Discourse Markers as the Classificatory Factors of Speech Acts

Da Qi[1]**, Chenliang Zhou**[1]**, Haitao Liu**[1,2,🖂]

[1]Department of Linguistics, Zhejiang University, China
[2]Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, China
{da.qi,cl.zhou}@zju.edu.cn, lhtzju@yeah.net

Abstract

Since the debut of the speech act theory, the classification standards of speech acts have been in dispute. Traditional abstract taxonomies seem insufficient to meet the needs of artificial intelligence for identifying and even understanding speech acts. To facilitate the automatic identification of the communicative intentions in human dialogs, scholars have tried some data-driven methods based on speech-act annotated corpora. However, few studies have objectively evaluated those classification schemes. In this regard, the current study applied the frequencies of the eleven discourse markers (*oh, well, and, but, or, so, because, now, then, I mean*, and *you know*) proposed by Schiffrin (1987) to investigate whether they can be effective indicators of speech act variations. The results showed that the five speech acts of *Agreement* can be well classified in terms of their functions by the frequencies of discourse markers. Moreover, it was found that the discourse markers *well* and *oh* are rather efficacious in differentiating distinct speech acts. This paper indicates that quantitative indexes can reflect the characteristics of human speech acts, and more objective and data-based classification schemes might be achieved based on these metrics.

1 Discourse Markers and the (Dis)agreement Continuum

A discourse marker (DM) is a word or phrase that people often use in the process of communication, and its main function is to coordinate and organize discourse to ensure the smooth flow of conversation. In addition, as a carrier of pragmatic information, it usually reflects speakers' mental states and communicative intentions, thus facilitating pragmatic inference (Furkó, 2020). In this regard, Fraser (1996, p.68) defined DMs as "linguistically encoded clues which signal the speaker's potential communicative intentions." Although scholars have never reached a consensus on the definition of DMs, no one would doubt their diverse discursive functions and the capability to transmit communicative intentions.

When analyzing the functions of DMs, scholars also differ considerably in terms of their frameworks and research paradigms. Ariel (1998) distinguished DMs from a semantic perspective: a DM either possesses a semantic meaning, which is interpreted in a particular context with some connection to its form (e.g., *and* and *I mean*); or it does not contain any semantic information (e.g., *well* and *oh*). However, Matei (2010) pointed out that although some DMs contain rich semantic information, there are particular contexts in which the communicative intention it conveys is not related to the semantic information it carries. For example, in some cases, the DM *and* can be used as a discourse continuative, filler word, and buffer term, etc.

Some scholars analyzed the range of functions through the functional-cognitive approach, which shows that DMs have a specific rather than a completely arbitrary range of functions (Redeker, 2006; Fischer, 2006). Ariel (1998, pp.242-243) also expressed support for the non-arbitrary nature of DM functions. She explicated this view in terms of the correspondence between form and function and argued that there are two probabilistically similar possibilities for the form-function correspondence, one in which a form corresponds to multiple functions and the other in which a function corresponds to

🖂 Corresponding Author

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

multiple forms. She further claimed that these two possible relationships do not indicate the syntactic arbitrariness but are characterized by unpredictability since the same form may evolve to express many innovative meanings. In this sense, functionalists argue that the universality of DM forms (as opposed to the uniformity of forms) is functionally driven.

The above investigations of DM functions have helped us to gain a deep and broad understanding of DMs' nature and their functional orientations in various contexts. However, as Matei (2010) mentioned, there is a great deal of uncertainty in DMs' functions, and even those with a relatively fixed semantic meaning may produce new and rare uses in some contexts. In addition, the *one form – many functions* and *one function – many forms* nature of DMs, as well as the innovative nature of their functions, also make their functions perform in a variety of ways. Thus, it is difficult to assess all the functions of DMs through an in-depth analysis of the discourse material one by one (the workload is too large). If we want to characterize all aspects of certain DMs and explore the patterns of these linguistic units full of uncertainties and probabilities, it is better to apply an approach that is suitable for approximating all the features possessed by the DMs.

Another consideration in employing this approach is that human communicative intentions are themselves fraught with probabilities and uncertainties. As pointed out by the Speech Act Theory, there is not always a clear correspondence between the words people express and their functions, and speech acts are also characteristic of *one form – many functions* and *one function – many forms* as mentioned by Ariel (1998) (Holtgraves, 2005). A more extreme example, such as *Kennst du das Land wo die Zitronen blühen?* (Knowest thou the land where the lemon trees bloom?), can even express the communicative intention "I am a German soldier" (Searle, 1969). By the same token, the various DMs proposed by previous authors, such as the eleven DMs by Schiffrin (1987) (*oh, well, and, but, or, so, because, now, then, I mean*, and *you know*), may occur in various speech acts depending on the specific speech context.

