

Representing 'how you say' with 'what you say': English corpus of focused speech and text reflecting corresponding implications

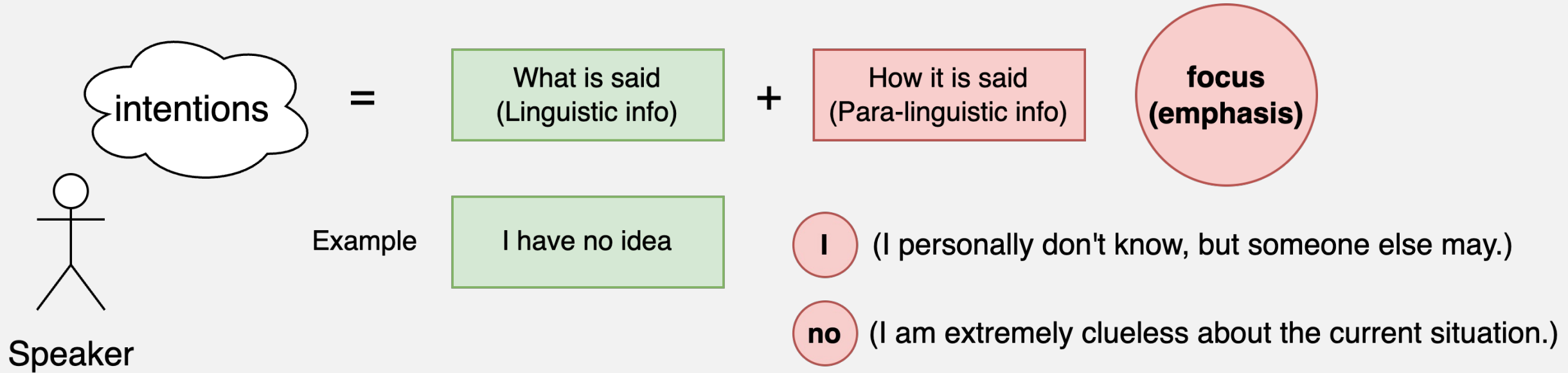
Naoaki Suzuki, Satoshi Nakamura

Nara Institute of Science and Technology, Japan

INTERSPEECH 2022

Introduction

- Speech communication



- Implications can be different

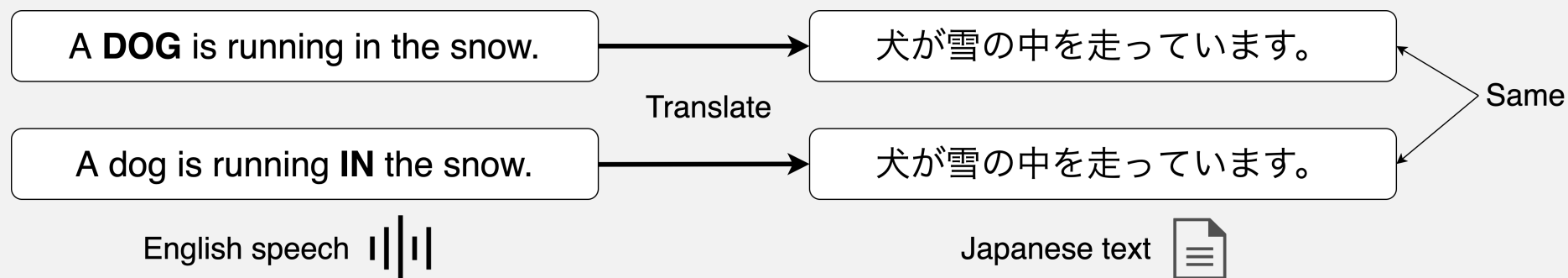
- Even with the same linguistic information

