UTTERANCE UNITS AND EXCHANGES IN SPONTANEOUS JAPANESE DIALOGUE

Masahito KAWAMORI
Nippon Telegraph and Telephone,
NTT Basic Research Laboratories
kawamori@atom.brl.ntt.co.jp

Akira SHIMAZU
Japan Advanced Institute
of Science and Technology
shimazu@jaist.ac.jp

ABSTRACT

In dialogue, speakers are engaged in various types of speech acts other than just those frequently encountered such as question, command, promise. The speaker often tries to make sure that his interlocutor has received the information that he is conveying, that his intentions are understood, and that his attitudes toward his interlocutor and the general surrounding environment are clear. Especially in spontaneously spoken discourse or dialogue, these various speech acts are not necessarily expressed in full-fledged well-formed sentences; they are often couched in rather fragmentary, truncated semi-sentences or phrases. In an analoguous manner to Parsons' subatomic event structures, we call these speech acts corresponding to non-full-fledged sentences SUBATOMIC SPEECH ACTS. We show in this paper how these subatomic speech acts are used to form mutual beliefs and to coordinate discourse. We use examples from actual Japanese spoken dialogue to demonstrate that our theory accounts for some of the characteristic phenomena in discourse and assigns correct interpretations and explanations to these phenomena.

1 INTRODUCTION

It is often said that ordinary conversations are full of irregularities. Some say that there is no 'grammar' in every day spoken language. In theoretical linguistics, such irregularities in every day conversation has been considered due to various limitations on the performance language use. These phenomena have been regarded as rather peripheral and not a central concern of theoretical linguistics, at least that of a grammatical theory. Such an attitude, however, has made it rather difficult to apply the scientific results of theoretical linguistics to the language in actual use. Grammatical theories are particularly vulnerable to the vulgarity of conversational language because many people speak in incomplete fragmentary sentences.

But the fact still remains that people do understand everyday language and that they usually do not find their conversational language anyway more difficult to understand than its written counterpart; on the contrary, it is usually the case that people find the written language spoken to them harder to interpret than spontaneously spoken casual language. So there seems to be a sort of paradox; everyday conversational language seems to lack proper 'grammar' and regularity enjoyed by its written counterpart but is not particularly difficult to understand, while its written counterpart with a proper grammar and regular structure is usually harder, when spoken, to understand.

We believe that the paradox is only superficial. The paradox ensues if one assumes that since conversational language does not always follow the rules of its written counterpart, it lacks rules and grammar. If one accepts that conversational language has its own grammar and rules, somewhat different from those of its written counterpart, then there will be no paradox involved. In other words, what we need is a theory of performance, a grammar of language in actual use.

We will in this paper explore this aspect of spoken language looking specifically into what consititutes its basic unit.

1.1 Why Utterance Units Are Important: Characteristics Of Spoken Dialogue

Japanese is an agglutinative language with predominantly verb final sentence patterns. It is to be expected, therefore, that Japanese sentence units are rather easy to recognize, and textbooks and written forms of Japanese indeed vindicate such an expectation.

For example, the following would be a typical Japanese sentence in a Japanese grammar book [2]:

(1) *Sensei-ga   hon-o   kai-ta*
Teacher-SUBJ book-OBJ write-PAST
'The theacher wrote a book.'

Since most of those sentences appearing in a grammar book are of this form, in which verbal forms come at the end, Japanese is typically considered a SOV language. On the other hand, such features as dropping case and conjunctive particles, frequent use of end particles and interjections, and abnormal ordering of words have long been considered signs of 'agrammaticism'[5]. These facts seem to corroborate the view that in Japanese grammatical sentence units are well established and rather easy to recognize.

When one observes Japanese in actual use, however, one can see that such a view is too simple, because a spontaneously spoken Japanese discourse is full of features which are totally beyond the grasp of grammar books: phrases ending with conjunctions, interjections with verbs attached, unfinished utterances and other disfluencies.

