

A REVIEW OF SPEECH ACT RECOGNITION

Zhao Chong-yuan

School of Foreign Languages and Literature, Lanzhou University , Lanzhou, China

E-mail: zhaozhy2019@lzu.edu.cn

ABSTRACT: *Recognizing the speech act that a speaker performs with an utterance is fundamental of pragmatic competence. Theories about pragmatic language propose that speech comprehension involves speech act recognition which relies on the hearer's inference of the speaker's intention. Speech act recognition, which can be triggered by specific context, adjacency pair or performative verb, plays a crucial role in the comprehension of communicative acts and the success of conversation. The present study provides a detailed review of the researches on speech act recognition and pays attention to a possible aspect of further study on speech act recognition.*

KEYWORDS: *speech act theory, intention recognition, conversation comprehension*

INTRODUCTION

Language comprehension covers not only understanding of the propositional content, but also the contextualization of discourse. Theories about pragmatic language propose that speech comprehension involves the recognition of the speaker's intentions. With regard to intention recognition one important approach is the Speech Act Theory. The present study introduces the origin and development of Speech Act Theory, including different linguists' views on speech act categorization and their proposals of intention recognition in conversation comprehension.

Speech Act Theory

Speech Act Types

Speech act theory (Searle, 1969) has been an influential approach to language use. It represents one of the major lines of research in pragmatics and discourse analysis (Holtgraves and Ashley, 2001). John Austin's book *How to Do Things with Words* (1962) represented the beginning of speech act theory. According to Austin, a speaker may perform three simultaneous acts in one utterance, that are the locutionary act, the illocutionary act and the perlocutionary act. The locutionary act is the actual words that the speaker is saying; the illocutionary act is the intention of the speaker conveyed by means of illocutionary force; while the perlocutionary act is the effect of the utterance on the hearer. Austin distinguished implicit speech act and explicit speech act and he divided speech acts into five types: verdictives, exercitives, commissives, behabitives and expositives.

John Searle (1969) further revised and developed Austin's theory and proposed the concept of indirect speech act and illocutionary force. According to Searle, utterances have both a propositional content and an illocutionary force symbolized as $F(p)$, where F stands for illocutionary force, the action side of every speech act, and p refers to proposition, the content side of every speech act. The illocutionary force of an utterance is the speaker's intention in producing that utterance (Liu, 2011). Different utterances, which convey the same propositions under varied situations, may express different illocutionary forces. Searle criticized Austin's taxonomy in several aspects, one aspect was that Austin's classification of speech acts was merely the categorization of speech act verbs, and his theory lacked theoretical foundations and principles. First, Austin regarded illocutionary act as conventional while only a small part of illocutionary acts in real life were conventional. Second, Austin's taxonomy was largely based on the differences of speech act verbs. Furthermore, Austin mismatched perlocutionary acts with results as the relationship of cause and effect. While in order to judge whether the speaker has performed a certain kind of speech act, one must take the speaker's intention in expressing that utterance into consideration. Combined with the illocutionary force, Searle classified speech acts into five groups: assertives, directives, commissives, expressives and declarations, and he based this classification on four criteria, which are generally called "felicity conditions":

- a. illocutionary point/speech act type
- b. direction of fit
- c. expressed psychological state
- d. the content of the act/ the proposition content

Explicit Speech Act vs. Implicit Speech Act

Austin distinguished implicit speech act and explicit speech act. Explicit speech acts (referred to as explicit performatives by Austin) are utterances that contain the performative verb, the verb that names the speech act being performed with the utterance. One can promise to shut the door, for example, by simply saying "I promise to shut the door". In contrast, implicit speech acts are utterances performing a speech act that do not contain the performative verb. The former explicit speech act can also be performed implicitly with "I guarantee that I will shut the door". Furthermore, implicit speech acts may be more common than utterances that contain speech act verbs (Thomas, 2008).

Speech Act Recognition

Speech act recognition is the process of recognizing the action of an utterance in a given context, it is also the hearer's inferencing process of the speaker's intention underlying that specific utterance. Thus, speech act recognition usually associates with implicatures, which are heavily related to implicit speech acts or indirect speech acts.