To address the function of DMs and the probability and uncertainty of human communicative intentions, the present paper tries to introduce some basic probabilistic and statistical methods, such as the hierarchical cluster analysis (HCA), to quantitatively analyze the DMs contained in specific communicative intentions. Our aim is to examine whether certain indicators of DM (e.g., their percentage of frequency of occurrence in different speech acts) can effectively distinguish the communicative functions embodied in differing speech acts to propose a new research methodology for DM-related studies.

In the current study, the frequency of different DMs in differing speech acts was investigated as a possible defining feature for the distinction of communicative intentions. The reason for doing so is that DMs carry diverse pragmatic and contextual information (Redeker, 2006). In this regard, the frequency of DMs may reflect the pragmatic characteristics of different speech acts, which may help us better explore the patterns of human communicative intentions.

Since we want to examine whether the frequency of DMs can effectively distinguish different speech acts, these DMs should first be able to reflect the differences between speech acts that differ significantly, e.g., agreement and disagreement, thanks and apology, etc. Next, we may examine whether it can reflect the slight differences between similar speech acts. Therefore, in the current paper, we applied the continuum of (dis)agreement (*accept, partially accept, hedge, partially reject,* and *reject*) as the object of study to explore whether the frequency distribution of DMs can accurately capture the nuanced differences in speech acts.

When analyzing the agreement-disagreement continuum, scholars have mostly focused on the perception of agreement- and disagreement-like speech acts by people in specific types of discourses. For example, Mulkay (1985) found that strong disagreement is easier to declare in writing than face to face after examining the written letters by biochemists. When investigating the arguments of mentally disabled people, Hewitt et al.'s (1993) study showed that regarding conflict resolution as the primary goal of arguments detracts from the true nature of verbal conflicts – they reflect a social continuum of agreement and disagreement (Jacobs and Jackson, 1981). Trimboli and Walker (1984) compared dyadic discussions following initial agreement and disagreement and found that disagreement was more competitive, characterized by high rates of verbalization, increased numbers of turns, more frequent interruptions, and reduced back channels.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.

From the studies above, it can be seen that agreement and disagreement are complex and influenced by various socio-cultural factors, but the specific mechanisms of their intricacies have been seldom studied, and a more systematic and comprehensive understanding has yet to be developed. In the current study, we attempted to employ the frequency of DMs as well as probabilistic and statistical methods to examine the speech acts of agreement and disagreement, complementing the existent findings in discourse analysis.

The research questions of this paper are as follows.

1. How is the frequency distribution of different DMs under the differing speech acts in the agreement-disagreement continuum?

2. Can the frequency of DMs effectively reflect the similarity and peculiarity of the different speech acts?

2 Methods and Materials

2.1 The Hierarchical Cluster Analysis (HCA)

The HCA is an algorithm for clustering the given data. It regards all the data input as a single cluster and then recursively divides each cluster into two subclasses. It enjoys a relatively long history in the study of communicative intentions, including the Speech Act Theory. In the 1960s, scholars had already proposed that human communicative behavior could be structured hierarchically (Scheflen, 1965; Scheflen, 1967). Some researchers then innovatively employed hierarchical organizations for speech acts to analyze specific types of discourse, e.g., therapeutic discourse (Labov and Fanshel, 1977) or interpersonal behavior (D'Andrade and Wish, 1985).

Furthermore, some pragmaticians in recent years started to analyze the data in their experiments with the HCA, especially when they probed into the relationship between existing classificatory schemes and people's perception of a given set of speech acts (Holtgraves, 2005; Liu, 2011). Though word frequency and other textual indices were not applied in their studies, it can be revealed that the HCA may be effective in speech act-related research.

As DMs indicate contextual information and pragmatic relationship, their frequency of use in utterances could be seen as an indicator of speech act. It is then plausible to examine whether these objective indices can be hierarchically clustered in a way that demonstrates the functional similarities and variations between different speech acts.

2.2 The Switchboard Dialog Act Corpus (SwDA)

The SwDA consists of 1,155 five-minute conversations, including around 205,000 utterances and 1.4 million words from the Switchboard corpus of telephone conversations (Jurafsky et al., 1997; Potts, 2022). The dialogs in this corpus all happened between two individuals of different ages, genders, and education levels, and the speech acts of speakers were annotated according to how participants might expect one sort of conversational units to be responded to by another. One of the SwDA's merits is that there can be more than one speech act within each utterance. This annotation scheme perfectly corresponds with the ideas of Labov and Fanshel, who criticized the one-utterance-to-one-speech act method of identifying speech acts in dialogs (Labov and Fanshel, 1977). In this regard, the results obtained through the SwDA may be an accurate reflection of the speech act patterns in human beings' daily dialogs.