These characteristics are often taken as the manifestation of the irregularity and non-systematic nature of spoken language. Although one may decide to ignore these as unimportant performance issues outside the domain of linguistics proper, if one wants sincerely to analyze language in its totality, one cannot avoid meeting these issues.

Especially when one wants to recognize 'sentences' in discourse, these characteristics inevitably stand in one's way, destroying the sentential structures expected by grammar. Because of these, from the stand point of the conventional grammar, clarifying such matters as which utterances constitute one sentence and which utterances belong to two different sentence units becomes extremely difficult. If the task of discerning sentence units, the building blocks of discourse, is difficult, the difficulty of discerning discourse structure is enormous.

2   UTTERANCE UNITS, SPEECH ACTS, AND DIALOGUE EXCHANGE

In a coordinated dialogue, the fundamental aspects of the intentions of the participants are conveyed by speech acts conveyed by the utterances in the dialogue.

Speech acts, or more properly illocutionary acts, are usually regarded as consisting of an illocutionary force and a proposition. In this view, an illocutionary force is comparable to a modality attached to the proposition. If we represent the proposition by P and an illocutionary force by $F(.)$ then an illocutionary act can be represented by $F(P)$[12].

For example, consider the following sentence:

(2) John is a student.

If the propositional content of this utterance is represented as:

(3) Student(John).

When this sentence is uttered with the illocutionary force ASSERT, the entire illocutionary act would be represented as:

(4) ⊢ Student(John).

Or, consider another sentence:

(5) Open the door.

Supplying the unnamed subject of this utterance, one can represent the propositional content of this utterance as:

(6) Open(you, the_door).

When this sentence is uttered with the illocutionary force COMMAND, the entire illocutionary act may be represented as:

(7) ! Open(you,the_door),

where ! represents the logical operator corresponding to the illocutionary force COMMAND.

When the propositional content of the utterance is rather clear and straightforward, such an understanding of illocutionary forces and speech acts is unproblematic, and most of the work in this area has been based on examples taken from neatly structured, well-formed sentences. But when the propositional content is unclear or when the sentence in question is not quite so neatly constructed, there would be much reconstructing to do to apply this orthodox view of illocutionary acts. Notice that even in the simple case of (5), we assumed that we safely understand the unmentioned subject of the utterance.

As has been observed above, an ordinary Japanese dialogue seldom consists of well-formed sentences like (2) that corresponds to a proposition.

For example, in the following dialogue:

(8) A1: *eeto, boku-no tukue no-ue-ni desune*
      well my desk   upon

  'Well, on my desk'
  B1: *hai*
      yes

  'Yes'
  A2: *hon   aru-deshou   yon-satu*
      book   there is, isn't there ? four

  'There are four books, aren't there?'
  B2: *a hai*
      oh yes

  'Oh, yes'
  A3: *sore-odesune*
      that

  'Those books'
  B3: *hai*
      yes

  'uh-hun'
  A4: *tyotto   motte kitekuremasuka?*
      a little   bring   can you please?

  'Can you bring them to me?'
  B4: *hai*
      yes

  'OK'

A's utterances can be interpreted as collectively expressing the speech act essentially the same as the one expressed by the following sentence:

(9) *boku no tukue no ue ni aru yon satu no hon o motte kite kuremasuka*
    my   desk existing on   four books   bring me please

  'Can you bring me the four books on my desk?'

As can be seen from these examples, utterances like *boku no tukue no ue ni* ('on my desk') and *sore o* ('that one') do not correspond to sentences but phrases or parts of sentences. But these utterances constitute, in some sense, "units" of speech.

In a way, one can see that in this sense, an actual dialogue is not a static, well-ordered act which only involves a well-formed, so-called grammatical sentence which expresses an illocutionary act corresponding to a whole proposition, but rather a dynamic act in which pieces of utterances that are smaller than what are usually considered sentences are gathered together to convey, as a whole, a spontaneous intention or speech act.

It is therefore reasonable to consider a unit of speech smaller than sentences.