- Speech translation

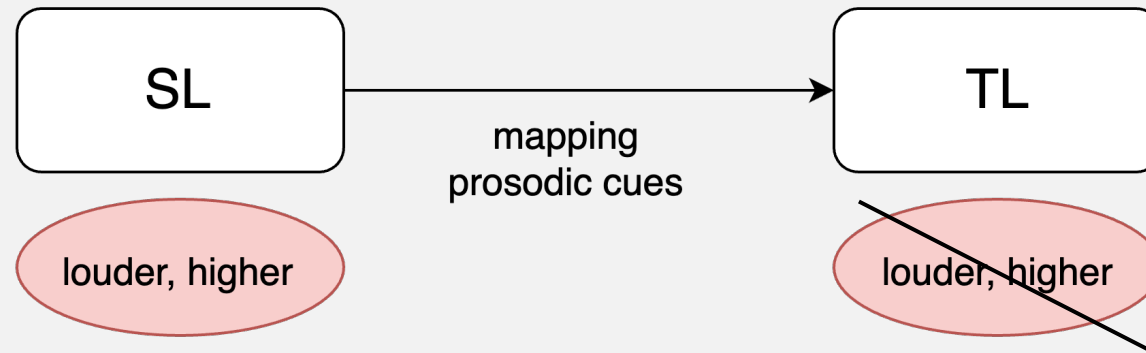
- Translates speech in one language into text/speech in another

- Limitation

- Unable to consider paralinguistic info
- If the linguistic info is the same, so are the translations



- Acoustic to acoustic mapping
 - Mapping acoustic cues in the source language (SL) to the counterparts in the target language (TL) [Aguero+ 2006, Kano+ 2013, Do+ 2018]



- What if prosodic counterparts do not exist in the TL?

Proposed Method

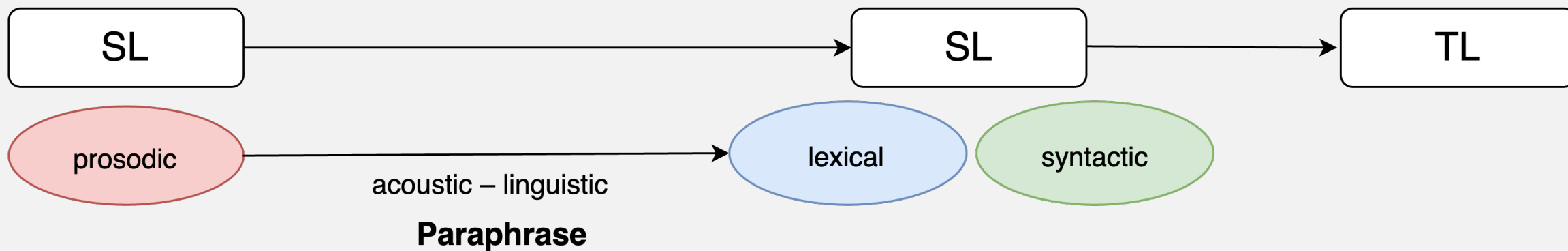
- Three devices for focus [Cruttenden 1997]

Prosodic e.g., higher, louder, longer

Lexical e.g., *very*, *even*

Syntactic e.g., passive, cleft

- Acoustic – linguistic mapping by paraphrasing



example

A DOG is running in the snow

The creature that is running in the snow is a dog

- Fundamentals for the achievement of acoustic – linguistic focus transformation

1. Corpus construction

- Speech having different items in focus
- Text reflecting the relevant implications

2. Relationships between focused speech and focused text

- What kind of methods are used for paraphrasing?

