Structural Analysis of Instruction Utterances

Tomohide Shibata, Daisuke Kawahara, Masashi Okamoto,
Sadao Kurohashi, and Toyoaki Nishida

Graduate School of Information Science and Technology, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{shibata, kawahara, okamoto, kuro, nishida}@kc.t.u-tokyo.ac.jp

Abstract. Toward designing a system which teaches various works interactively and visually, this paper proposes a method of analyzing instruction utterances. One of the biggest problem in dealing with spoken language is ellipsis/anaphor resolution. We resolve it using a domain-specific case frame dictionary constructed automatically from a large amount of texts. Then, we attach utterance-type to distinguish actions from notes, tips, etc. Based on the attached type, we analyze discourse structure of utterances and detect a unit of actions.

1 Introduction

With the advance of computers, networks, and media processing techniques, it has been possible for computers to help human with intelligent works. For example, "Dialog Navigator" is a system which can answer questions interactively based on large text knowledge-base, such as support technical information of Microsoft [1].

As an extension of such text-based QA systems, we can design a system that teaches methods, notes, and tips interactively and visually about various works. Such a system is so effective in many fields, such as handiwork, cooking, sports, and so on, where ideally, we want to ask experts/teachers and to follow their examples.

To realize such a system, it is necessary to collect, analyze, and structure video contents given by experts/teachers. As the first step toward such an intelligent video archiving, we need to analyze instruction utterances of experts/teachers. The resultant linguistic information can guide to correspond visual information with linguistic information, summarize video contents, and find related information in other videos or in the web.

In this paper, we focus on Japanese cooking instruction, and describe a method of analyzing structure of cooking instruction utterances. We do not deal with speech recognition but start with transcriptions and closed captions of a TV cooking program as shown in Figure 1.

2 Structure of Cooking Instruction Utterances

In cooking instruction utterances, while explanations of actions are the core, there are a declaration of beginning of series of actions, tips of actions, notes,

Hello.
Welcome to "Today's Cooking".
Today, I introduce a dish suitable for summer.
...
Next, cucumber.
I will put this in a soup.
Peel it.
Peel it with this peeler.
As summer has come, cucumber appear on the market, and I think it is delicious in a soup.
Cut it lengthwise.

Fig. 1. An example of cooking instruction utterances (NHK "Today's Cooking").

etc. among them. We classify cooking instruction utterance into the following 7 types (by referring to Izuno [2]):

Action declaration:	e.g. Then, we cook a steak.
Individual action:	e.g. Pour water into a pan.
Food state:	e.g. There is no water in a carrot.
Food/Tool presentation:	e.g. Ingredients are 150g minced beef.
Substitution:	e.g. You may squeeze it by hand.
Note:	e.g. Be careful so that seeds won't go into.
Miscellaneous:	e.g. Hello.

Action sequence in cooking is regarded as a hierarchical tree structure. For example, cooking of "fried pork" consists of seasoning, preparing sauce, deep frying, etc., and preparing sauce consists of slicing a ginger and a leek, putting in spice, etc. In such a tree structure, utterance corresponding to an inner node is a declaration of beginning of series of actions under the node.

3 Basic Analysis of Cooking Instruction Utterances

3.1 Pre-Processing Cooking Instruction Utterances

The outline of processing cooking instruction utterances is shown in Figure 2. First, we make morphological analysis using JUMAN, and syntactic analysis using KNP [3]. JUMAN, the Japanese morphological analyzer, and KNP, the Japanese parser, have been developed for written language. However, they are designed robustly enough to be able to analyze spoken language in quotation of newspaper. We found they can deal with utterances just by eliminating interjections and sentence-final particles.

3.2 Constructing Case Frame Dictionary in Cooking Domain

In Japanese, zero pronouns are often used, especially, in spoken language. Ellipsis resolution requires a case frame dictionary.

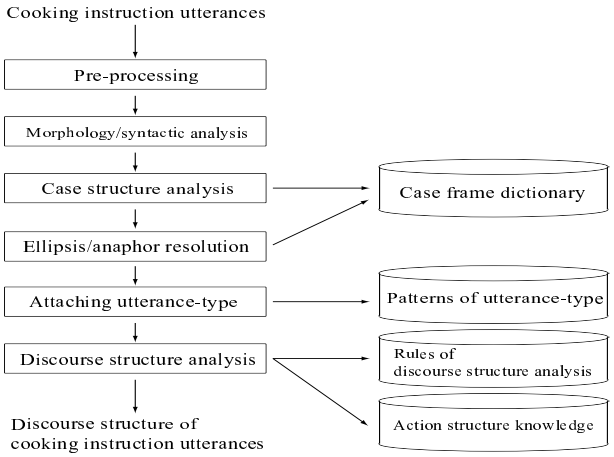

Fig. 2. The outline of processing cooking instruction utterances.

Table 1. An example of Case frame.

