Sec 4.2

Utterance Segmentation

Utterance segmentation

SLT evaluation has an additional level of complexity compared to machine translation.

Utterance segmentation

SLT evaluation has an additional level of complexity compared to machine translation.

Machine Translation:

Document:

This is an audio signal. In the training data it was split using strong punctuation. Three sentences in total.

Utterance segmentation

SLT evaluation has an additional level of complexity compared to machine translation.

Machine Translation:

Document:

This is an audio signal. In the training data it was split using strong punctuation. Three sentences in total.

Source sentences:

This is an audio signal.

In the training data it was split using strong punctuation.

Three sentences in total!

Utterance segmentation

SLT evaluation has an additional level of complexity compared to machine translation.

Machine Translation:

Document:

This is an audio signal. In the training data it was split using strong punctuation. Three sentences in total.

Source sentences:

This is an audio signal.

In the training data it was split using strong punctuation.

Three sentences in total!

Reference sentence:

Questo e’ un segnale audio.

Nei dati di training e’ stato diviso usando la punteggiatura forte.

Tre frasi in totale!

Utterance segmentation

SLT evaluation has an additional level of complexity compared to machine translation.

Machine Translation:

Source sentences:	MT sentences:	Reference sentence:
This is an audio signal.	Questo è un segnale audio.	Questo e’ un segnale audio.
In the training data it was split using strong punctuation.	Nei dati di allenamento è stato suddiviso utilizzando una forte punteggiatura.	Nei dati di training e’ stato diviso usando la punteggiatura forte.
Three sentences in total!	3 frasi in totale!	Tre frasi in totale!

Utterance segmentation

SLT evaluation has an additional level of complexity compared to machine translation.

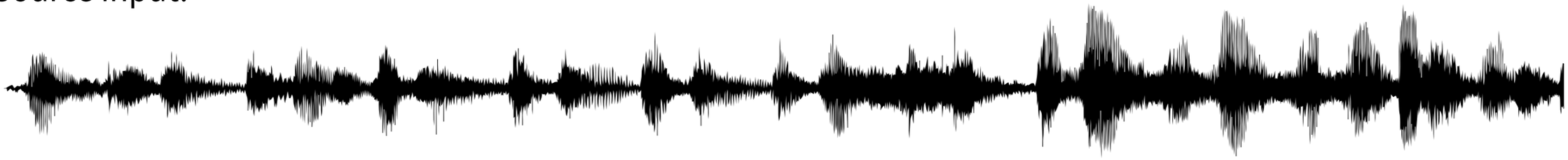
Machine Translation:

Source sentences:	MT sentences:	Reference sentence:
This is an audio signal.	Questo è un segnale audio.	Questo e’ un segnale audio.
In the training data it was split using strong punctuation.	Nei dati di allenamento è stato suddiviso utilizzando una forte punteggiatura.	Nei dati di training e’ stato diviso usando la punteggiatura forte.
Three sentences in total!	3 frasi in totale!	Tre frasi in totale!

Utterance segmentation

Spoken Language Translation:

Source input:

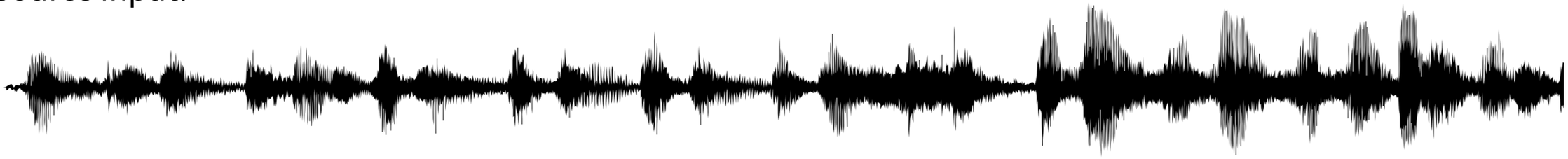


t h i s i s a n a u d i o s i g n a l i n t h e t r a i n i n g d a t a i t w a s s p l i t u s i n g s t r o n g p u n c t u a t i o n t h r e e s e n t e n c e s i n t o t a l

Utterance segmentation

Spoken Language Translation:

Source input:



t h i s i s a n a u d i o s i g n a l i n t h e t r a i n i n g d a t a i t w a s s p l i t u s i n g s t r o n g p u n c t u a t i o n t h r e e s e n t e n c e s i n t o t a l

Reference sentences:

This is an audio signal.