If a sentence corresponds to a proposition, then these units smaller than sentences would correspond to parts of a proposition. One can call these parts of a proposition *subatomic proposition*.

If a stretch of discourse, consisting of several fragmentary utterances, proposes something then the discourse as a whole may be viewed as expressing the same speech act of proposal that would be expressed using some single well-formed sentence. If this single well-formed sentence represents a proposition, then each of the fragmentary utterances that constitute the stretch of discourse would be representing a part of the proposition, a *subatomic proposition*. And to make the paralell complete, each of the subatomic propositions would, with a certain type of, not necessarily illocutionary, force, correspond to *a subatomic speech act*, which is part of the act of proposal.

2.1 Subatomic propositions and subatomic speech acts

The term subatomic proposition is inspired by Parson's SUBATOMIC SEMANTICS [11], the framework of semantics that studies entities beneath the level of atomic sentences.

Parson's theory is close in spirit to Davidson's event logic [3]. In the theory, the sentence

(10) Brutus stabbed Caesar in the back with a knife

is interpreted as having the following logical form:

(11) $(\exists e)[\text{Stabbing}(e)\&\text{Subj}(e,B)\&\text{Obj}(e,C)\&\text{In}(e,b)\&\text{With}(e,k)]$

The logical form says in effect that there is an event of stabbing whose subject is Brutus and whose object is Caesar and whose location is in the back and whose instrument is a knife. Notice that each of $\text{Stabbing}(e)$, $\text{Subj}(e,B)$, $\text{Obj}(e,C)$, $\text{In}(e,b)$, $\text{With}(e,k)$ can be thought of as a proposition and together they constitute the proposition corresponding to the sentence (10). To the extent that (10) represents an atomic proposition, these propositions are subatomic.

Although Parson's subatomic semantics is concerned with events, and is, in that respect, committed to ontology of subatomic entities, our theory is more representational, in that we do not commit ourselves to the exact nature of the subatomic propositions, whether they are actual entities out there or they are just expedient mechanism to allow proper inference and representation. The exact nature of subatomic entities is not especially crucial to our current concern.

The concept of subatomic proposition may be considered related to the traditional grammatical notion of *bunsetu*, though the latter is specific to Japanese while the former is universal. According to Shinkichi Hashimoto [16] a *bunsetsu* is "the shortest unit in a sentence when the sentence as language in actual use is divided into as many parts as possible." Although Hashimoto does not mention anything similar to the concept of speech act, his understanding of *bunsetsu* is intuitively similar to ours. Note, however, *bunsetsu* is part of an actually uttered sentence, whereas subatomic proposition is what is conveyed by such a unit. Sometimes *bunsetsu* is regarded as a phonological unit, as an accent phrase [10]. In a sense one might understand a *bunsetsu* as a means in Japanese to represent a subatomic proposition.

Now let us consider applying the theory above to Japanese dialogue and speech acts. For example, consider the following sentence:

(12) *Tokyo-ni   itte kudasai*
　　 Tokyo-TO go　please

'Please go to Tokyo'

In a Searlean manner, the propositional part of the sentence would be:

(13) GO_TO(you,TOKYO).

But regarding this sentence as consisting of its subatomic parts, the whole proposition could be expressed as:

(14) $(\exists e)[\text{TO}(\text{TOKYO},e) \wedge \text{GO}(\textbf{you},e)]$

where TO(TOKYO) and GO(**you**,e) are subatomic propositions composing the whole proposition.

It would be feasible to assign this logical form to a sentence that means corresponds to a proposition such as 'You went to Tokyo' or 'You are going to Tokyo', because (14) states that there is actually an event in which 'you' go to Tokyo. But since (12) does not describe an actual state of affairs but a possible state of affairs in the future, (14) is not quite appropriate as the logical form for (12).

We propose, instead, that the entity in question is the event-type:

(15) $\{e|\text{TO}(\text{TOKYO},e) \wedge \text{GO}(\textbf{you},e)\}$

where any event of you going to tokyo counts as describing the state of affairs in question.