According to Jurafsky et al. (1997), there are four sets of speech act hyper-categories that have enough data and meaningful sub-categories – *Agreement, Understanding, Answer,* and *Information Request*. With the 27 kinds of speech acts and the 11 DMs in the four hyper-categories, statistical tests can be conducted to get reliable results. Moreover, traditional speech act classifications such as Searle's (1976), though important, may have some defects, e.g., their abstractness and the overemphasis on speakers. Thus, the SwDA can serve as an ideal research material by virtue of the following attributes.

First, the corpus makes a more detailed and clear distinction between the speech acts of agreement and disagreement. According to Jurafsky et al.'s (1997) classification criteria, speech acts expressing speakers' attitudes are distinguished into a continuum containing five subcategories – direct approval (Agree/Accept), partial approval (Maybe/Accept-part), hold before positive answers (Hedge), partial

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.

negation (Dispreferred Answers/Reject-part), and direct negation (Reject). All of them were annotated based on Allen and Core's (1997) decision tree (see Figure 1), which helped control the subjectivity and the disagreements of the annotators.

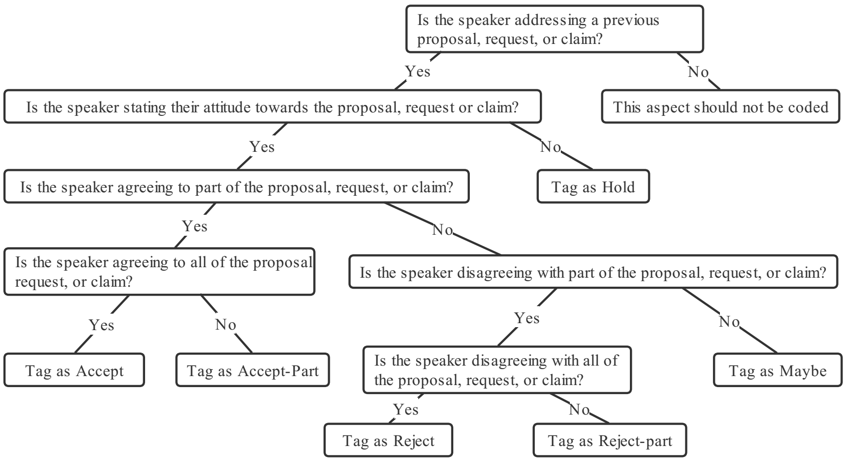


Figure 1: The decision tree for annotating speech acts in the agreement-disagreement continuum.

Second, the SwDA was annotated based on a shallow discourse tag set, which can reduce the abstractness of the speech acts owing to the more direct description of human communicative intentions. In addition to that, eight labelers involved in the project spent about half an hour on labeling each conversation (the conversations lasted five minutes on average). The labeling accuracy and the impact of labelers' subjectivity was evaluated by the *Kappa statistic* (Carletta, 1996; Carletta et al., 1997; McHugh, 2012), and the average pair-wise *Kappa* was .80, which indicated that the annotating results were acceptable (Jurafsky et al., 1997).

We could thus explore not only whether the frequencies of DMs can effectively distinguish speakers' affective attitudes through the HCA, but also whether they can distinguish properties such as the degree of indirectness in communicative acts.

3 Results and Discussions

In this paper, the DM system (*oh, well, and, but, or, so, because, now, then, I mean,* and *you know*) proposed by Schiffrin (1987) was employed to explore whether the frequencies of DMs can reflect the affinity relationship between different speech acts. This system, containing commonly occurring DMs and widely accepted by the academic community, can capture how the diverse DMs with distinctive functions demonstrate the similarities and peculiarities among different communicative intentions. As mentioned in Section 2, the current study analyzed the five speech acts in the SwDA that express agreement or disagreement because they present a typical continuum, which facilitates a more detailed examination of the results of data analysis.

The original text files of the five speech acts were firstly compiled. The five speech acts of *Agreement* contain altogether 24,816 words, among which Agree/Accept has 19,942 words, Maybe/Accept-part 528 words, Hedge 2,703 words, Dispreferred Answers 1,772 words, and Reject 942 words). Next, we applied Antconc 4.0.5 (Anthony, 2021) to automatically get the total word count in the five speech acts, in which the frequency data of the eleven selected words/phrases proposed by Schiffrin (1987) were extracted. Since the automatic process could not distinguish between the 11 words/phrases as DMs and other cases, the authors manually checked the automatically collected data to obtain the exact DM frequencies for follow-up analyses.