In the training data it was split using strong punctuation.

Three sentences in total!

Utterance segmentation

SLT outputs depend on the segmentation of the audio input:

This is an audio

Signal in the training data was split.

Using strong punctuation, 3 sentences in total!

Utterance segmentation

SLT outputs depend on the segmentation of the audio input:

This is an audio

Signal in the training data was split.

Using strong punctuation, 3 sentences in total!

This is an audio signal in the training data.

It was split using strong punctuation.

Three sentences in total!

Utterance segmentation

SLT outputs depend on the segmentation of the audio input:

This is an audio

Signal in the training data was split.

Using strong punctuation, 3 sentences in total!

This is an audio signal in the training data.

It was split using strong punctuation.

Three sentences in total!

This is a signal.

In the training data.

It was split in three sentences.

Utterance segmentation

SLT outputs depend on the segmentation of the audio input:

This is an audio

Signal in the training data was split.

Using strong punctuation, 3 sentences in total!

This is an audio signal in the training data.

It was split using strong punctuation.

Three sentences in total!

This is a signal.

In the training data.

It was split in three sentences.

This is

Signal. In the training data

it was split using strong punctuation.

Three sentences

in total!

Utterance segmentation

SLT outputs depend on the segmentation of the audio input:

This is an audio

Signal in the training data was split.

Using strong punctuation, 3 sentences in total!

This is an audio signal in the training data.

It was split using strong punctuation.

Three sentences in total!

This is a signal.

In the training data.

It was split in three sentences.

This is

Signal. In the training data

it was split using strong punctuation.

Three sentences

in total!

Reference sentences:

This is an audio signal.

In the training data it was split using strong punctuation.

Three sentences in total!

SLT output - reference alignment

1. How to compare the automatically split SLT outputs with the manually split references?
2. How to compare different systems splitting the SLT outputs in different ways?

SLT output - reference alignment

1. How to compare the automatically split SLT outputs with the manually split references?
2. How to compare different systems splitting the SLT outputs in different ways?

Issues:

- Different number of sentences
- Truncated SLT sentences
- Insertion of additional text in the SLT outputs
- Missing large parts in the SLT outputs

Concatenation

SLT output:

This is | Signal. In the training data | it was split using strong punctuation. | Three sentences | in total!

Reference sentences:

This is an audio signal. | In the training data it was split using strong punctuation. | Three sentences in total!

Concatenation

SLT output:

This is Signal . In the training data it was split using strong punctuation . Three sentences in total !

Reference sentences:

This is an audio signal . In the training data it was split using strong punctuation . Three sentences in total !

The concatenated STL outputs (references) are considered as a single sentence.

Automatic metrics applied on two strings.

Much less precise than working at segment level, but fast to implement

Automatic re-segmentation algorithm

SLT output:

This is Signal . In the training data it was split using strong punctuation . Three sentences in total !

Reference sentences:

This is an audio signal . In the training data it was split using strong punctuation . Three sentences in total!