Corpus Construction

How to build the corpus

- Flickr8k [Rashtchian+ 2010]

- 8000 images which depict actions relating to people or animals
- Five text descriptions are given for each image



A beagle and a golden retriever wrestling in the grass
Two dogs are wrestling in the grass.
Two puppies are playing in the green grass.
two puppies playing around in the grass
Two puppies play in the grass

× 8000

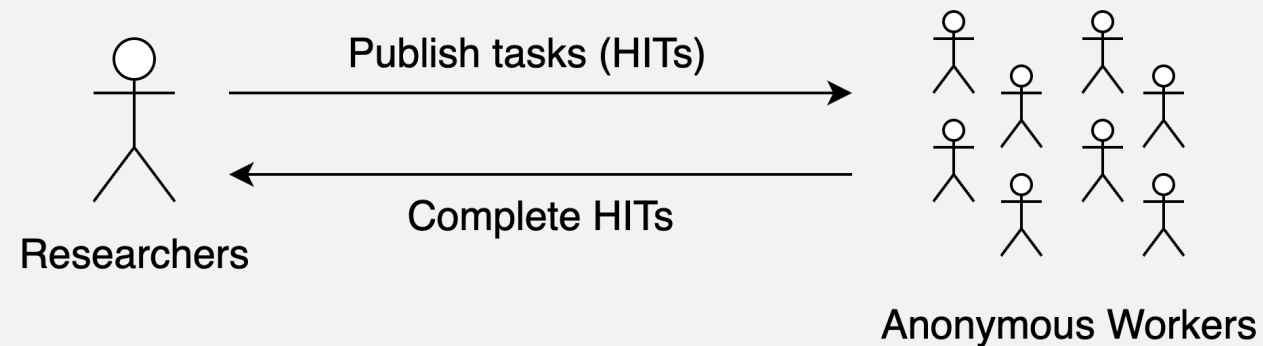
- Selected 196 short sentences as the source (words length: max six words)

- Focus placement

- every word to be the target of focus

A dog runs in the snow
A dog runs in the snow
A dog runs in the snow
A dog runs in the snow
A dog runs in the snow
A dog runs in the snow

- Amazon Mechanical Turk (MTurk)
 - a crowd sourcing platform
 - allows researchers to create tasks called HITs and anonymous users (Workers) to complete them for a small monetary fee



- Underlined text



A biker enjoys a coffee.

A biker enjoys a coffee.

A biker enjoys a coffee.

A biker enjoys a coffee.

A biker enjoys a coffee.

A biker enjoys a coffee.

- Instructions

Make a recording emphasising the underlined word
if it does not sound unnatural


- Subjects

- 10 British native English speakers
- 3 speakers/caption

- Results

- focused: 2800
- normal (without focus): 600

- Focused speech

- a 

- biker 

- Subjects

- 16 native English speakers
 - 2 paraphraser / focused speech

- Instructions

Paraphrase the speech by writing your own sentence in a way that clearly conveys the meaning implied by the emphasised word.

- Results

- 2100 paraphrases
 - e.g. *'It is a biker enjoying a coffee'*

- Pairs of

- focused speech
- paraphrase

- Subjects

- 16 native English speakers
- 3 participants/paraphrase

- Results

- 1700 paraphrases

- Instructions

- Rate how accurately does the written sentence convey what is implied in the speech?
- Is the written sentence formatted well as an English sentence?

- Five scales (1 – 5) for both

Low scores



Collect and evaluate
paraphrases again

High scores



Accept the paraphrases

Focused speech	Paraphrase
A biker enjoys a coffee	One biker enjoys a coffee
A biker enjoys a coffee	There is one biker enjoying a coffee
A biker enjoys a coffee	It's a biker enjoying a coffee
A biker enjoys a coffee	The person enjoying a coffee is a biker
A biker enjoys a coffee	A biker drinking a coffee is enjoying it
A biker enjoys a coffee	The biker seems to enjoy a coffee
A biker enjoys a coffee	A biker enjoys one coffee
A biker enjoys a coffee	A biker has one coffee he enjoys
A biker enjoys a coffee	What the biker enjoys is a coffee
A biker enjoys a coffee	It is coffee the biker enjoys

Analysis

What kind of transformation methods were used for paraphrasing?