It should be noted here that the raw number of DM frequencies may affect the results of the statistical analysis due to the large variation in the total number of words in each speech act. In this regard, this paper calculated the percentages of each DM's frequency relative to the total word number to standardize the data. When converting the numbers to percentages, the authors distinguished two different kinds of

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.

DMs: one for single words (unigram), such as *oh, well*, etc., and the other for two consecutive words (bigram), *I mean* and *you know*. For the former types of DMs, we counted the percentages of DM frequencies relative to those of all unigrams in each speech act; for the latter, we calculated those of all bigrams, with the aim of making the standard uniform.

After collation and calculation, the frequency data of the eleven DMs proposed by Shiffrin (1987) under each speech act were obtained, as shown in Table 1.

Speech act	oh	well	and	but	or	so	because	now	then	I mean	you know
Agree/Accept	0.0318	0.0117	0.0036	0.0015	0.0003	0.0011	0.0004	0.0004	0.0000	0.0031	0.0005
Maybe/Accept-part	0.0019	0.0455	0.0019	0.0019	0.0038	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019
Hedge	0.0137	0.0396	0.0133	0.0074	0.0007	0.0067	0.0011	0.0037	0.0007	0.0033	0.0004
Dispreferred Answers	0.0045	0.0796	0.0034	0.0028	0.0000	0.0017	0.0000	0.0017	0.0006	0.0051	0.0006
Reject	0.0149	0.0658	0.0042	0.0138	0.0032	0.0011	0.0011	0.0021	0.0011	0.0053	0.0053

Table 1: The proportion of each DM to the total unigrams or bigrams under each speech act.

As can be seen from Table 1, there are significant differences in the proportion of DMs under each speech act, especially the difference between the proportion of *well* in the speech act of agreement and that of disagreement, in which the frequency of *well* is significantly higher than that in the speech act of agreement. In addition, the frequency of *well* in indirect speech acts is higher than that in the direct ones (Dispreferred Answers > Reject > Maybe/Accept-part > Hedge > Agree/Accept). This pattern may indicate a face-saving strategy at work in the politeness principle.

The following excerpts from the SwDA further illustrate the differences between agreement and disagreement as well as those between direct and indirect speech acts.

A. *Dispreferred Answers*
1) Well, I, I think, uh, my background is probably what absolutely turned me off with sixty minutes.
2) Well, I heard tonight on the news that he is willing to come down.
3) Well, I, I, I come from kind of a biased opinion because I'm a, a therapist and a drug and alcohol.
4) Well, that was, you know, with a, with a circular saw.

From the utterances containing *well* in the speech act of Dispreferred Answers, we can see that *well* mainly serves to provide a buffer for the subsequent words. In addition, since the speaker wants to express opposition to the words spoken by the hearer without completely opposing them, he or she tends to use the strategy of repetition (e.g., the repetition of *I* in A. 1) and A. 3)) or continue to apply other DMs as filler words to further moderate the illocutionary force of the speech act of opposition (e.g., *you know* in A. 4)). This phenomenon shows that people would frequently resort to the buffer DM *well* along with other means to minimize the force of opposition they are expressing.

B. *Reject*
1) Well, I don't think you can mail thing, guns through the mail.
2) Well, I doubt that.
3) Well, yes.

When expressing direct opposition to another speaker's opinions, the frequency of *well* is also higher due to the principle of politeness and the consideration of face-saving strategy. Although Reject and Agree/Accept are both direct speech acts, the use of buffer words like *well* in direct disagreement is still significantly higher than that in direct agreement (Reject: 0.0658 > Agree/Accept: 0.0117).

C. *Maybe/Accept-part*
1) Well, even if it's not technical. If it's, uh, some social thing or whatever. It doesn't matter.