The notion of “speaker meaning” has remained a core focus of research in pragmatics since Grice’s seminal work on so called non-natural meaning and speaker intention has engendered a move to study meaning goes beyond what is said. Grice defined the speaker meaning as the speaker meant something by uttering an utterance and the speaker uttered that utterance with the intention to produce some effect in an audience by means of the recognition of this intention (Haugh, 2013).

The recognition of other’s intention can lead to a successful communication. How is it that addressees will recognize the speech act that is being performed with this remark? In general, speech act theorists have suggested that sentence type and mood, intonation, back-ground knowledge, and other relevant features of the context come into play and aid the addressee’s recognition of the speech act performed (Austin, 1962; Searle, 1969). Speech act recognition is also related to cultural differences since culture is closely connected with our communication styles. Edward Hall (1976) proposed the concept of high-context and low-context communication. In high-context cultures, communication style is influenced by the closeness of human relationships, well-structured social hierarchy, and strong behavioral norms. Internal meaning is usually embedded in the information, thus not everything is stated in writing or spoken. The communication in high-context cultures is indirect and ambiguous. Hearers need to infer the nonverbal aspects of communication and the cognitive load of speech act recognition is much higher than that in low-context cultures. While in low-context cultures meanings are explicitly stated through language. Direct and linear communication and constant use of words represent the typical features of low-context culture. Communication is direct, precise, dramatic, open and based on feelings and true intentions (Hall and Hall, 2001; Nishimura et.al., 2008).

LITERATURE REVIEW OF SPEECH ACT RECOGNITION

A large amount of study has focused on whether utterance comprehension involves speech act recognition. Holtgraves (2001, 2005, and 2008) examined speech act classification and the automatic recognition of illocutionary acts for English native speakers and second language learners. His research turned out that English native speakers could recognize speech acts in an automatic way, which was true for both spoken and written forms. On the contrary, foreign language learners could not perform in the same way as native speakers. Besides, he found that English speakers tended to classify speech acts based on perlocutionary effects rather than the illocutionary points, which was how Searle classify speech acts. Holtgraves (2008) also proposed that speech acts captured in a single word the action a speaker is performing with an utterance, and this allows for efficient processing of conversation turns. Later, Si Liu (2011) took Holtgraves’ methods and tested how Chinese speakers classify speech acts and if utterance comprehension was involved speech act recognition when the language was Chinese. She found that illocutionary force recognition was involved in speech act

interpretation and Chinese native speakers tended to classify speech acts based on perlocutionary acts. Licea-Haquet et al (2019) used Holtgrave's methods and conducted experiments to explore Spanish speakers' recognition and processing of speech acts. His research revealed that compared to native English and Chinese speakers, native Spanish speakers took longer to identify speech acts and he found evidence for automatic speech act recognition in Spanish speakers in one of the two tasks. He further explored the cognitive functions behind speech act recognition and found that the Theory of Mind and Executive Functions were two main cognitive functions that closely related to pragmatic inference during conversation comprehension. Other studies also examined the role of context and the co-context (such as the adjacency pair) in speech act recognition. The following are the general three aspects concerning previous researches on speech act recognition.

The automaticity of speech act recognition: priming effect

In linguistic studies, priming refers to the phenomenon in which prior exposure to language, to some extent, influences subsequent language processing. It is a subconscious process and one manifestation of human implicit memory that is responsible for skills, habits. Semantic priming refers to a general tendency for language users to show facilitation in their processing of words due to a previous experience with words similar in meaning. Tasks in semantic priming experiment include lexical decision task, pronunciation (naming) task and semantic categorization task. Lexical decision task presents to participants with a certain interval (stimulus onset asynchrony, SOA) and participants see (or hear) one letter strings (prime) and are asked to decide whether the second letter strings (target) constitutes a real word in a given language.

According to Posner and Snyder (1975) there are two types of processing: automatic (rapid) and attentional (time-consuming). Based on this, if speech act activation is an automatic rather than controlled process, then activation must occur very quickly. Although there are no exact cut-offs, priming effects that occur at SOAs longer than 1500 ms are generally taken as reflecting a controlled process. Those occurring at 250 ms or less are viewed as automatic processes (Posner and Snyder, 1975).