Furthermore, in the Searlean analysis of speech act, the illocutionary act represented by the utterance would be:

(16) $\text{REQUEST}(\text{GO\_TO}(\textbf{you},\text{TOKYO}))$

while in the subatomic framework, it would become something like:

(17) $\text{REQUEST}(\{e|\text{TO}(\text{TOKYO},e) \wedge \text{GO}(\textbf{you},e)\})$

Notice that even though the illocutionary force REQUEST has a scope over the whole proposition, the subatomic analysis makes it possible to extract a subatomic proposition (type) of the form, e.g.:

(18) $\{e|\text{TO}(\text{TOKYO},e)\}$

in which the concept of going to Tokyo is simply represented.

With this recourse at hand, we can analyze a series of utterances such as the following:

(19) *a, anoo, Tokyo ni desune, eeto, itte kuremasuka?*
　　　　　　　　to Tokyo　　　　　　　　go　would you please

　　'Well, would you please go to Tokyo?'

Observing that *desune* and *ne* can be attached to forms corresponding to subatomic propositions, one can presume that Japanese 'sentences' may be taken as semantically denoting subatomic propositions, and not full propositions. On the other hand, it might be argued that in English and other European languages employ 'sentences' corresponding more closely to full propositions.

3　DOMAIN AND ACTION IN DISCOURSE.

We have shown that Japanese dialogue can be understood to be based on subatomic propositions, and hence that the basic unit of analysis for Japanese dialogue may be utterances corresponding to subatomic propositions rather than full-fledged sentences corresponding to propositions since the latter is often not obvious in actual Japanese dialogue.

We will see in this section how these subatomic propositions are related to illocutionary acts, and how speech acts in dialogue are to be construed.

So let us consider applying the theory presented in the last section to the dialogue (8), which is taken from a real conversation.

The speaker A's utterance in (8) is made up from four turns, each of which is, in the traditional grammatical sense, not quite well-formed. In our analysis, these turns are understood as representing subatomic propositions.

The first turn, which is repeated here as (20),

(20) A1: *eeto, boku no tukue no ue ni desune*
　　　　　well　my desk　　　　upon

　　　'Well, on my desk'

can be interpreted as a speech event in which something, whose identity is unknown yet, on the desk is mentioned. Notice that unlike in the case of *Tokyo-ni*, this cannot be rendered as $\{e|\text{ON}(e,my\_desk)\}$, since that something in question that is on the desk is obviously not an event, but rather some object still unspecified. So it is a type of object that is, possibly, on the desk. In other word this speech event is describing an event in which unspecified entity on the desk is mentioned. Let us represent this described proposition, albeit somewhat cumbersomely, as:

(21) $\{e|\{x|\{\text{ON}(x,my\_desk,e)\}\}$.

This in effect says that the utterance in question refers to a subatomic proposition, which is an event type, in which some type of object on my desk is described.

The next turn, which is repeated here:

(22) A2: *hon   aru deshou   yon satu*
         book   there is, isn't there ? four

    'There are four books, aren't there?'

in effect says that the object in question was in fact a book, and there are four of them. If we represent the subatomic proposition as: $\{e|book(b,e)\}$, it would imply that any one argument predicate becomes relational in our analysis. This is not a very satisfactory situation. So instead of writing every event type involving a predicate P as $\{e|P(z,e)\}$, let us write $e \models P(z)$, which can be informally read '$P(z)$ in e'. In this notation the representation for the first turn would now become, not (21), but rather:

(23) $e_1 \models \{x|on(x,my\_desk)\}$

and that for (the first part of) the second turn would become:

(24) $e_2 \models \{x|book(x)\}$

and the latter part of the second turn would be:

(25) $e_2 \models \{z|card(z,4)\}$.