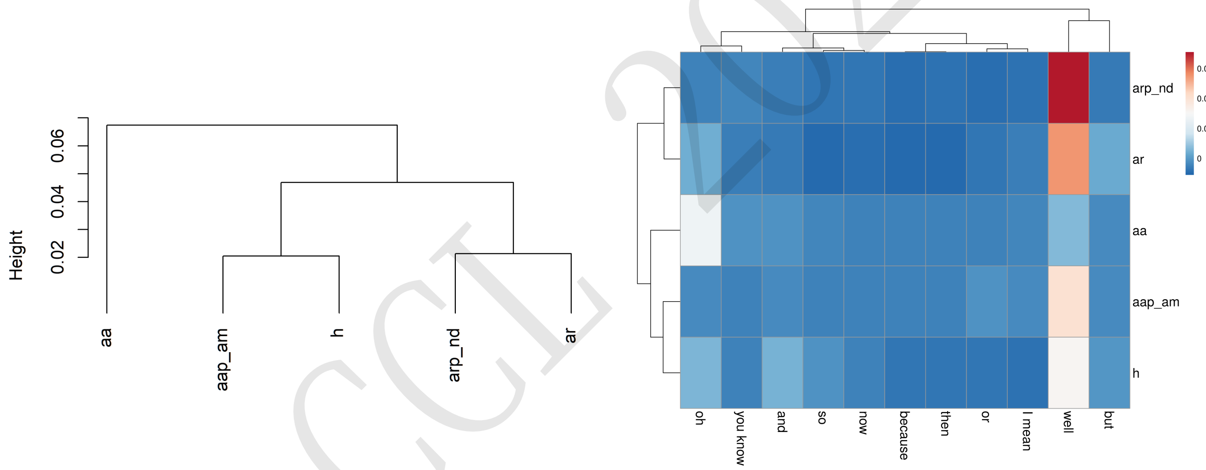
D. *Hedge*
1) Well, uh, it's funny, when I tried, to make the call the other days,

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.

E. *Agree/Accept*
1) Oh, well yeah.
2) Well, that's true.

Among the three speech acts concerning agreement (Agree/Accept, Maybe/Accept-part, and Hedge), the use of *well* is more convergent, serving as a simple tone buffer, and does not involve a strategy of face protection for the other interlocutor. According to previous studies on *well*, it is often employed as a delay device and a pragmatic marker of insufficiency, indicating the problems with the content of the current or the previous utterances, or as a face-threat mitigator, showing the conflicts in the interpersonal level (Jucker, 1993). Although *well* has a relatively fixed spectrum of discourse functions, its frequency of occurrence varies across discourses expressing different communicative intentions, depending on the specific context and the nature of probability within speakers' language use. Therefore, to accurately capture how the frequency of *well* in different speech acts reflects their affinities, it is best to apply a more suitable method to study these probabilistic linguistic units.

Moreover, from the above analysis, *well* is a DM that can effectively distinguish between agreement and disagreement; however, people cannot merely use *well* when expressing these communicative functions; DMs such as *you know* and *and* also frequently occur in these speech acts. In order to comprehensively and systematically grasp how the frequency of DMs reflects the differences of each speech act, we included in the present study a more comprehensive DM system (that proposed by Schiffrin). Meanwhile, to avoid the overwhelming workload caused by manual qualitative analysis, we adopted established statistical methods to grasp the characteristics embodied in DMs accurately. The *factoextra* and *cluster* packages in *R* were applied to perform an HCA on the data in Table 1. The results are shown in Figure 2.



(a) (b)

Figure 2: (a) The HCA results of the five speech acts in *Agreement*. (b) The HCA results of speech acts (in rows) using the Manhattan distance and the Ward.D2 method. *aa* is referred to as Agree/Accept, *aap_am* is Maybe/Accept-part, *h* is Hedge, *arp_nd* is Dispreferred Answers, and *ar* is Reject.

Figure 2a demonstrates that the clustering results based on the frequency of the eleven DMs neatly reflect the functions of the five speech acts under the *Agreement* hyper-category. The results show a tripartite classification, with Dispreferred Answers and Reject clustered together, Maybe/Accept-part and Hedge in the same cluster, and Agree/Accept in a separate cluster out of the above four speech acts. Hence, we can roughly get a "reject" cluster and an "accept" one in *Agreement*. Nevertheless, this result still has some imperfections: Agree/Accept is clustered out of the other four speech acts, while its function is similar to the "accept" category. After trying different method-distance combinations of the HCA, it was found that the aforementioned classification enjoys the highest probability of occurrence.

We then further altered the combination of clustering methods and distances and found that using the Manhattan distance together with the Ward.D2 method produced a clustering result consistent with the functional division of the speech acts in *Agreement* (see Figure 2b)[0]. Moreover, the top panel in Figure 2b displays the clustering result of each DM based on their frequency of use. The cluster of *well* and *but* further corroborates our previous analysis of *well*'s frequent appearance in the speech acts concerning disagreement.

After obtaining the above clustering results, we employed the *cluster* package in *R* to get the proportion of each DM in each cluster for a more detailed analysis. The distribution of DMs' frequency proportions when there are two clusters (henceforth Type A clustering) and three ones (henceforth Type B clustering) are shown in Table 2 and Table 3, respectively.

