generation. Given the age range of our speakers, this may explain the excessive use of this marker in general. In addition, with an open social network, speakers have more contact with young native speakers of the same age group.

As for the DM *just*, according to Tagliamonte (2005), it is also a DM whose use is more associated with young speakers. It seems that young people are unanimously innovators. (See, for example, (Andersen, 1997, 1998, 2001; Buchstaller, 2001; Erman, 2001; Siegel, 2002) Its widespread presence in our corpus suggests a linguistic behaviour of non-native speakers similar to that of the native speakers.

It is worth noting that the DM *then* is one DM for which both the gender factor and the social network factor all seem to be significant for its use. It seems that social factors play an essential role in the use of this marker. This remains to be confirmed later with other social factors.

As for *yes*, it is the group with a restrained social network that uses it more often than the other group; one may ask if it is due to L1 transfer. Its Chinese equivalent *dui* is considered to be one of the most used DMs by Chinese native speakers. (Liu, 2009) The high frequency of this DM in the speech of the restrained group suggests the possibility of L1 transfer. Since for native speakers of English, it is rather its variant *yeah* that is used as DM.

### Social Status

For social status, we distinguish between two groups: students and professionals. This factor is proven to be statistically significant for seven DMs: *still*, *never*, *for example*, *during*, *of course*, *otherwise*, and *at the beginning*.

Table 6.  
Effect of social status

|               | No. | <i>still</i>    | <i>never</i>    | <i>for example</i> | <i>during</i>   | <i>of course</i> | <i>otherwise</i> | <i>at the beginning</i> |
|---------------|-----|-----------------|-----------------|--------------------|-----------------|------------------|------------------|-------------------------|
| Students      | 12  | 110             | 52              | 21                 | 16              | 13               | 9                | 9                       |
| Professionals | 17  | 142             | 87              | 42                 | 32              | 28               | 0                | 0                       |
| Total         | 29  | 252             | 139             | 63                 | 48              | 41               | 9                | 9                       |
| P             |     | <b>0.018402</b> | <b>0.009872</b> | <b>0.011955</b>    | <b>0.037803</b> | <b>0.017902</b>  | <b>0.04066</b>   | <b>0.04066</b>          |

These seven discursive markers all fall into the category of formal DMs. Only *otherwise* and *at the beginning* markers are more used by the student group, while the other five are more used by professionals. It seems that professional life in general favours the use of more forms of DMs since interpersonal contact with native speakers is inevitable for professionals, whether in an academic environment or other areas of work.

However, the results do not suggest a difference between the two groups regarding the informal DMs. In fact, even if professionals use more forms of DMs to ensure discursive fluidity, the professional situation would require some formality of their speech and, therefore, disadvantages of excessive use of overly informal DMs in their daily conversation with native speakers at work. The same is true for students. The academic environment restricts the use of overly informal DMs, at least in classes. If students only have contact with native speakers in the classroom, learning informal DMs would be challenging to accomplish. From this point of view, other factors would play a more critical role than professional status, at least in the use of informal DMs.

### Duration of Stay

As for the duration of stay in the host country, we divided our speakers into four groups based on the number of years they spent in the States. Table 7 shows that this factor is statistically significant only for four DMs: *then*, *during*, *eventually*, and *in the end*.

Table 7.  
*Effect of duration of stay*

|       | No. | <i>then</i>     | <i>during</i>      | <i>eventually</i> | <i>in the end</i> |
|-------|-----|-----------------|--------------------|-------------------|-------------------|
| 1-3   | 4   | 23              | 0                  | 0                 | 0                 |
| 4-6   | 8   | 84              | 13                 | 4                 | 0                 |
| 7-9   | 14  | 149             | 12                 | 2                 | 0                 |
| 10+   | 3   | 112             | 23                 | 7                 | 2                 |
| Total | 29  | 368             | 48                 | 13                | 2                 |
| P     |     | <b>0.018955</b> | <b>&lt; .00001</b> | <b>0.002094</b>   | <b>0.024024</b>   |

On the one hand, it must be admitted that the factor of the length of stay in the host country is not as significant as expected. Its influence on the DMs most used by our speakers is not statistically significant. The four markers to which this factor is significant do not have many occurrences in our corpus. Therefore, it seems that the length of stays does not so much influence learning DMs in the host country. On the other hand, it should be pointed out that the longer we stay in the host country for all four forms, the more we use these DMs. Especially for the group who stayed in the target country for over 10 years, we found the use of all four forms, while the newcomer group only use *then*. This suggests that a longer duration of stay in the host country facilitates the use of a more extensive range of DMs. In other words, the newcomers often rely on a small subset of forms, while those who spent more extended time in the target community usually have a good command of a much more extensive range of available forms.