	Case marker	Examples
cut (1)	*ga*	<agent>
	wo	pork, cucumber, ···
	ni	rectangle, diamonds, ···
cut (2)	*ga*	<agent>
	wo	card, check ···

We have constructed a case frame dictionary from newspaper articles of 20 years. To deal with cooking instruction utterances, we gathered texts on cooking as follows:

- We collected words that co-occur with a word "cooking" by the Web and newspapers, such as egg, peel, knife, meat, etc.
- From the Web, we obtain texts including two words: "cooking" and above each word.

The case frame dictionary is automatically constructed by the following procedure (Table 1 shows an example) [4]:

1. A large raw corpus is parsed by KNP, and reliable predicate-argument examples are extracted from the parse results.
2. The extracted examples are bundled according to the verb and its closest case component.
3. The case frames are clustered using a similarity measure, resulting in the final case frames. The similarity is calculated by using a Japanese thesaurus [5], and its maximum score is 1.0.

3.3 Ellipsis/Anaphor Resolution

We resolve ellipsis/anaphor by using the case frame dictionary. We built a zero pronoun resolution system which utilizes the case frame dictionary and the dis-

tance tendency that a zero pronoun has its antecedent in its close position. We examine candidate antecedents in increasing order of distance from a zero pronoun. To measure the distance, we introduce *location classes*, which capture locational relations between zero pronouns and antecedents structurally. The location classes are established and ordered by investigating zero pronouns in an annotated corpus on a large scale. The algorithm is as follows:

1. Parse an input sentence using KNP, and process each predicate in it by the following steps.
2. Select a case frame corresponding to the predicate and its closest case component.
3. Match each input case component to an appropriate case slot of the selected case frame.
4. Regard case slots that have no correspondence as zero pronouns. Each zero pronoun is analyzed by the next step.
5. Estimate an antecedent of a zero pronoun. Candidate antecedents are examined in the location class order. If a candidate is classified as positive by a binary classifier and its similarity exceeds a threshold, this candidate is selected as the antecedent. We employ Support Vector Machines as the classifier.

4 Organizing Cooking Instruction Utterances

4.1 Attaching Utterance-Type

As mentioned in the section 2, cooking instruction utterance is classified into action declaration, individual action, food state, food/tool presentation, substitution, notes, and miscellaneous.

Among them, action declaration, food/tool presentation, substitution, notes, and miscellaneous can be recognized by patterns of sentence-end. Concretely, we prepared the patterns of morphology sequence, and about each morphology, we can check its basic form, part-of-speech, conjugational form, and semantic features on thesaurus. As for individual action and food state, it is possible to enumerate all the predicate in the cooking domain. However, considering the portability of the system, we use general rules regarding intransitive verbs or adjective + *"naru"*(become) as food state, and others as individual action.

Examples of patterns for attaching utterance-type are shown in Table 2.

4.2 Discourse Structure Analysis of Utterances

Next, based on types attached in the previous section, we analyze discourse structure of utterances. As a model of the discourse structure, we suppose a graph structure that each utterance is one node and is linked with related utterances. In the discourse structure of task-oriented utterance like cooking, a task reflecting the tree structure is the core, and such utterance as substitution and notes modify it.

We have to detect a unit of actions from among action declarations and individual action. However, it is difficult to analyze without knowledge about

Table 2. Examples of patterns for attaching utterance-type.

Pattern	Example
action declaration	
∼*ni-kakarimasu*(begin to ∼)	Then, I begin to cook a steak.
∼*te-ikimasu*(be going to ∼)	And then, I will add spice.
food state	
intransitive verb	Water boiled.
adjective + *naru*(become)	Oil heated up sufficiently.
food/tool presentation	
<food/tool> *wo-tsukaimasu*(use)	I use this handy mixer.
substitution	
∼*shitemo-kekkoudesu*(may)	You may use lemon juice.
∼*demo-kekkoudesu*(may)	You may use sliced meat.
note	
∼*wasurenaide-kudasai*(forget)	Don't forget to add salt first.

```
<ActionStructure label="cut" type="same ingredient" common="vegetable">
    <Action id="1" type="opt">peel</Action>
    <Action id="2">cut lengthwise</Action>
    <OR>
      <Action id="3">cut into small pieces</Action>
      <Action id="3">cut into diamonds</Action>
      <Action id="3">cut into rectangles</Action>
    </OR>
</ActionStructure>
```

Fig. 3. An example of action structure knowledge in cooking.

action structure of cooking. Then, we constructed action structure knowledge by hand. An example of action structure knowledge is shown in Figure 3. It is described in XML format and <Action>...</Action> expresses an individual action. It has an ability almost equal to regular expression as follows:

- **type="opt"** matches zero or one time. (optional action)
- **type="repeat"** matches one or more times.
- **<OR>...</OR>** matches one of the alternatives.

We analyze the discourse structure of utterances combining action structure knowledge with general rules of discourse structure analysis.

A method of the discource structure ananlysis is based on [6]. As a new sentence comes in, by checking surface information, we find a connected sentence and the coherence relation between them. Examples of rules for discourse structure analysis are shown in Table 3. Each rule specifies a condition for a pair of a new sentence and a possible connected sentence: the range of possible connected sentences (how far from the new sentence) and patterns for the two sentences. If a pair meets these condition, the relation and score in the rule are given to it. As a final result, we choose the connected sentence and the relation that have the maximum score.