Cluster	oh	well	and	but	or	so	because	now	then	you know	I mean
A	0.0158	0.0322	0.0063	0.0036	0.0016	0.0026	0.0005	0.0014	0.0002	0.0021	0.0009
B	0.0097	0.0727	0.0038	0.0083	0.0016	0.0014	0.0005	0.0019	0.0008	0.0052	0.0029

Note: Cluster A is the cluster of Agree/Accept, Maybe/Accept-part, and Hedge, and Cluster B is Dispreferred Answers and Reject.

Table 2: The percentages of DM frequencies in different clusters (Type A).

From the data in Table 2 and Table 3, the reason Agree/Accept is separated as an individual cluster in Figure 2a is probably because *oh* appears significantly more frequently in it than in other speech acts. After analyzing the original corpus data, it was found that *oh* usually appears in expressions such as “Oh yes” or “Oh yeah”, which constitute a typical feature of Agree/Accept compared with other speech acts. Also, the high frequency of *oh* indicates that most clustering methods are influenced by individual salient values, which lead to the changes in specific cluster branches. In addition, the results in Table 2 and Table 3 show that the frequency of *well* is significantly higher when it expresses negative views than when it expresses positive ones. Since the other DMs accounted for lower frequencies and contributed less to the clustering results, the results obtained in this study may indicate that the two DMs, *well* and *oh*, are more effective in distinguishing between the speech acts of agreement and disagreement. This result also further complements the previous studies on the principle of politeness and the face theory, providing new perspectives for future systematic research on pragmatic principles with large-scale corpus data.

Cluster	oh	well	and	but	or	so	because	now	then	you know	I mean
C	0.0318	0.0117	0.0036	0.0015	0.0003	0.0011	0.0004	0.0004	0.0000	0.0031	0.0005
D	0.0078	0.0425	0.0076	0.0046	0.0023	0.0033	0.0006	0.0018	0.0004	0.0017	0.0011
E	0.0097	0.0727	0.0038	0.0083	0.0016	0.0014	0.0005	0.0019	0.0008	0.0052	0.0029

Note: Cluster C is the cluster of Agree/Accept, Cluster D is Maybe/Accept-part and Hedge, and Cluster E is Dispreferred Answers and Reject.

Table 3: The percentages of DM frequencies in different clusters (Type B).

In summary, from the above analysis, it can be concluded that by employing a method that can accurately grasp the statistical patterns of linguistic units, we may be able to better capture the tendency of each speech act in using DMs and establish the connection between the two important constructions (*speech acts* and *DMs*) with the support from real data. This approach can complement well-developed qualitative analyses of DMs, provide more comprehensive and theoretically supported results (e.g., the DM classification system proposed by Schiffrin), and introduce the advantages of quantitative analysis (big data, objectivity, and accuracy) into the research related to pragmatics and discourse analysis.

4 Conclusions and Implications

In this study, we adopted a quantitative approach to analyze whether DMs, the discourse units that possess discursive and pragmatic information, can effectively distinguish the speech acts of different communicative functions. After calculating the frequencies of the 11 DMs proposed in Schiffrin (1987), we

[0]For all the clustering results using different methods and distance metrics, see Appendix A.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.

conducted an HCA using *R* for the examination of such effects. The results showed that the frequencies of DMs were efficacious in differentiating the speech acts of agreement and disagreement. Moreover, the frequencies of DMs also well reflect the intricacies within the indirectness of the five speech acts in the agreement-disagreement continuum, corroborating that DMs are rather precise indicators of speech acts' differences.

The results also indicated that the frequencies of *well* and *oh* might be the key indicators to distinguish between the speech acts of agreement and disagreement, especially *well*, the frequencies of which echo the previous findings in the principle of politeness and the face theory. In this regard, the application of quantitative measures for testing and generalizing the existent theoretical framework may help the research related to pragmatics and discourse analysis develop in a scientific and precise direction. The deficiencies of traditional qualitative research in terms of data size can thus be supplemented by conducting research on the authentic data from large-scale corpus.

In addition, since the current study only examined the five speech acts under the continuum of agreement and disagreement, the patterns found may not fully reflect the patterns in all types of speech acts. Therefore, subsequent studies can further collect the natural corpus data of human conversations and examine more types of speech acts to further explore the effectiveness of DM frequency in reflecting human conversational behaviors. In this way, we may establish a more comprehensive framework for quantitative research in pragmatics and discourse analysis.

Acknowledgements

The authors are grateful to the three anonymous reviewers for providing helpful feedback on this paper.

References

James Allen and Mark Core. 1997. Draft of DAMSL: Dialog act markup in several layers. http://www.fb10.uni-bremen.de/anglistik/ling/ss07/discourse-materials/DAMSL97.pdf.

Laurence Anthony. 2021. Antconc 4.0.5. Waseda University.

