

Incremental Spoken Language Translation Utilizing Grammatically Ill-formed Expressions

Shigeki Matsubara Satoru Asai Yasuyoshi Inagaki

Department of Computational Science and Engineering

Nagoya University

Furo-cho, Chikusa-ku, Nagoya 464-01 Japan

{matu,asai,inagaki}@inagaki.nuie.nagoya-u.ac.jp

Katsuhiko Toyama

Department of Computer Science

Chukyo University

101 Tokodate, Kaizu-cho, Toyota 470-03 Japan

toyama@sccs.chukyo-u.ac.jp

Abstract

This paper proposes a system for incremental English-Japanese spoken language translation, the system behaving like simultaneous interpretation. Since spontaneous speech appears continuously, each stage of a spoken language translation system should work incrementally. However, machine translation systems which have been proposed so far can not achieve high degree of incrementality because of the difference in word-order between the source language and the target one. In this paper, the system utilizing grammatically ill-formed expressions characterizing Japanese spoken language such as repetitions, inversions, ellipses, repairs and hesitations, we have succeeded in the incremental transfer from English to Japanese. An experimental system on the basis of incremental chart parsing framework has been implemented. To evaluate the system we have made an experiment with 278 English dialogue sentences. 228 of them are translated correctly, providing a success rate of 82.0%, and 47msec on the average was taken to process one word. These results show our technique to be useful for spoken language translation with acceptable accuracy and high real-time nature.

1 Introduction

Spontaneously spoken language, which is different from written language, has the feature that the appearance process is continuous. In spoken language translation, therefore, highly real-time processing is strongly required. However, most machine translation systems which have been proposed so far work on a sentence-by-sentence basis. This means that their systems cannot start transferring it until the utterance of the whole sentence finishes, so that should lose real-time nature of processing.

In order to develop advanced speech-to-speech translation systems, a technique for translating spoken language incrementally and synchronously is required [Menzel, 94]. As such frameworks, Kitano has

proposed a method of incremental sentence production for modeling simultaneous English-Japanese interpretation [Kitano, 90]. As well, Amtrup has introduced a method of chart-based incremental transfer for head switching in German-English machine translation [Amtrup, 95]. We, however, should point out that their methods do not achieve high degree of incrementality and synchronicity because of the difference in word-order between the source language and the target one. For example, let us consider to incrementally translate the following English simple sentence into Japanese, the word-order of which is different from that of English.

(1.1) Ken met her in the park near the school yesterday.

The standard translation of (1.1) is considered to be

the following Japanese sentence:

(1.2) *ken-wa* (Ken) *kinoo* (yesterday) *gakko-no*
soba-no (near the school) *koen-de* (in the park)
kanojo-ni (her) *atta* (met).

Noted that it is impossible to utter “*kinoo*” appearing secondly in (1.2) until “yesterday” appearing lastly in (1.1) is heard, that is to say, (1.2) can not be outputted except “*ken-wa*” until the input of (1.1) finishes. This example illustrates that the difference in word-order between the source language and the target one might cause the system to lose real-time nature. In fact, both of Kitano and Amtrup adopt just such the translation method as they have the problem discussed above. In order to overcome this difficulty, it is necessary to develop a technique for incremental translation whose processing synchronizes with the input [Inagaki and Matsubara, 95].

This paper proposes a system for incremental English-Japanese spoken language translation, the system translating English input words into Japanese according to the order of appearances of them. The system utilizes grammatically ill-formed expressions characterizing Japanese spoken language, that is to say, produces Japanese translations including the expressions such as repetitions, inversions, ellipses, errors, repairs and hesitations. The frequent utilization of these expressions is a key to success of the exceedingly incremental translation of English into Japanese. There have been many discussions about the way of robustly analyzing grammatically ill-formed expressions peculiar to spoken language so far. In contrast, since they are used frequently in human daily conversations, it seems to be quite all right for spoken language translation systems to utter grammatically ill-formed sentences.

An experimental system which translates English dialogues into Japanese on a possibly word-by-word basis has been implemented on the basis of a chart processing framework. The system is mainly composed of two modules: incremental parsing and transfer, which work synchronously with each other. Working in an exceedingly incremental fashion, the system can be expected to be available for advanced real-time machine translation with simultaneous interpretation.