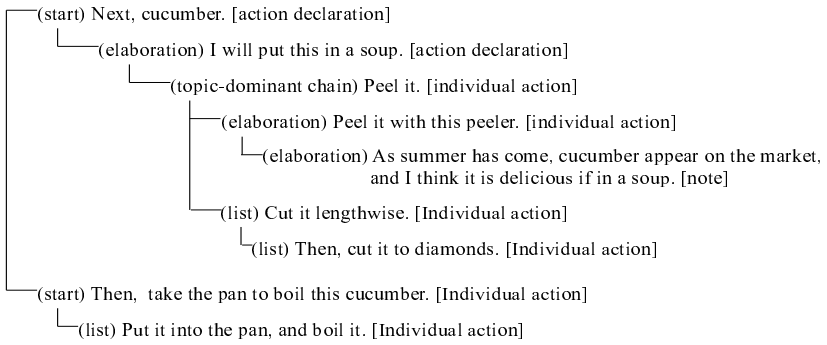

Fig. 4. An example of discourse structure.

Table 3. Examples of rules of discourse structure analysis.

Coherence relation	Score	Applicable range	Patterns for a connected sentence	Patterns for a new sentence
list	5	1	*	*soshite*(then)$\cdots$
contrast	30	1	*	*mushiro*(rather than)$\cdots$
contrast	40	*	X$\cdots$	X'($\simeq$ X)$\cdots$
elaboration	15	1	*	<note>
reason	30	1	*	$\cdots$karada(because)

While checking such general rules, we check sequence action matches action structure knowledge. If matched, the structure is applied.

If we make analysis of utterances in Figure 1, we get discourse structure in Figure 4.

5 Experiment and Discussion

We used the corpus of NHK TV program, "Today's Cooking". A program consists of about 200 utterances, and the average length of an utterance is about 20 characters.

We evaluated the precision of ellipsis/anaphor resolution and attaching an utterance-type on each 60 sentences of three cooking programs. The result is shown in Table 4. The threshold of the similarity in ellipsis/anaphor resolution (in the section 3.3) was set to 0.60 based on preliminary experiment.

As for ellipsis/anaphor resolution, the accuracy of "*ga*" is especially low. The case "*ga*" of transitive verbs are often omitted, and the antecedent is mostly the speaker. The current system prefers to search it in sentences, and this causes many mistakes. As for utterance-types, most of them are correctly recognized by checking surface expressions. It is important to distinguish action declaration/individual action from others. As for discourse structure analysis, supposing action structure knowledge, the accuracy is 67%. We constructed action structure knowledge by hand, but we need to learn it automatically in the future.

Table 4. Result of ellipsis/anaphor resolution and attaching utterance-type.

Ellipsis/anaphor resolution	
ga	25/68(37%)
wo	40/62(65%)
ni	43/67(65%)
Total	108/197(55%)

Attaching utterance-type	
action	101/106(95%)
food state	1/2(50%)
food/tool presentation	15/17(88%)
substitution	5/5(100%)
note	19/21(90%)
others	47/65(72%)
Total	188/216(87%)

6 Conclusion

As the first step of designing a system which teaches various works interactively and visually, we analyzed the structure of cooking instruction utterance. We resolve the problem of ellipsis and anaphor in spoken language with knowledge acquired by automatic learning. Then, we recognize utterance unrelated with action, and analyze discourse structure of utterance using action structure knowledge from the top down.

We are planning to use the resultant structured linguistic information to correspond visual information with linguistic information and to summarize video contents.

References

1. Youji Kiyota, Sadao Kurohashi, and Fuyuko Kido. Dialog Navigator: A Question Answering System based on Large Text Knowledge Base, In *Proceedings of 19th COLING (COLING02)*, pp.460-466, 2002.
2. Hidekatsu IZUNO, Yuichi NAKAMURA, and Yuichi OHTA. QUEVICO: A Framework for Video-based Interactive Media. In *Working Notes WS-5 International Workshop on Intelligent Media Technology for Communicative Reality, PRICAI-02 (Seventh Pacific Rim International Conference on Artificial Intelligence)*, pp. 6-11, August, 2002.
3. Sadao Kurohashi and Makoto Nagao. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. *Computational Linguistics*, 20(4), pp.507-534.
4. Daisuke Kawahara and Sadao Kurohashi: Fertilization of Case Frame Dictionary for Robust Japanese Case Analysis, In *Proceedings of 19th COLING (COLING02)*, pp.425-431, 2002.
5. Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentarou Ogura, Yoshifumi Oyama, and Yoshihiko Hayashi, editors. 1997. Japanese Lexicon. Iwanami Publishing.
6. Sadao Kurohashi and Makoto Nagao. Automatic Detection of Discourse Structure by Checking Surface Information in Sentences, In *Proceedings of 15th COLING*, Vol.2, pp. 1123-1127, 1994.