Mira Ariel. 1998. Discourse markers and form-function correlations. In Andreas H. Jucker and Yael Ziv, editors, *Discourse Markers: Descriptions and Theory*, pages 223–260. John Benjamins, Philadelphia.

Jean Carletta, Amy Isard, Stephen Isard, Jacqueline C. Kowtko, Gwyneth Doherty-Sneddon, and Anne H. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

Jean Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Roy Goodwin D'Andrade and Myron Wish. 1985. Speech act theory in quantitative research on interpersonal behavior. *Discourse Processes*, 8(2):229–259.

Kerstin Fischer. 2006. Frames, constructions, and invariant meanings: The functional polysemy of discourse particles. In Kerstin Fischer, editor, *Approaches to Discourse Particles*, pages 427–447. Elsevier, Oxford.

Bruce Fraser. 1996. Pragmatic markers. *Pragmatics*, 6(2):167–190.

Péter B. Furkó. 2020. *Discourse Markers and Beyond: Descriptive and Critical Perspectives on Discourse-Pragmatic Devices across Genres and Languages*. Springer International Publishing, Cham.

Lynne E. Hewitt, Judith F. Duchan, and Erwin M. Segal. 1993. Structure and function of verbal conflicts among adults with mental retardation. *Discourse Processes*, 16(4):525–543.

Thomas Holtgraves. 2005. The production and perception of implicit performatives. *Journal of Pragmatics*, 37(12):2024–2043.

Scott Jacobs and Sally Jackson. 1981. Argument as a natural category: The routine grounds for arguing in conversation. *Western Journal of Speech Communication*, 45(2):118–132.

Andreas H. Jucker. 1993. The discourse marker well: A relevance-theoretical account. *Journal of Pragmatics*, 19(5):435–452.

Proceedings of the 21st China National Conference on Computational Linguistics, pages 728-737, Nanchang, China, October 14 - 16, 2022.

Daniel Jurafsky, Liz Shriberg, and Debra Biasca. 1997. Switchboard swbd-damsl shallow-discourse-function annotation coders manual, draft 13. https://web.stanford.edu/˜jurafsky/ws97/manual.august1.html.

William Labov and David Fanshel. 1977. *Therapeutic Discourse: Psychotherapy As Conversation*. Academic Press, New York.

Si Liu. 2011. An experimental study of the classification and recognition of chinese speech acts. *Journal of Pragmatics*, 43(6):1801–1817.

Mădălina Matei. 2010. Discourse markers as functional elements. *Bulletin of the Transilvania University of Braşov*, 3(52):119–126.

Mary L. McHugh. 2012. Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3):276–282.

Michael Mulkay. 1985. Agreement and disagreement in conversations and letters. *Text - Interdisciplinary Journal for the Study of Discourse*, 5(3):201–228.

Christopher Potts. 2022. Switchboard dialog act corpus with penn treebank links. https://github.com/cgpotts/swda.

Gisela Redeker. 2006. Discourse markers as attentional cues at discourse transitions running head: Discourse transitions. In Kerstin Fischer, editor, *Approaches to Discourse Particles*, pages 339–358. Brill, Leiden, The Netherlands.

Albert E. Scheflen. 1965. *Stream and Structure of Communicational Behavior: Context Analysis of a Psychotherapy Session*. Eastern Pennsylvania Psychiatric Institute.

Albert E. Scheflen. 1967. On the structuring of human communication. *American Behavioral Scientist*, 10(8):8–12.

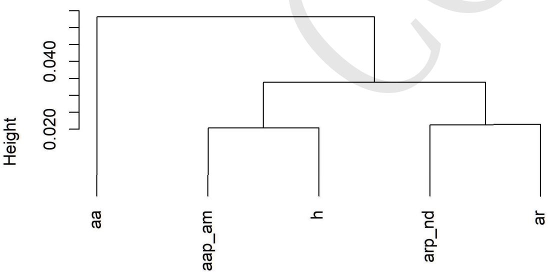
Deborah Schiffrin. 1987. *Discourse Markers*. Studies in Interactional Sociolinguistics. Cambridge University Press, Cambridge.

John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge.

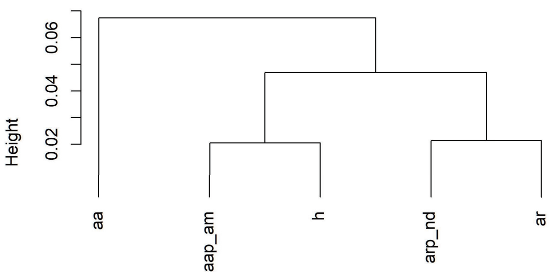
John R. Searle. 1976. A classification of illocutionary acts. *Language in Society*, 5(1):1–23.

