Translation Unit Concerning Timing of Simultaneous Translation

Hideki KASHIOKA
ATR Spoken Language Translation Research Laboratories

2-2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-0288 Japan
hideki.kashioka@atr.co.jp

Abstract

This paper discusses and proposes a translation unit for simultaneous translation using a machine translation (MT) system. Monologues, such as lectures or broadcast news, are used as the target of simultaneous speech translation. To date, a lot of research on speech translation has dealt with dialogues, especially travel conversations. Most of the speech translation systems in MT have treated a sentence as a translation unit. In the ATR travel conversation database, sentence length is less than 10 words on average. Therefore, most of the sentences are simple and almost all of the utterances are constructed in one or two sentences. However, the sentences of monologues are longer than travel dialogues. They have over 30 words (as in "ASU-wo-YOMU," a TV news commentary program) on average, and most of the sentences are complex or compound. Accordingly, it is difficult to treat a sentence as a translation unit for monologues, and thus an appropriate translation unit needs to be found. Considering this, we hypothesized that an adequate translation unit of speech translation systems relates to the translation unit of a human simultaneous translator. Therefore, we collected simultaneous translation data from lectures by human translators and investigated the characteristics of monologues and simultaneous translation.

1. Introduction

This paper discusses and proposes a translation unit for simultaneous translation using a speech translation system for monologues, such as lectures or broadcast news.

To date, a lot of research on speech translation has dealt with dialogues, especially travel conversations.(Nakamura et al., 2001; Levin et al., 2000) Most speech translation systems treat a sentence as a translation unit for sentences less than 10 words in length (in the ATR travel conversation database) on average. Therefore, most of the sentences are simple, and almost all of the utterances are constructed in one or two sentences. However, the sentences of monologues are longer than travel dialogues. They have over 30 words (as in "ASU-wo-YOMU," a TV news commentary program) on average, and most of the sentences are complex or compound. Accordingly, it is difficult to treat a sentence as a translation unit for monologues.

Considering another point of view, human translators, especially simultaneous translators, translate constituents incrementally, instead of by sentence, over original speech. Therefore, we hypothesized that an adequate translation unit of speech translation systems relates to the translation unit of a human simultaneous translator. We collected simultaneous translation data from lectures by human translators and investigated the characteristics of the monologues and simultaneous translation.

In this paper, we describe how to collect simultaneous translation data and the features of these data. Then, we observe the simultaneous translation data for finding an adequate translation unit with two points of view. We then discuss and propose an appropriate translation unit.

2. Simultaneous Translation Data

We chose monologues from a TV news commentary program, called "ASU-wo-YOMU." "ASU-wo-YOMU" is a weekday program broadcast daily by NHK. Each program is 10 minutes in length and deals with hot topics (i.e., economy, politics, international circumstances, etc.). In this section, we describe how to collect simultaneous interpretation data, and set forth the statistics of our collected data and a comparison between monologues and dialogues.

2.1. How to make simultaneous translation data

We recorded the simultaneous interpretation that was translated by a simultaneous interpreter for a video of "ASU-wo-YOMU." While recording simultaneous interpretation, we note the number of the interpreter and the recording environment.

Each program was translated by one simultaneous interpreter without an assistant. Usually, simultaneous interpreters consist of a team with two or three interpreters. The main interpreter translates the original speech and the other members support the main interpreter. The roles were changed alternately about every 30 minutes. In the case of our recordings, the programs were not so long and the original speaker did not hand out any papers during his/her talk. Therefore, there was no requirement to obtain those papers or to arrange something to drink, etc. In addition, a week before the recording, we notified the simultaneous interpreter of the keywords, as shown in Table 1, for the target program. The keywords chosen were considered important and/or necessary words. Thus, we recorded simultaneous translated data by one interpreter.

These recording speech data are constructed with two channel data: one channel is an original commentator's speech and the other channel is a simultaneous interpreter's speech. These two channels are synchronised. Therefore, we can analyse the relationship of these two types of speech.

We transcribed these two speech data with time information. In transcriptions as shown in Tables 5 and 6, we divided utterances with pauses and fillers. The pause is defined as a silence length of over 200 ms. The speech data of the original speaker was aligned with a transcription by using a module of a speech recognition tool (i.e., ATR sprec).