To evaluate the effectiveness of the system, we have made an experiment with 278 English spoken sentences in ATR Dialogue Database [Ehara et al., 90]. As a result, 228 of the sentences are translated correctly, providing a success rate of 82.0%. Moreover, we have confirmed that the system outputs the corresponding Japanese speech as soon as a spoken English word is inputted. These show the system to be useful for spoken language translation with acceptable accuracy and high real-time nature.

This paper is organized as follows: Section 2 explains the idea of using ill-formedness for the incremental translation. Section 3 describes the overview of the incremental spoken language translation system. Section 4 reports on the experimental results. Section 5 discusses that the system produces grammatically ill-formed sentences.

2 Utilizing Grammatically Ill-formed Expressions

The expressions such as repetitions, inversions, ellipses, repairs and hesitations are grammatically ill-formed but natural in Japanese daily conversation. Japanese people can usually understand the speech including them. For the purpose of translating simultaneously, the system which this paper proposes utilizes these expressions in an effective way. This section illustrates the way of utilizing them.

2.1 Ill-formed Expressions: How to Utilize

2.1.1 Repetitions

In general, repetitions are used for adding the more detailed information. For example, the system translates an English sentence (2.1) into the Japanese (2.2).

(2.1) I live in a university dormitory.

(2.2) *watashi* (I) *wa* *sunde-imasu* (live). *daigaku-no* (university) *ryoo-ni* (in a dormitory) *sunde-imasu* (live).

As soon as an English verb “live” is inputted, the system outputs the corresponding Japanese verb “*sunde-imasu*”. Then, since a verb usually appears at the end of the sentence in Japanese, here ends the first translation once. The system reflects the expressions which follow the source verb into the second sentence of the translation, and additionally, outputs the target verb “*sunde-imasu*” last again.

The standard translation of (2.1) is considered to be the following Japanese sentences:

(2.3) *watashi-wa* (I) *daigaku-no* (university) *ryoo-ni* (in a dormitory) *sunde-imasu* (live).

Here, we emphasize, as well as (2.3), that (2.2) represents the semantic content of (2.1) exactly and most Japanese people can comprehend it correctly.

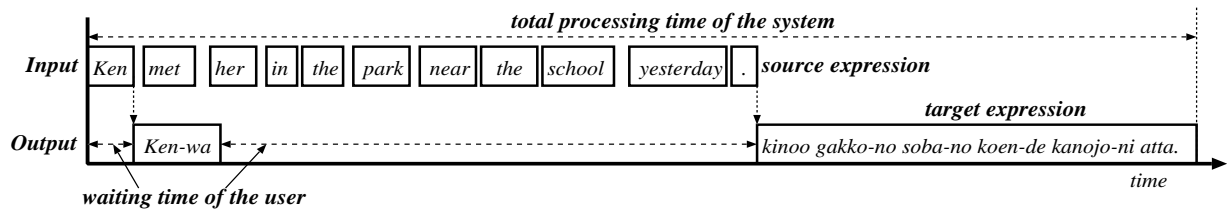


Figure 1: Timing of the output of (1.2)

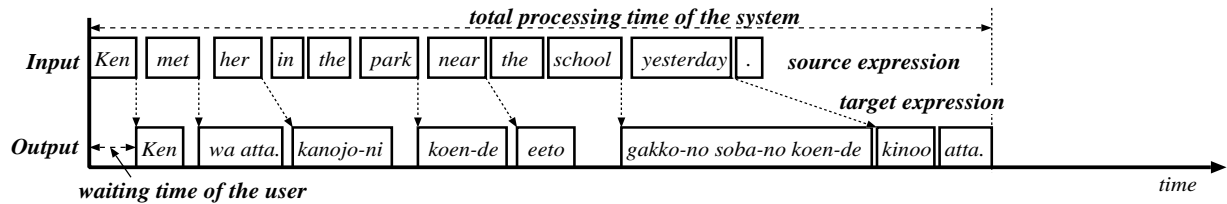


Figure 2: Timing of the output of (2.10)

2.1.2 Inversions

The word-order of Japanese is very different from that of English. However, some else to be borne in mind here is that the word-order of Japanese is considerably flexible. This means that, even if the system produces the corresponding Japanese speech according to the order of appearances of English words, the user could understand it easily. For example, although the conventional Japanese translation of (2.4) is “*kono tsuaa-wa ikura desu-ka*”, most Japanese people can understand (2.5) which the system outputs.

(2.4) How much is this tour?