Past research has demonstrated that native speakers of English automatically recognize speech acts when they comprehend utterances (Holtgraves & Ashley, 2001). Holtgraves (2007 b) examined whether automatic speech act recognition occurs for participants learning English as a second language and he came to the conclusion that speech act recognition might be an automatic process for L1 but not L2 individuals. Holtgraves (2008) explored the time course of speech act recognition through lexical decision task, and he examined speech act activation at short (250 ms) and long (2000 ms) SOAs. He found priming effect of 50.7ms at the short delay, which turned out to be consistent

with his previous hypothesis: the activation of speech acts was an automatic process that occurred quite early. Thus, the activation of an implicit speech act facilitates participants to respond faster at this task.

Furthermore, recognition of speech acts is not made at the final stage in the comprehension process, occurring at the last word of incoming utterances, but takes place early on when the turn has only been partially processed (Rosa S. Gisladdottir et al. 2015). Therefore, triggers of the speech act play a key role in the process of activation of speech acts.

Illocutionary Force Indicators

An important feature of speech acts is that there is no one-to-one mapping between illocutionary force and a specific utterance (Holtgraves, 2005). The same utterance can be used to convey different meanings under varied scenarios, how can speakers' intention or the illocutionary force be recognized by the listeners? One important aspect is the illocutionary force indicators. Speech act can be revealed directly by performative verbs and indirectly through contexts, syntactic features (such as wh-question words, imperatives and interrogative word order). These devices are referred to as the illocutionary force indicators. Whereas speakers do not usually use these indicators in daily interaction, instead, implicit speech acts that without performative verbs are quite common. Speech act inference and intention recognition occur frequently when the utterances are underspecified and lack of direct illocutionary force indicators.

Context

The propositional content and the context generally provide the hearer with the basis for identifying the speaker's intention. Context matters a lot in speech act recognition, but there are also evidences that speech act recognition occurs quite early and even before the utterance is presented. Here the notion of implicature should be mentioned since it associates with speech act recognition closely. Grice (1975) explored the generalized and particularized implicatures and made a distinction between them. According to Grice, generalized implicatures are implicatures that arise without the need for any reference to the context, such as metaphors, idioms, and conventional indirect requests. Particularized implicatures require a relatively time-consuming inference process and cannot be recognized apart from the context within which they are used. It is clear that the recognition of implicit speech acts represents a generalized rather than a particularized implicature (Grice, 1975). Recognition of the intended meaning of generalized implicatures appears to be a relatively automatic process (Holtgraves, 2008). Based on Grice's theory, Holtgraves (2008) conducted experiment to explore the role of the conversation context in speech act activation by means of deleting the contextual descriptions. Participants were asked to judge the speech acts only by the target utterances. He concluded that a context-based inference process was

not required for speech act recognition. Other studies have also demonstrated that automatic speech act activation for native speakers can occur in the absence of any conversational context, suggesting that a context-based inference process is not necessarily required for speech act recognition, but that does not mean that context does not matter. Speech acts, of course, never occur in a vacuum; there is always a context. Additionally, it might be that the background context plays a relatively greater role in L2 speech act comprehension (Taguchi, 2002). Furthermore, implicit speech acts function as generalized implicatures for native speakers and particularized implicatures for non-native speakers, which attend more closely to the context in order to interpret the speaker's meaning (Holtgraves, 2007 b).

Performative verbs

One important distinction in research on speech acts can be made between explicit and implicit speech acts. Owing to this, speech act verbs cannot be thoroughly analyzed without inspecting their pragmatic use. Implicit speech acts are speech acts conveying the illocutionary force (speaker's intention) not by means of performative verbs, but instead by the specific context or the auditory indications. Thus, the transition from the surface form to the underlying meaning represents the hearer's inference of the speaker's purpose for producing that utterance. Explicit speech acts are speech acts that are relatively clear and direct and contain the performative verb, the verb that names (in the appropriate contexts) the speech act it performs (Holtgraves, 2005). And they are easier to recognize compared with the implicit one with the assistance of performative verbs.

Acoustics cues for speech act perception

Hellbernd and Sammler (2016) investigated whether and how prosody (the vocal tone) contributes to the identification of "unspoken" intentions. Their research showed that characteristic prosodic feature configurations for different intentions that were reliably recognized by listeners. The speaker's intentions are represented by the prosodic signal which can determine the success of the interpersonal communication.