### Extracurricular Contact with Native Speakers

Many researchers confirm that extracurricular activity promotes the learning of discursive markers. (Sankoff et al., 1997) Based on the self-reported response of our non-native speakers during the interview, we established three groups of extracurricular contact with native speakers outside the classroom: those who have almost no contact with native speakers after class/out of work, those who have occasional contact, and those who have frequent contact with the native

speakers. The statistical result shows that this factor is not as significant as expected. However, the result is still interesting and therefore worth a discussion.

Table 8.

*Effect of extracurricular contact*

|            | No. | <i>and</i>      | <i>just</i>     | <i>yes</i>      | <i>then</i>     |
|------------|-----|-----------------|-----------------|-----------------|-----------------|
| Rare       | 7   | 489             | 289             | 206             | 33              |
| Occasional | 9   | 688             | 211             | 126             | 61              |
| Frequent   | 13  | 1620            | 738             | 163             | 274             |
| Total      | 29  | 2797            | 1238            | 495             | 368             |
| P          |     | <b>0.010023</b> | <b>0.010117</b> | <b>0.042682</b> | <b>0.021226</b> |

First, Table 8 shows three different patterns of learning concerning these four DMs. For *and*, the more frequent the extracurricular contact with native speakers, the more one uses this marker. This is also consistent with what is found in the social network factor. Second, for *just*, we see a “U-shaped development”. That is, *just* is most used by the group with frequent contact with native speakers, followed by the group who have rare contact with native speakers. The group who have occasional contact with native speakers use this DM the least. This could be interpreted by an L1 transfer at the beginning. Then, with a little more contact, this influence of L1 transfer decreases, shown by a sharp decrease in the use of this DM. Finally, excessive contact with native speakers allows the speakers to approach the native pattern with a new increase of use. As for *yes*, its use also corresponds to what was found earlier in the social network factor and therefore corroborates our hypothesis of a possible L1 transfer. What makes it different from the use of *just* is that *just* is used frequently by native speakers, especially the younger generation, while native speakers do not use *yes* as a DM. This explains why excessive extracurricular contact does not lead to an increase in the frequency of this DM. Finally, regarding the DM *then*, it is similar to the DM *and*: the more frequent the extracurricular contact is, the more it is used. Frequent extracurricular contact facilitates its use. What is interesting about this DM is that it appears that most extralinguistic factors are proven statistically significant for the use of this marker. The influence of social factors seems extremely big on the use of this DM among non-native speakers.

## Conclusion

By exploring the data collected with 29 Chinese speakers of English in the States, the current study, on the one hand, provided a complete list of 72 English discourse markers most used by non-native speakers in our corpus; on the other hand, it investigated the impact of some social factors most relevant to non-native speakers.

The statistical results showed that the social factors do have an impact on the use of DMs by non-native speakers. In general, the non-native speakers demonstrate a similar pattern concerning the newly emerged DMs, such as *like* when it comes to the difference between male speakers and female speakers. The age factor does not seem to be very influential, given the relatively homogeneous age of our speakers. The duration of stay in the target country is not necessarily correlated with more use of informal DMs, but it does allow speakers to master a

wider variety of forms. Lastly, the extracurricular contact with native speakers does facilitate a native-like pattern. The L1 transfer is found the most common in the newcomers with little extracurricular contact with native speakers outside the classroom.

For the future study on the same community, it would be interesting to include more middle-aged and older speakers to observe if there is any age effect on the use of DMs. At the same time, it is necessary to interview speakers from a more diverse social background, such as workers, employees, etc., to see how social-economic status could impact the use of DMs in non-native speech. Lastly, it would also be ideal for conducting a longitudinal study on the same group of speakers to trace the change of the use of DMs over the lifespan.

## References

- Andersen, G. (1997). They like wanna see like how we talk and all that: the use of like as a discourse marker in London teenage speech. Dans M. Ljung, *Corpus-based Studies in English: Papers from the 17th International Conference on English Language Research on Computerized Corpora*, 37-48. Amsterdam: Rodopi.
- Andersen, G. (1998). The pragmatic marker like from a relevance-theoretic perspective. *Pragmatics and beyond new series*, 147-170.
- Andersen, G. (2001). *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents* (Vol. 84). John Benjamins Publishing.
- Andersen, H. (2007). Marqueurs discursifs propositionnels. *Langue française*, 154(2), 13-28.
- Beeching, K. (2002). *Gender, Politeness and Pragmatic Particles in French*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Beeching, K. (2007). La co-variation des marqueurs discursifs bon, c'est-à-dire, enfin, hein, quand même, quoi et si vous voulez: une question d'identité? *Langue Française*, 2, 78-93.
- Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (1996). Are apposition markers discourse markers? *Journal of linguistics*, 32(2), 325-347.
- Blakemore, D. (2001). Discourse and relevance theory. *The handbook of discourse analysis*, 100-118.
- Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Version 6.0. 37. Retrieved February, 3, 2018.
- Bolden, G. B. (2008). "So what's up?": Using the discourse marker so to launch conversational business. *Research on Language and Social Interaction*, 41(3), 302-337.
- Bolden, G. B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation. *Journal of Pragmatics*, 41(5), 974-998.
- Buchstaller, I. (2001). An alternative view of like: its grammaticalisation in conversational American English and beyond. *Edinburgh Working Papers in Applied Linguistics*, 11, 21-41.
- Cheshire, J., Kerswill, P., & Williams, A. (2005). Phonology, grammar, and discourse in dialect convergence. *Dialect Change: Convergence and Divergence in European Languages*, 135-167.
- Croucher, S. M. (2004). Like, you know, what I'm saying: A study of discourse marker frequency in extemporaneous and impromptu speaking. *National Forensic Journal*, 22(2), 38-47.
- Dailey- O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser like and quotative like. *Journal of Sociolinguistics*, 4(1), 60-80.
- D'Arcy, A. (2005). *Like: Syntax and Development* (Doctoral dissertation, University of Toronto).
- Dostie, G. (2004). *Pragmaticalisation et marqueurs discursifs: Analyse sémantique et traitement lexicographique*. Louvain-la-Neuve, Belgique : De Boeck Supérieur.

- Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics*, 33(9), 1337-1359.
- Fernandez, M. J. (1994). *Les particules énonciatives dans la construction du discours*. Presses Universitaires de France-PUF.
- Field, A. (2007). Analysis of variance (anova). In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 1, pp. 33-35). SAGE Publications, Inc.
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1-4), 19-33.
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167-190.
- Fraser, B. (1998). Contrastive discourse markers in English. *Pragmatics and Beyond New Series*, 301-326.
- Fraser, B. (1999). What are discourse markers? *Journal of pragmatics*, 31(7), 931-952.
- Fuller, J. M. (2003). Use of the discourse marker like in interviews. *Journal of sociolinguistics*, 7(3), 365-377.
- Hansen, M. B. (1998). *The function of discourse particles: A study with special reference to spoken standard French*, 53. John Benjamins Publishing.
- Hobbs, J. R. (1985). *On the coherence and structure of discourse*.
- Huang, L. F. (2011). *Discourse markers in spoken English: A corpus study of native speakers and Chinese non-native speakers* (Doctoral dissertation, University of Birmingham).
- Jucker, A. H., & Ziv, Y. (1998). *Discourse markers: Descriptions and theory*, 57. John Benjamins Publishing.
- Knott, A., & Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse processes*, 18(1), 35-62.
- Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. Academic Press.
- Labov, W. (2001). *Principles of linguistic change*. Vol. 2: Social Factors (vol. 29). Language in Society, Oxford & Malden, MA.
- Levinson, S. C. (1983). Pragmatics Text. *Cambridge Textbooks in Linguistics*. Cambridge University Press, 420.
- Levey, S. (2006). The sociolinguistic distribution of discourse marker like in preadolescent speech. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication*, 25(4), 413-441.
- Liu, B. & Xiao, E. (2009). Chinese discourse markers in oral speech of mainland Mandarin speakers. *In Proceedings of the 21st North American conference on Chinese linguistics (NACCL-21)*, 2, 358-374.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organisation. *Text*, 8(3), 243-281.
- Maschler, Y., & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. *The handbook of discourse analysis*, 1, 189-221.
- Miller, J., & Weinert, R. (1995). The function of LIKE in dialogue. *Journal of Pragmatics*, 23(4), 365-393.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. John Benjamins Publishing.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of pragmatics*, 14(3), 367-381.
- Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29(6), 1139-1172.
- Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M. O., & Gagnon, L. (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*, 9(2), 191-217.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schourup, L. C. (1985). *Common discourse particles in English conversation* (1st ed.). Routledge.
- Siegel, M. E. (2002). Like: The discourse particle and semantics. *Journal of Semantics*, 19(1), 35-71.

- Smith, S. W., & Jucker, A. H. (1998). Interactive aspects of reference assignment in conversations. *Pragmatics & Cognition*, 6(1), 153-187.
- Snedecor, W., & Cochran, G. (1989). *Statistical Methods*. Iowa State University Press.
- Tagliamonte, S. (2005). So who? Like how? Just what?: Discourse markers in the conversations of young Canadians. *Journal of Pragmatics*, 37(11), 1896-1915.
- Trillo, J. R. (2002). The pragmatic fossilisation of discourse markers in non-native speakers of English. *Journal of pragmatics*, 34(6), 769-784.
- Underhill, R. (1988). Like is, like, focus. *American Speech*, 63(3), 234-246.
- Zwicky, A. M. (1985). Clitics and particles. *Language*, 283-305.

### **Acknowledgements**

Not applicable.

### **Funding**

Not applicable.

### **Ethics Declarations**

### **Competing Interests**

No, there are no conflicting interests.

### **Rights and Permissions**

### **Open Access**

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. You may view a copy of Creative Commons Attribution 4.0 International License here: <http://creativecommons.org/licenses/by/4.0/>.