(2.5) *ikura* (how much) *desu-ka* (is). *kono* (this) *tsuaa-wa* (tour).

2.1.3 Ellipses

The various words are frequently omitted in Japanese. The system also makes up more natural translations by omitting various target words. Note that in (2.2) the locative case of the first sentence and the nominative case of the second one are omitted.

2.1.4 Repairs

Many input sentences include some structural/lexical ambiguities. In incremental language processing, particularly, the ambiguities increase because of incompleteness of the input information. The system, however, is forced to select the most plausible one for the purpose of achieving high incrementality. This means that the translations may include errors which could be found later. Finding the errors, the system makes

up a correct translation by repairing them. For example, “these” in (2.6) can usually take two parts of speech: a *demonstrative pronoun* and a *demonstrative adjective*. Even if the system outputs a pronoun “*korera-wa*”, it can repair the translation by uttering an adjective “*korera-no*” later.

(2.6) And are these tours expensive?

(2.7) *sore-to* (and) *korera-wa* (these) *tsuaa* (tours), *korera-no* (these) *tsuaa-wa* (tours) *takai-desu-ka* (are expensive?).

2.1.5 Hesitations

The hesitations have a tendency of appearing before the repair expressions [Matsubara and Inagaki, 96]. In the translation of (2.8), the system outputs the hesitation “*eeto*” before outputting the repair expression “*nijuu-yokka-ni syuppatsu*”.

(2.8) Leaving on the 24th and leaving on the 31st?

(2.9) *nijuu-yokka-ni syuppatsu-wa* (leaving on the 24th) *eeto*, *nijuu-yokka-ni syuppatsu* (leaving on the 24th) *sore-to* (and) *sanjuu-ichinichi-ni syuppatsu?* (leaving on the 31st?).

The system can produce more natural utterances by utilizing hesitations effectively.

2.2 Typical Example

According to the idea illustrated in the previous subsection, the system translates (1.1) as follows:

(2.10) *ken* (Ken) *wa* *atta* (met). *kanojo-ni* (her) *koen-de* (in the park), *eeto*, *gakko-no soba-no koen-de* (in the park near the school) *kinoo* (yesterday) *atta* (met).

Figure 1 and 2 show the timing of the output of (1.2) and (2.10) respectively. Comparing Figure 2 with 1, it is obvious that the incremental system can output the target language synchronously with the input, and therefore both the waiting time of the user and the total translation time are getting shorter.

Comparing (2.10) with (1.2), what is more, the characteristic features of (2.10) are the following points:

- The verb “*atta*” in the first sentence is repeated at the last in the second one.
- The locative case “*koen-de*” and the objective case “*kanojo-ni*” are inversed in word-order.
- The objective case in the first sentence is omitted and the nominative case in the second sentence is also done.
- The error “*koen-de*” is corrected by speaking “*gakko-no soba-no koen-de*” afterward.
- The hesitation “*eeto*” is inserted.

The Japanese sentence (2.10) includes the grammatically ill-formed expressions: repetition, inversion, ellipses, error, repair and hesitation. However, we would like to emphasize that (2.10) does not only represent the semantic content of (1.1) correctly, but can also be understood easily by most Japanese people. Though we showed a translation example of a simple declarative sentence here, the system can cope with various type of English sentences by utilizing various ill-formed expressions effectively.

3 Overview of the System

To demonstrate the feasibility and the usefulness of the incremental English-Japanese spoken language translation system, we have developed an experimental system in GNU Common Lisp 2.2. For the purpose of the translation experiment with ATR Dialogue Database [Ehara et al., 90], the system has been implemented in the scale of English lexicon 476 words and 204 grammar rules, and also the transfer rule has been established corresponding to each grammar rule.