Of course, the simultaneous translation data is parallel data. However, it is difficult to construct sentence alignment on such data because human interpreters summarise

Teiki syakuya (fixed term rental housing)
Chintai apaato (rental houses)
Chintai jimusyo (rental offices)
akewatasi (evacuation)
syakuti syakuyahou (leasehold and tenancy law)
seitou jiyuu (just cause)
yachin tainou (the delinquency in the payment of rents)
tachinoki ryo
(the payment of compensation for evacuation)
giin teiann (the bill proposed by the parliamentarians)
sinki keiyaku (the new contracts)
tyuuto kaiyau (cancel the contract before the expiry)
teiki syakuti ken (fixed term system)
houmu syou (the Ministry of Justice)
fudousan sijo (the real estate market)
juutakukensetu sijo (housing construction market)
kasseika (activate)
jyakusya kiri sute (the under-privileged)
suisin
roukyuu syakuya (the old rental houses)
kasi siburi (the reluctance to rent)
kizonn keiyaku (the existing contracts)
juutaku jijou (the housing situations)
juutaku seisaku no hinkon
fukusi seisakuno yugami
seitou jiyuu seido
syanimuni (do it up)

Table 1: Sample of Keyword for the Program "TEIKI SYAKUYA SEIDO DOUNYUU E"

the data, use paraphrases, use ellipses, and reorder the contents.

2.2. The statistics of simultaneous translation data

The Japanese transcription(Table 5) of each program has about 2000 words and about 60 sentences on average. The maximum sentence length is 178. The segmentation of the sentence depends on the person who transcribed the speech data. Vocabulary size is about 500 per transcription of one program. On the other hand, the English transcription (Table 6) of each program has about 1,200 words and 70 sentences on average. Vocabulary size is about 400 per transcription of one program. It is likely that the contents of the original sentence are divided into two or more sentences in interpretation and that some of them are deleted. Additional information on filler (included intersections) appeared 120 times in the Japanese transcription and 164 times in the English transcription on average per program.

2.3. Comparison between monologue and dialogue

We now compare the features of the monologues with the features of the dialogues. The monologues' data are the transcriptions of "ASU-wo-YOMU," and the dialogues' data are the travel conversation's transcriptions collected by ATR.

Dialogue (Travel conversation):
about 10 words/sentence,
80% single sentences,
20% complex or compound sentences, and
0.3 dependent clauses in a sentence.

Monologue (TV news commentary program):
about 30 words /sentence,
20% single sentences,
80% complex or compound sentences, and
2.1 dependent clauses in a sentence.

The sentence length of the monologue data is three times that of the dialogue data. The ratio of single sentences and complex sentences is exchanged between dialogues and monologues. Moreover, the number of dependent clauses differs greatly between them. It is clear that the monologues are more difficult to analyze than dialogues.

3. The Features Of Simultaneous Translation

In the previous section, we described the features of monologues and it is clear that monologues are constructed with more difficult sentences. Therefore, monologues are hard to treat using a sentence as a translation unit. Other translation units for simultaneous speech translation are required. Simultaneous translators do not translate sentence by sentence. Consequently, we examined two points in these data for considering the translation unit for simultaneous speech translation:

1. The timing of the starting point of a translator's utterance in this parallel corpus,

2. The delay time between a keyword and a translator's corresponding word.

3.1. Translation timing

Our first point of view is "The timing of the starting point of a translator's utterance in this parallel corpus."

We presume that a PAUSE in a translator's utterance has a specific meaning for the simultaneous translation data. Of course, a translator's speech includes PAUSEs for breathing purposes or for ordinarily emphasising points. However, a translator in simultaneous translation is pressed for time, and he/she cannot freely control the content of his/her utterances. Accordingly, we presume that a PAUSE, especially a long PAUSE, indicates a translator's sign to wait and translate the contents of the original speaker's utterance. In other words, the end of a PAUSE by the translator tells us the content unit of the original speech. In this situation, the starting point of the translator's utterance is defined as the end of the translator's PAUSE.

We found about 8,500 PAUSEs in total in the 50 programs. 1,155 of the PAUSEs are more than 1,000 ms in length. The length of most pauses are 250 ms – 350 ms. The maximum length of the pause is about 7,000 ms.