Strictly speaking, there is another element to be noticed. It is the verb *aru* 'be' in the second turn. We can denote the fact that this verb is used in this turn as an independent subatomic proposition of the form:

(26) $e_2 \models \mathsf{Action}(be)$,

though whether 'be' is an action is an issue too problematic for us to go into here.

Notice, however, that the 'illocutionary force' of these subatomic propositions is not really that of assertion or claim, but rather that of confimation, as can be seen from the phrase *phrase* 'isn't there?' The whole point of these turns is to ascertain that the hearer recognizes that there is indeed such an entity, and to construct the mutual understanding that the speaker is talking about that entity. For the sake of clarity, let us write this in the following manner:

(27) $e_2 \models \mathsf{Force}(\textsc{confirm})$

But this notation is misleading since we do not want to claim that this is also part of subatomic propositions constituting the whole proposition; rather the force is carried by the proposition. But we do not go into this either.

In a similar manner, the third turn:

(28) A3: *sore-o   desune*
         that-OBJ

    'Those books'

would be rendered as:

(29) $e_3 \models \mathsf{Obj}(\mathbf{dem})$,

where **dem** is the all-purpose demonstrative, here standing for *sore*, whose its referent or antecedent is to be resolved from the discourse context.

The fourth turn is also rendered as before:

(30) A4: *tyotto   motte kitekuremasuka?*
         a little   bring   can you please?

   'Can you bring them to me?'

where the subatomic proposition is:

(31) $e_4 \models \mathsf{Action}(bring)$

while the illocutionary force is:

(32) $e_4 \models$ Force(REQUEST).

One can notice that this series of turns can be broken into two parts, with the two verbs, *aru* and *motte-kite*, at each end.

We can build a discourse structure representation for that portion of dialogue, converting our subatomic propositions into conditions in Discourse Representation Theory (DRT). The result would be as in the following:
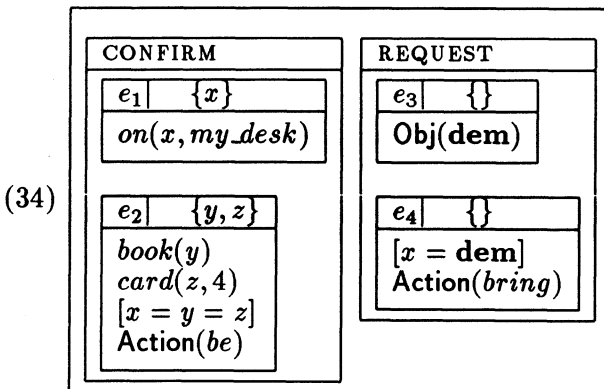
(33)

e_1	$\{x\}$
$on(x, my\_desk)$	

e_2	$\{y, z\}$
$book(y)$	
$card(z, 4)$	
$[x = y = z]$	
Action(be)	
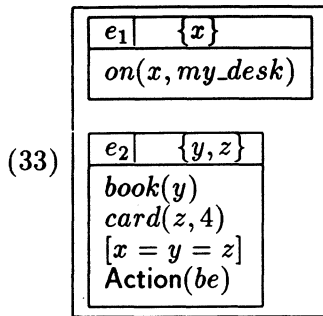
The second part of the dialogue, with the third and fourth turns, can also be converted into a discorse structure representation using DRT notation.

e_3	$\{\}$
Obj(dem)	

e_4	$\{\}$
$[x = \mathbf{dem}]$	
Action($bring$)	
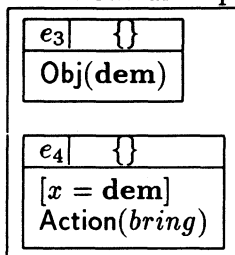
In a way, one can say that both of these two portions are 'governed' by the two verbs that are related to actions; each of the subatomic propositions constituting the turns contributes to and culminates in the verb at each end. These actions can be taken to represent a sort of unit that somehow corresponds to a speech act: the first corresponding to confirmin and the second to requesting.