Figure 3 shows the configuration of the system. The system is composed of eight components: speech recognition, speech synthesis, incremental parsing, incremental transfer, parsing rules, transfer rules, a dictionary and a chart. To represent incomplete struc-

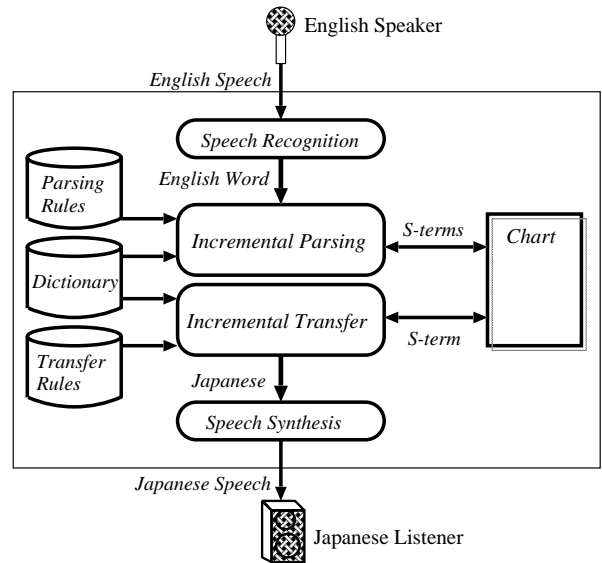


Figure 3: Configuration of the system

tures gained by incremental syntactic analysis, we have adopted a chart-based framework in both the parsing module and the transfer one. The chart component is the data structure which represents possible phrase structures of the source language on the way of the input. The system translates English sentences incrementally and synchronously by executing the below (1)-(2) sequentially for each word.

- (1) **Incremental parsing** makes edges with the terms, whose syntactic categories are sentences, as the labels. From now on, we call such the term, simply, *s-term*. The parsing procedure follows the bottom-up chart processing framework. The altering point for the orthodox bottom-up chart parsing method [Kay, 80] is that for embodying the *s-term* gradually, the syntactic operation of replacing an undecided term in the *s-term* with the other active one has been defined. Therefore, the parsing module behaves in the same way as combinatory categorial grammar [Steedman, 87] which is well-known as a framework of incremental parsing. If no *s-term* can be gained newly, the system fails in the translation.
- (2) **Incremental transfer** transforms the sentence structure which incremental chart parser has made, to the target expressions by applying transfer rules in a top-down fashion. How the system utilizes grammatically ill-formed expressions is described as the transfer rules. In the spoken language translation, the source expressions transferred once does not have to be transferred any longer. Therefore, the incremental transfer marks such the expressions in *s-term* as *transferred*.

Table 1: Translation results of 278 sentences

type	sentences	rate(%)
(A) correct	228	82.0
(B) unnatural	33	11.9
(C) incorrect	16	5.7
(D) failed	1	0.4

Table 2: Incorrect translations

cause	sentences	rate(%)
too many repairs	33	11.9
structural ambiguity	8	2.9
lexical ambiguity	3	1.1
idiomatic phrase	3	1.1
ill-formed source sentence	1	0.4
euphonic changes	1	0.4
parsing error	1	0.4

4 Evaluation

We have made an experiment on the prototype system. The conversations consisting of 4 dialogues and 278 English sentences in ATR Dialogue Database [Ehara et al., 90] have been used. The average sentence length is 6.8 words. The dialogue task is applications of travels. However, since the purpose of this study is to enhance the real-time nature of spoken language translation, the inputs are restricted on grammatically well-formed sentences from which the ill-formedness such as hesitations and errors were excluded.

First, the success rate was examined. As Table 1 shows, we have classified the sentences according to the translation results. 228 sentences are translated correctly, providing a success rate of 82.0%. Appendix shows typical correct translations. The result shows the system to be acceptable for spoken language translation. The translations gained on the system are different from those on a conventional system in the sense that they include many ill-formed expressions. However, the translations represent the semantic contents of the source sentences correctly. As Table 2 shows, furthermore, unsuccessful 50 sentences can be also classified according to the causes.

Next, the translation time of one input word was measured on SparcCenter 1000 (SuperSPARC 50MHz). The average time, 47msec, is less than the average speech time for one word, therefore, we can conclude the system to be usable for speech-to-speech translation with high real-time nature.