The translator needs time to translate the original speech he/she hears to the speech he/she translates. Concerning this delay time, we need to check the Part of Speech (POS) around the starting point of the translator's utterance. Therefore, the following investigation assigned 200 ms to this delay time.

POS	# of total	# at that timing	%
PAUSE	9816	1937	20
COMMON NOUN	17577	2091	12
VERBAL NOUN	5267	639	12
ADJECTIVE NOUN	1235	136	11
CONJUNCTION	862	84	10
VERB	7204	674	9
ADJECTIVE	836	109	13
ADVERB	1421	143	10

Table 2: The frequency of POS

We checked the POS or PAUSE of each word uttered by the original (Japanese) speaker with the timing at the end of the interpreter's pause.

Table 2 indicates the frequency in the data and the frequency of the POS at that timing. The total running words in the commentator's speech is about 96,000. The interpreter's pauses are about 8700. Thus, if the timing of the starting translation did not relate the POS of the original speech, each POS can be found at that timing with almost the same percentage of total frequency of each POS (in this case about $9\% = 8700/9600$). However, the POS listed in Table 2 can be found more than 9%. Therefore, we must draw attention to these POSs.

POS	%
PAUSE	22
COMMON NOUN	24
VERBAL NOUN	7
ADJECTIVE NOUN	2
CONJUNCTION	1
VERB	8
ADJECTIVE	1
ADVERB	1

Table 3: The frequency of POS that appears exactly at the starting point of a translator's utterance in the original data

As shown in Table 3, the most frequent POS that appears exactly at the starting point of a translator's utterance in the original data is a COMMON NOUN (24%). The next is a PAUSE (22%). The sum of the predicates (i.e., VERB, AUXILIARY VERB, etc.) is about 17%. The VERBAL NOUN is 7%. The other POS rates are less than these POS rates.

POS	Frequency	%
PAUSE	586	28
COMMON NOUN	252	12
ADNOMINAL PARTICLES	403	19

Table 4: The previous POS frequency of a NOUN that appears exactly at the starting point of a translator's utterance in the original data

Moreover, we checked the POS of one more previous word in the case of a NOUN at that timing. (Table 4) We found that about 28% of all NOUNs are PAUSEs, while 19% of all NOUNs are ADNOMINAL PARTICLES. Therefore, these observations suggest that the original speaker's pause is a key to the interpreter to start the translation. In addition, the AUXILIARY VERB is the most frequent POS of one more previous word in the case of a PAUSE at that timing. AUXILIARY VERBs covered 30% of the PAUSEs at that timing.

3.2. Keyword translation delay

Our second point of view is "The delay time between a keyword and a translator's corresponding word."

Simultaneous translation data is parallel data with utterance time information. If sentence alignment on such data performs well, we can easily observe the delay time that the interpreter translates from the word to the corresponding word. However, the performance of the sentence alignment is not very good because human interpreters summarise the data, use paraphrases, use ellipses, and reorder the contents. Therefore, observation of the delay time from any word to its corresponding word is difficult. In spite of this difficulty, it is highly probable that the keywords can be observed in the delay time because we provided the translator with a list of keywords for each program one week before recoding the simultaneous translations.

We checked the delay time against the appearance of each of these keywords. The number of keywords was about 20 for each program. Almost all of the keywords could be found in the translator's utterances, and these keywords were sometimes found not only once but a few times. The utterances that included keywords had five or six minute delays on average. Sample utterances that include the keyword and its corresponding word are shown in Table 7.

4. Discussion

On the assumption that the translation unit of the simultaneous translator can be observed in the pause of the translator's speech, we can define the translation unit for a speech translation system using the features of the commentator's speech related to the simultaneous translator's translation unit. Usually, pauses are used for breathing purposes or for emphasising points. However, the translator has no time to waste following the original speech. In such a situation, translators do not use long pauses for breathing purposes or for emphasising points. We thus hypothesised that the pause of a translator, particularly a long pause, is related to the translation unit of the simultaneous translator. As a result we checked the relationship the pause of the translator with the original speech in Section 3.1., and found that the PAUSE, COMMON NOUN, and VERB in the original speech is a key to finding the translation unit of the simultaneous translator. Even then, the PAUSE, COMMON NOUN and VERB frequently appear in original speech. 30% of the PAUSEs appeared at the starting point of the translation, followed by the AUXILIARY VERB. It should be noted that the VERBAL NOUN and the VERB frequently occur at the starting point of the translation. In Japanese, a clause boundary can be found around a predicate. These POSs (AUXILIARY VERB, VERBAL NOUN and VERB) are classified as predicates. Thus, we suggest that the PAUSE and predicate in the original speech is an important signal for the translation unit, and that the clause has the ability of being a translation unit.