Sequential context: adjacency pair

Prior talk in conversation is not merely the background to speech act comprehension, but rather contains a rich structure of action sequences that could proactively funnel possible interpretations of upcoming talk (Rosa S. Gisladdottir et al. 2015). Adjacency pair, such as question-answer, invitation-acceptance/rejection, greeting-greeting, apology-forgiveness et al., provide crucial sequential cues for inferencing speaker's intentions. Besides, the same target utterance under different prior turns can convey different verbal actions. For example:

- a) -How are you going to pay for the ticket?

-I have a credit card. (question-answer)

b) -I can lend you money for the ticket.

-I have a credit card. (offer-rejection)

c) -I don't have money for the ticket.

-I have a credit card. (request-pre-offer)

Researches on adjacency pair offered a meaningful way to study speech act recognition under co-context content.

Cognitive functions behind speech act recognition: TOM, EF and Empathy

Theory of mind (TOM) is proposed by Premack and Woodruff (1978) to underscore the unobservability of mental states. It is a cognitive process which refers to the ability to make inferences about the others' mental states and use them to understand, judge and predict others' behavior. TOM ability can be categorized into two subsystems: cognitive TOM, which denotes the ability to attribute thoughts, beliefs, and intentions; and affective TOM, which reflects the understanding of feelings, emotions, and affective states of others (Yildirim, 2020). Besides, there are another distinction needed being explained between explicit and implicit TOM. Explicit TOM reasoning investigates a person's ability to infer the mental state of another person. Implicit TOM reasoning investigates the ability to comprehend what a person thinks, knows or believes about another person's mental state, and they require a greater cognitive load in order to be understood. Experimental evidences support that patients with autism and RHD (right-hemisphere damage) have difficulties in pragmatic communication. The existing tests for TOM fall into three classes: emotion recognition tests; cognitive and affective mentalizing tasks that measure attribution of beliefs, intentions, desires, and emotions; and multidimensional measures that combine these features (Turner & Felisberti, 2017). Executive functions (EF) are argued to drive motivated adaptative behavior, and allow healthy individuals to respond to novel and challenging tasks. They coordinate behavior, enabling a person to use their cognitive abilities in various situations in a flexible manner. An intact EF system is necessary for normal individuals to engage in motivated, adaptive, and effective communication (Cummings, 2017). There are three core EFs: inhibitory control, cognitive flexibility and working memory. Inhibition involves suppression of irrelevant information or responses and allows readers to forget or suppress information that is not relevant to text meaning. Cognitive flexibility, enables readers to switch back and forth between text elements or processes and has been found to be particularly important for reading comprehension, contributing to comprehension beyond decoding skill and language comprehension and beyond other EFs. Working memory involves two simultaneous activities: storage or maintenance of information, and active processing or transformation of that stored information, which Miyake et al. (2000) call updating. Storage and active manipulation play critical roles in integrating information during reading comprehension.

Empathy refers to the ability to recognize and identify what someone else is feeling (cognitive aspect) and to share that emotional state (affective aspect) in order to react properly to social situations (Lucas-Molina et al., 2017).

Comparison of Previous Researches

First, researches have revealed that the illocutionary force recognition is involved in utterance comprehension and play a big role in conversational interaction. Second, speech act frequently presents itself as purposeful, which entails the hearer's inferencing process of the speaker's intention during communication. Third, speech act recognition entails the activation of implicit speech acts and this recognition can be assisted by the appropriate context, single performative verb, adjacency pair and so on. Last, the automatic process of recognizing implicit speech act provides a meaningful view toward study under different language backgrounds. Besides, researches have utilized experimental equipment, such as ERP, EEG and MEG to investigate pragmatic language comprehension, and there are also FRMI studies on indirect speech acts and multi-modality studies on speech acts applying pictures, utterances and audio materials. Furthermore, some psychometric and cognitive tests are used to explore the cognitive reasons behind speech acts recognition.

CONCLUSION

The present study has provided a review of the previous researches on speech act recognition. Speech act theory serves as the theoretical foundation for studies on speech act recognition. We have covered the origin and development of speech act theory, including the analysis of explicit and implicit speech acts. Researches generally focus on the automaticity of speech act recognition, illocutionary force indicators and the cognitive functions behind speech act recognition. Detailed analysis about illocutionary force indicators include the context, performative verbs, acoustic cues and sequential context. Finally, we have made a comparison among these studies and mentioned some experimental equipment and devices used. Through the time course of previous researches, cognitive functions behind speech act recognition has emerged as a new direction that has gained attention.