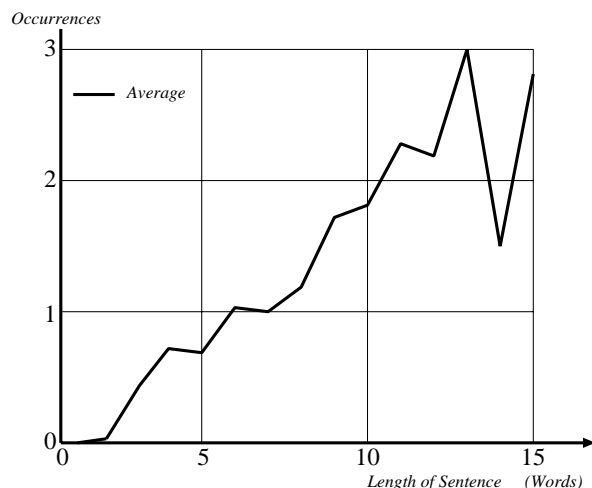


Figure 4: Relation between source sentence length and occurrences of repairs

5 On Producing Grammatically Ill-formed Sentences

Spontaneous spoken language, which is different from written language, can be characterized as the frequent appearances of ill-formed expressions [Sagawa et al., 94]. Accordingly, the methods of robustly analyzing ill-formed sentences have been required. Likewise, the effective techniques for producing spontaneous speech should be developed. To this end, utilizing grammatically ill-formed expressions must be one of such the techniques. In particular, as this paper demonstrates, it is considered that the appropriate use of repairs is effective for various real-time speech production systems.

The problem of the appearance of too many ill-formed expressions, however, has been left. As Table 2 shows, 11.9% translations have been judged as incorrectness because of too frequent appearances of repairs. In particular, the longer the source sentence becomes, the more repairs appear in the translation, as Figure 4 illustrates. In order to solve this problem, (1) the elaborate studies on the frequency of ill-formed expressions in utterances of human speakers/interpreters, and (2) the investigations that the users evaluate the Japanese speech which the system produces are required.

6 Concluding Remarks

This paper has proposed a system for incremental English-Japanese spoken language translation. Utilizing ill-formed expressions characterizing Japanese

speech has been a key to success of translating English speech to Japanese speech incrementally. Since the high degree of incrementality is required rather than that of quality in spoken language translation, it seems to be quite all right for spoken language translation systems to produce the sentences including ill-formed expressions.

We have implemented a prototype system in Common Lisp. The experiment with the system has shown the system to be usable for spoken language translation with acceptable accuracy and high real-time nature.

In this paper, the system utilizing the grammatical ill-formedness as the output has been described. On the other hand, we have not been concerned with the grammatically ill-formed input. In spontaneous speech, however, the ill-formedness as utilized in this paper appears frequently. Though it is not evident whether the chart parser introduced in this paper can process ill-formed sentences incrementally, it might be possible to construct incremental parser for ill-formed inputs by extending the methods of parsing ill-formed inputs on a sentence-by-sentence way [Mellish 89, Kato 94]. It is necessary to investigate a method of incrementally translating ill-formed input using the parser.

The achievement of this paper can be regarded as the first step toward the development of simultaneous interpreting system, which is just one of ambitious applications in the near future. As the related application system, we can consider multimodal dialogue translation. In order that the user may simultaneously understand various information through some modalities such as speech, text, face expressions, glances and gesture, the system is required to output the translation in a synchronous fashion. This system can be expected to be also available for this application.

Acknowledgements

The authors wish to thank Mr. Keiichi Iwashima, Mr. Yoshihide Kato and all members of their laboratory for their helpful discussions. The first author has been supported by Support Center for Advanced Telecommunications Technology Research, Foundation (SCAT).