- How focus in speech was mapped into focus in text?
 - We manually examined the original text – paraphrase pairs
- Lexical

Types	Original Phrase	Paraphrase
Substitution substitute the focused word with its synonyms	People sit <u>on</u> benches.	People sit <u>on top of</u> benches..
Modification modify the focused word or its phrase with modifiers	Two brown dogs <u>play</u> .	Two brown dogs play <u>enthusiastically</u> .
Negation explicitly state an alternative of the focused word and negate it	A <u>man</u> stands outside.	A man, <u>not a woman</u> , stands outside.

- How focus in speech was mapped into focus in text?
 - We manually examined the original text – paraphrase pairs
- Syntactical

Types	Original Phrase	Paraphrase
Leftward shift move the focused word towards the beginning of the sentence	Two dogs splash through the water.	<u>Splashing</u> is what two dogs are doing through the water.
Rightward shift move the focused word towards the end of the sentence	Two lizards fight in the water.	What the two lizards do in the water is <u>fight</u> .
Tense change change the tense from simple to progressive or vice versa	The woman is walking her dogs.	The woman <u>walks</u> her dogs.

- A certain part-of-speech was more likely to use a certain transformation method
 - randomly sampled 50 paraphrases for each part-of-speech
 - counted each transformation method for each paraphrase

		N	V	Adj	Num	Aux	P	Det
Lexical	Substitution	0.28	0.12	0.10	0.18	0.08	0.42	0.72
	Modification	0.00	0.02	0.12	0.06	0.60	0.18	0.50
	Negation	0.10	0.06	0.06	0.12	0.10	0.14	0.00
Grammatical	Leftward	0.10	0.20	0.04	0.08	0.00	0.08	0.00
	Rightward	0.46	0.44	0.70	0.52	0.08	0.26	0.02
	Tense	-	0.28	-	-	0.20	-	-

Mean occurrences of each transformation method per part-of-speech (N: Noun, V: Verb, Adj: Adjective, Num: Numeral, Aux: Auxiliary, P: Preposition, Det: Determiner)

Discussion and Conclusion

1. Corpus construction

- Speech having different items in focus
- Text reflecting the relevant implications

2. Relationships between focused speech and focused text

- What kind of methods are used for paraphrasing?
 - Broad categorization of transformation methods
 - Tendency dependent on part-of-speech
-
- Limitation
 - Lack of context

- A new direction for paralinguistic translation
 - Demonstrated the possibility of mapping paralinguistic info to the linguistic domain with lexical and syntactic devices
 - The corpus and insights from our analysis will lead us to construct a speech translation model which can preserve paralinguistic information

- Aguero, P. D., Adell, J., & Bonafonte, A. (2006, May). Prosody generation for speech-to-speech translation. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (Vol. 1, pp. I-I). IEEE.
- Cruttenden, A. (1997). *Intonation*. Cambridge University Press.
- Derwing, T. M., Thomson, R. I., Foote, J. A., & Munro, M. J. (2012). A longitudinal study of listening perception in adult learners of English: Implications for teachers. *Canadian modern language review*, 68(3), 247-266.
- Do, Q. T., Sakti, S., & Nakamura, S. (2018). Sequence-to-sequence models for emphasis speech translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(10), 1873-1883.
- Downing, L. J. (2006). The prosody and syntax of focus in Chitumbuka. *ZAS papers in linguistics*, 43, 55-79.
- Kano, T., Takamichi, S., Sakti, S., Neubig, G., Toda, T., & Nakamura, S. (2013, August). Generalizing continuous-space translation of paralinguistic information. In *INTERSPEECH* (Vol. 445, pp. 25-29).
- Rashtchian, C., Young, P., Hodosh, M., & Hockenmaier, J. (2010, June). Collecting image annotations using amazon's mechanical turk. In *Proceedings of the NAACL HLT 2010 workshop on creating speech and language data with Amazon's Mechanical Turk* (pp. 139-147).