Carmelina Trimboli and Michael B Walker. 1984. Switching pauses in cooperative and competitive conversations. *Journal of Experimental Social Psychology*, 20(4):297–311.

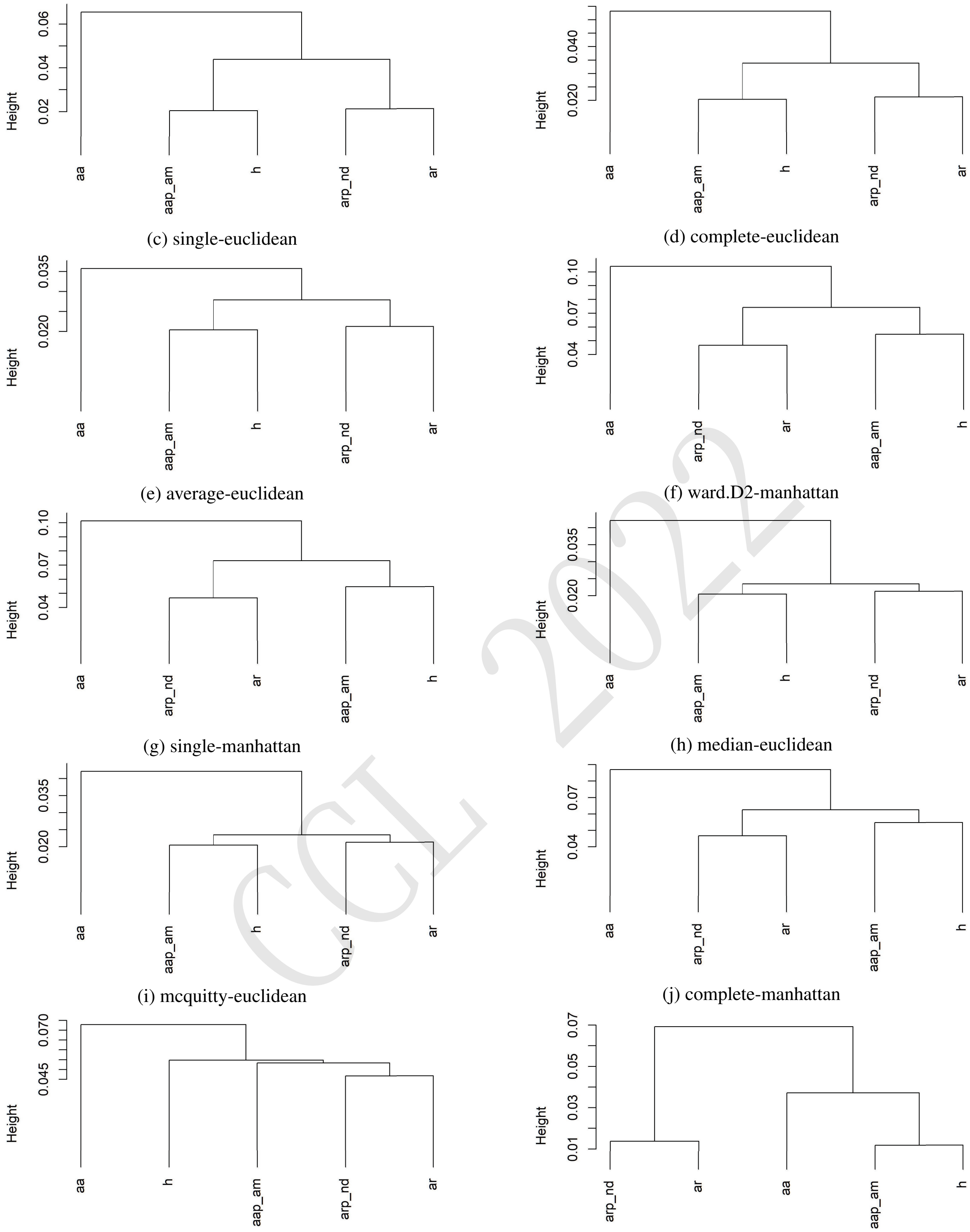
Appendix A. The Clustering Results of the Frequencies of Discourse Markers in the Speech Acts of Agreement.



(a) ward.D-euclidean



(b) ward.D2-euclidean



(c) single-euclidean
(d) complete-euclidean
(e) average-euclidean
(f) ward.D2-manhattan
(g) single-manhattan
(h) median-euclidean
(i) mcquitty-euclidean
(j) complete-manhattan
(k) average-manhattan
(l) ward.D2-maximum

Figure 3: The clustering results of the frequencies of DMs in the speech acts of *Agreement*.