References

- [Amtrup, 95] Amtrup J. W. Chart-based Incremental Transfer in Machine Translation. *Proc. of 6th Int. Conf. of Theoretical and Methodological Issues in Machine Translation (TMI'95)*, 188-195 1995.
- [Ehara et al., 90] Ehara T., Inoue N., Kohyama H., Hasegawa T., Shohyama F. and Morimoto T. Contents of the ATR Dialogue Database. *Technical Report TR-I-0186*, ATR Interpreting Telephony Research Laboratories, 1990. (in Japanese)
- [Kato 94] Kato T. Yet Another Chart-Based Techniques for Parsing Ill-Formed Input. *Proc. of 4th Conf. on Applied Natural Language Processing (ANLP'94)*, 107-112 1994.
- [Inagaki and Matsubara, 95] Inagaki Y. and Matsubara S. Models for Incremental Interpretation of Natural Language. *Proc. of 2nd Symposium on Natural Language Processing (SNLP'95)*, 51-60 1995.
- [Kay, 80] Kay M. Algorithm Schemata and Data Structures in Syntactic Processing. *Technical Report CSL-80-12*, Xerox PARC, Oct. 1980.
- [Kitano, 90] Kitano, H. Incremental Sentence Production with a Parallel Marker-Passing Algorithm. *Proc. of 13th Int. Conf. on Computational Linguistics, (COLING'90)*, 217-222 1990.
- [Matsubara and Inagaki, 96] Matsubara S. and Inagaki Y. A Study on Features and Phenomena of Japanese Spoken Language in Simultaneous Interpretation. *Proc. of IE-ICE Fall Information and Systems Society Conf. '96*, 55 1996. (In Japanese)
- [Mellish 89] Mellish C.S. Some Chart-Based Techniques for Parsing Ill-Formed Input. *Proc. of 27th Annual Meeting of Association for Computational Linguistics (ACL'89)*, 102-109 1989.
- [Menzel, 94] Menzel W. Parsing of Spoken Language under Time Constraints. *Proc. of 11th European Conf. on Artificial Intelligence (ECAI'94)*, 560-564 1994.
- [Sagawa et al., 94] Sagawa Y., Ohnishi N. and Sugie N. A Parser Coping with Self-Repaired Japanese Utterances and Large Corpus-based Evaluation. *Proc. of 15th Int. Conf. on Computational Linguistics, (COLING'94)*, 593-597 1994.
- [Steedman, 87] Steedman M. Combinatory Grammar and Human Sentence Processing. J. Garfield (ed.), *Modularity in Knowledge Representation and Natural Language Understanding*, MIT Press, 187-205 1987.

Appendix

Typical Correct Translations ¹

- Hi, yes, I am interested in booking a tour to Hong-Kong, please.
もしもし (hi), はい (yes), 興味があります (I am interested). ツアーを予約することに (in booking a tour) えーと, 香港への (to Hong-Kong), ツアーを予約することに (in booking a tour) 興味があります (interested).
- And what does that include?
それと (and) 何は (what), あっ, 何を (what) それは (that) 含みますか (include) ?
- I think it was 3days.
思います. (I think) それが (it) 3日間であったと (was 3days).
- I am looking forward to hearing from you.
楽しみにしています (I am looking forward). 話しを聞くこと (to hearing), えーと, あなたから話しを聞くことを (to hearing from you) 楽しみにしています (am looking forward).
- I would like to stay longer if possible.
滞在したいです (I would like to stay). より長く (longer) もし可能なら (if possible) 滞在したいです (would like to stay).
- Does that price include meals?
それは (that), えーと, その値段は含みますか (does that price include) ? 食事を (meals) 含みますか (does include) ?
- How much would this trip cost?
いくら (How much) この旅行は (this trip) 費用がかかる (cost) だろうか (would) ?
- Actually 6 or 7 would be best for me?
実際 (actually), 6時または7時は (6 or 7) 最もよい (be best), あのー, 私にとって (for me) 最もよいものであるだろう (would be best).
- I am in Japan for a short visit with my husband.
日本に (in Japan) 少しの間います (am for a short visit). えー, 私の夫といます (am with my husband).
- Well, I am kind of interested in seeing some Japanese temples and Buddas and things like that.
えーと (well), ちょっと興味があります (I am kind of interested). いくつかの日本の寺と仏陀, それとそのような類のものを見ることに (in seeing some Japanese temple and Buddas and things like that) ちょっと興味があります (am kind of interested).
- I like taking showers.
好きです (I like). シャワーをすることが (taking showers) 好きです (I like).
- I am staying at the Marunouchi Hotel.
泊まっています (I am staying). えーと, 丸の内ホテルに (at the Marunouchi Hotel) 泊まっています (am staying).
- Leaving on the 24th and leaving on the 31st?
24日に出発 (leaving on the 24th) と (and) 31日に出発 (leaving on the 31st) ?
- I don't think that's possible.
思いません (I don't think). それが可能であると (that's possible).
- Sure, the name is Lenelle Degenhardt.
ええ (sure), 名前は (the name) レネーレ デーゲンハルトです (is Lenelle Degenhardt).
- I don't have a phone.
もっていません (I don't have). 電話を (a phone). もっていません (don't have).
- I'm a university student here in Japan and I am interested in booking a tour to India.
大学生 (a university student), えー, ここ (here) 日本の大学生です (am a university student in Japan). それと (and) 興味があります (I am interested). ツアー (a tour), あ, インドへのツアーを (a tour to India) 予約することに (in booking) 興味があります (am interested).

¹“えーと”, “あっ”, “あのー”, “えー” and “あ” are kinds of hesitations in Japanese.