All told, the utterance of the speaker A would constitute the following speech act:

(34)

CONFIRM

e_1	$\{x\}$
$on(x, my\_desk)$	

e_2	$\{y, z\}$
$book(y)$	
$card(z, 4)$	
$[x = y = z]$	
Action(be)	

REQUEST

e_3	$\{\}$
Obj(dem)	

e_4	$\{\}$
$[x = \mathbf{dem}]$	
Action($bring$)	

Thus we have indicated that our theory of subatomic proposition is capable of capturing the traditionally irregular aspects of spoken Japanese dialogue and manages to construct a discourse strucutre that is amenable to formal treatment.

4   CONCLUSION

We have discussed the unit of utterance in spoken Japanese and its relation to dialogue exchange. We have shown that utterances in ordinary spoken Japanese dialogue are not 'well-formed' in the traditional grammatical sense but is governed by different rules. We have proposed one such is that spoken dialogue is often composed of subatomic propositions. We also showed how these subatomic propositions are related to discourse structures and utterance exchanges.

REERENCES

[1] J. Allwood, J. Nivre, and E. Ahlsén. "On the semantics and pragmatics of linguistic feedback." Technical report, Gothenburg Papaers in Theoretical Linguistics, 64., 1991.

[2] J. J. Chew. *A transformational analysis of modern colloquial Japanese*. Mouton, The Hague, 1973.

[3] D. Davidson. *Actions and Events*. Oxford University Press, Oxford, 1980.

[4] K. Dohsaka and A. Shimazu. "A computational model of incremental utterance production in task-oriented dialogues." In *Proceedings of the 16th International Conference on Computational Linguistics*, 1996.

[5] M. Iwata. *Noo-to Kotoba(Brain and Language)*. Kyouritsu Shuppan, Tokyo, 1996.

[6] H. Kamp. "A theory of truth and semantic representation." In J. Groenendijk, T. Janssen, and M. Stokhof, editors, *Truth, Interpretation and Information*, pp. 1–41. Foris, Dordrecht, 1981.

[7] M. Kawamori, T. Kawabata, and A. Shimazu. "Discourse markers in spontaneous dialogue: A corpus based study of Japanese and English." In M. Stede, L. Wanner, and E. Hovy, editors, *the Proceedings of Coling/ACL 98 Workshop on Discourse Relations and Discourse Markers*, pp. 93–99, 1998.

[8] M. Kawamori and A. Shimazu. "The notion of control in discourse." *TR-IEICE*, pp. 31–36, 1996.

[9] M. Kawamori, A. Shimazu, and K. Kogure. "Roles of interjectory responses in spoken discourse." In *the Proceedings of the 3rd International Conference on Spoken Language Processing*, pp. 955–958, 1994.

[10] S. Martin. *A Reference Grammar of Japanese*. Yale University Press, New Haven, 1975.

[11] T. Parsons. *Events in the Semantics of English: a Study in Subatomic Semantics*. MIT Press, Cambridge, Mass., 1990.

[12] J. Searle. *Speech Acts*. Cambridge University Press, Cambridge, 1969.

[13] A. Shimazu, K. Kogure, and M. Nakano. "Cooperative distributed processing for understanding dialogue" utterances. In *the Proceedings of the 3rd International Conference on Spoken Language Processing*, pp. 99–102, 1994.

[14] H. Suzuki and S. Tutiya. "A strictly incremental approach to Japanese grammar." In J. Barwise, M. Gawron, G. Plotkin, and S. Tutiya, editors, *Situation Theory and Its Applications*, volume 2, pp. 517–532. CSLI, Palo Alto, 1991.

[15] S. Tutiya, K. Shirai, H. Suzuki, and M. Kawamori. "In search of semantics of Japanese." *Gengo*, 1–12, 1990.

[16] M. Watanabe. *Kokugo Bunpouron (Theory of Japanese Grammar)*. Kasama, Tokyo, 1974.

[17] S. Whittaker and P. Stenton. "Cues and control in expert client dialogues." In *the Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pp. 123–130, 1988.