References

- Austin, John L. (1962). *How to Do Things with Words*. Harvard: Reprinted by Foreign Language Teaching and Research Publishing House, Beijing, 2002.
- Cummings, L. (2017). "Cognitive aspects of pragmatic disorders". In: Cummings, L. (Ed.), *Research in Clinical Pragmatics. Perspectives in Pragmatics, Philosophy & Psychology*. Springer, Cham, Switzerland, 587-616.
- Grice, H. P. (1975). *Logic and conversation*. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Liu, S. (2011). "An experimental study of the classification and recognition of Chinese

- speech acts". *Journal of Pragmatics*. 43 (6):1801-1817.
- Hall, E.T., 1976. *Beyond Culture*. Doubleday: New York.
- Hall, E.T., Hall, M.R., 2001. "Key concepts: underlying structures of culture". In: International HRM: Managing Diversity in the Workplace, 199-203.
- Haugh, M. (2013). "Speaker meaning and accountability in interaction". Special issue on focus on the speaker. *Journal of Pragmatics*. 48(1):41–56.
- Hellbernd, N., & Sammler, D. (2016). "Prosody conveys speaker's intentions: acoustic cues for speech act perception". *Journal of Memory and Language*. (88):70-86.
- Nishimura, S., Nevgi, A., Tella, S. (2008). Communication Style and Cultural Features in High/Low Context Communication Cultures: A Case Study of Finland, Japan and India. Teoksessa A. Kallioniemi (Toim.), Uudistuva Ja Kehittyvä Ainedidaktiikka. Ainedidaktinen Symposiumi (8):783-796. Retrieved from.
- Holtgraves, T. M., & Ashley, A. (2001). Comprehending illocutionary force. *Memory & Cognition*, 29: 83–90.
- Holtgraves, T. M. (2005). "The production and perception of speech acts". *Journal of Pragmatics*. (37): 2024–2043.
- Holtgraves, T. (2007 b). Second language learners and automatic speech act comprehension. *Language Learning*.
- Holtgraves, T. (2008). "Automatic intention recognition in conversation processing". *Journal of Memory and Language*. 58 (3): 627-645.
- Haquet G.L. Licea, E.P. Velásquez-Upegui, Holtgraves T., Giordano M. (2019). "Speech act recognition in Spanish speakers", *Journal of Pragmatics*. 141 (6): 44-56.
- Lucas-Molina, B., Pérez-Albéniz, A., Ortuño-Sierra, J., & Fonseca-Pedrero, E. (2017). "Dimensional structure and measurement invariance of the Interpersonal Reactivity Index (IRI) across gender". *Psicothema*. 29(4): 590–595.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). "The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis". *Cognitive Psychology*. (41): 49–100.
- Morton, J., Frith, U., & Leslie, A. (1991). "The cognitive basis of a biological disorder: Autism". *Trends in Neurosciences*. 14:434-438.
- Whiten, A. (1991). *Natural theories of mind*. Oxford: Basil Blackwell.
- Tong, Wong, Kwan, et al. (2020). "Theory of Mind as a Mediator of Reading Comprehension Differences between Chinese School-Age Children with Autism and Typically Developing Peers". *Scientific Studies of Reading*. 24(4):292-306.
- Premack, D., & Woodruff, G. (1978). "Does the chimpanzee have a 'theory of mind'?" *Behavior and Brain Sciences*. 4: 515-526.
- Gisladdottir, R. S., Chwilla, D. J. & Levinson, S. C. (2015). "Conversation electrified: ERP correlates of speech act recognition in underspecified utterances". *PLoS One*, 10(3):1-24.

- Searle, John R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, pp. 23–65.
- Taguchi, N. (2002). An application of relevance theory to the analysis of L2 interpretation processes: The comprehension of indirect replies. *International Review of Applied Linguistics*, 40:151–176.
- Turner R., Felisberti F. M. (2017). “Measuring mindreading: a review of behavioral approaches to testing cognitive and affective mental state attribution in neurologically typical adults”. *Frontiers in Psychology*. 8:47.