On the other hand, NOUNs occur frequently at the starting point of the translation. It appears that the NOUN at the starting point of the translation can be classified. Two types of these classification are convincingness. The first is constructed as a subject in the following sentence. The second is noun on which the clause depends.

5. conclusion

We observed the translation unit for simultaneous translation by taking a general view of the translator's data. We hypothesized that the translator's PAUSE is an important signal for the translation unit. Therefore, the pause or predicate appearances in the original speech are related to the translation unit. We are currently planning alignment using a simultaneous translation corpus, and using information about keyword delay times together with presumptions on the translation unit. We are also developing a simultaneous translation system.

6. acknowledgements

The research reported here was supported in part by a contract with the Telecommunications Advancement Organization of Japan entitled, "A study of speech dialogue translation technology based on a large corpus."

7. References

Lori Levin, Boris Bartlog, Ariadna Font Llitjos, Donna Gates, Alon Lavie, Dorcas Wallace, Taro Watanabe, and MonikaWoszczyna. 2000. Lessons learned from a task-based evaluation of speech-to-speech machine translation. In *LREC 2000*.

Nakamura, Naito, Tsukada, Gruhn, Sumita, Kashioka, Nakajima, Shimizu, and Sagisaka. 2001. A speech translation system applied to a real-world task/domain and its evaluation using real-world speech data. *IEICE Transactions on Information and Systems(English)*, E84-D(1):142–154.

Start	End	utterance
0	4219	PAUSE
4219	4310	[e]
4310	4782	konbanwa
4782	5272	[e]
5272	5399	kotosi mo syuntou no kisetsu ga megutte mairi mashita
5399	8221	[ma]
8221	8917	kinou nikkeiren ga rinji soukai wo hiraki masite
8917	9004	[e]
9004	11403	chingin yokusei no sisei wo taihen sennmei ni itasimasita
11403	11712	PAUSE
...		

Table 5: Sample Transcription of Commentator Speech(Japanese)

Start	End	utterance
0	5099	PAUSE
5099	7005	good evening, ladies and gentlemen.
7005	7495	PAUSE
7495	8534	the spring
8534	8802	PAUSE
8802	11819	bargaining season is coming to close.
11819	12286	PAUSE
12286	14369	the Japan Economic Federation
14369	14850	PAUSE
14850	19292	made a meetings to stress the importance of the suppression of the
19292	19568	[ah]
19568	20226	salaries,
20226	20612	PAUSE
20612	20714	and
20714	20773	[ah]
20773	23756	Central Struggle Committee of the labour unions
23756	24303	PAUSE
...		

Table 6: Transcription Sample of Translator Speech(English)

TEIKI SYAKUYA / fixed term rental housing
J:2990 - 6477 minasan-ha teiki syakuya toiu seido o gozonji desyouka
E:7996 - 10099 fixed term rental housing?
J:25864 - 31810 teiki syakuya toiu seido no naiyou aruiwa kono mondai kadai nituite konnya wa kangaete mairitai toiufuuni omoimasu
E:28672 - 31206 the program, fixed term rental housing
CHINTAI APAATO / rental houses
KASHI JIMUSYO / rental offices
J:7231 - 9263 chintai apaato kashi jimusyo nado
E:11225 - 15478 for rental offices or rental houses at the expiry of the contract
AKEWATASHI / evacuation
J:12263 - 16294 yanushi ga akewatashi o seikyuu suru kotoga dekiru toiu seido desu
E:19614 - 21045 request evacuation.
SYAKUTI SYAKUYA HOU / leasehold and tenancy law
J:37344 - 39249 ima no syakuti syakuya hou dewa
E:42061 - 44053 leasehold and tenancy law
SEITOU JIYUU / just cause
J:48270 - 51909 mondai ni sarete orimasu nowa seitouna jiyuu ga nai kagiri
E:62226 - 65409 just cause they are not allowed to request evacuation.

Table 7: Sample utterance including the keyword and its corresponding word