Automatic re-segmentation algorithm

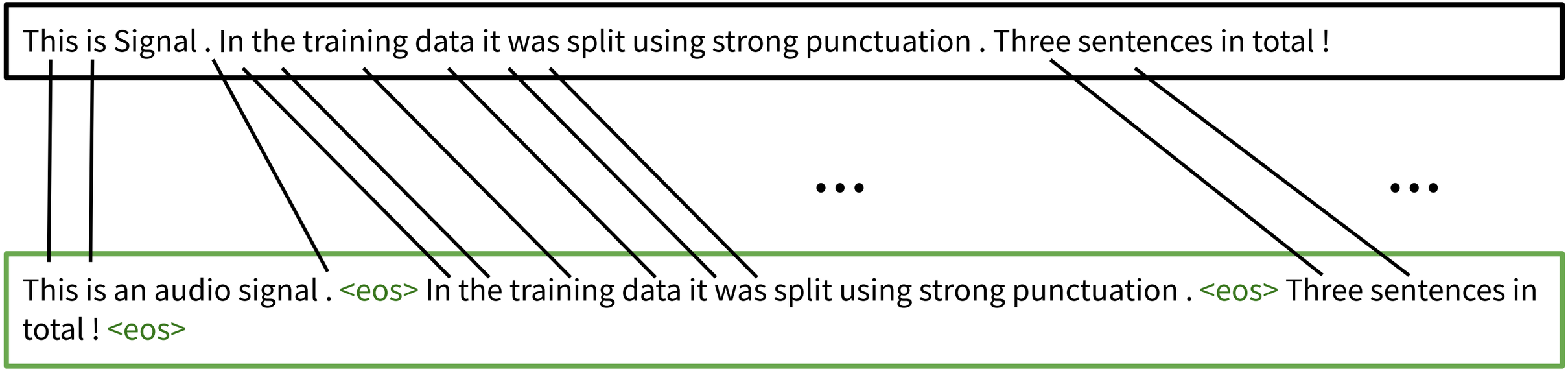
SLT output:

This is Signal . In the training data it was split using strong punctuation . Three sentences in total !

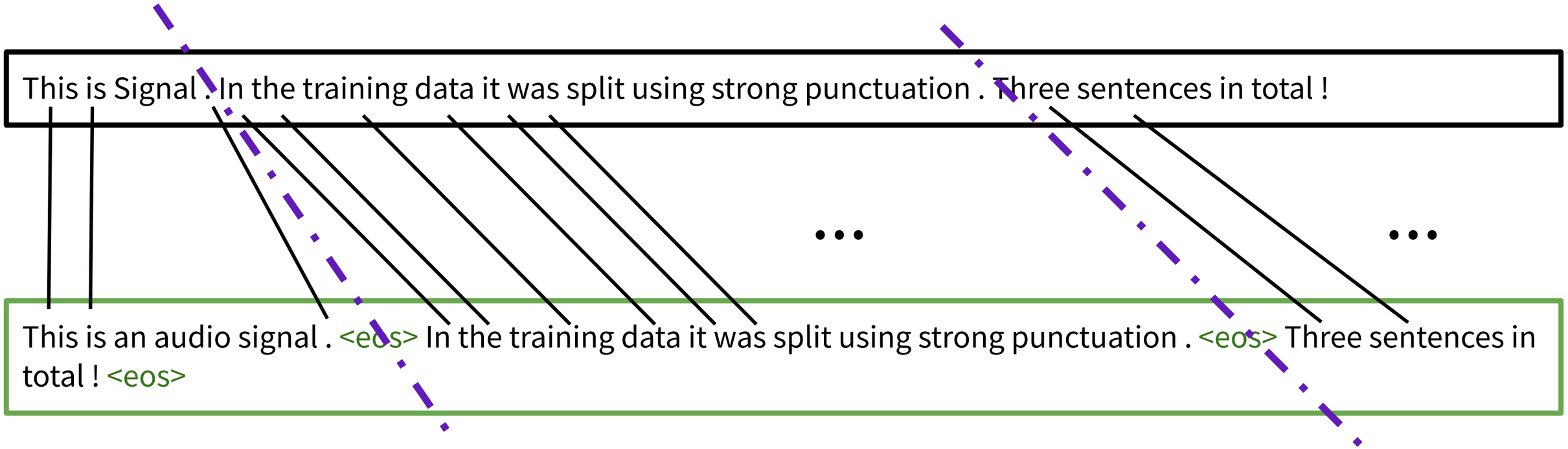
Reference sentences:

This is an audio signal . <eos> In the training data it was split using strong punctuation . <eos> Three sentences in total ! <eos>

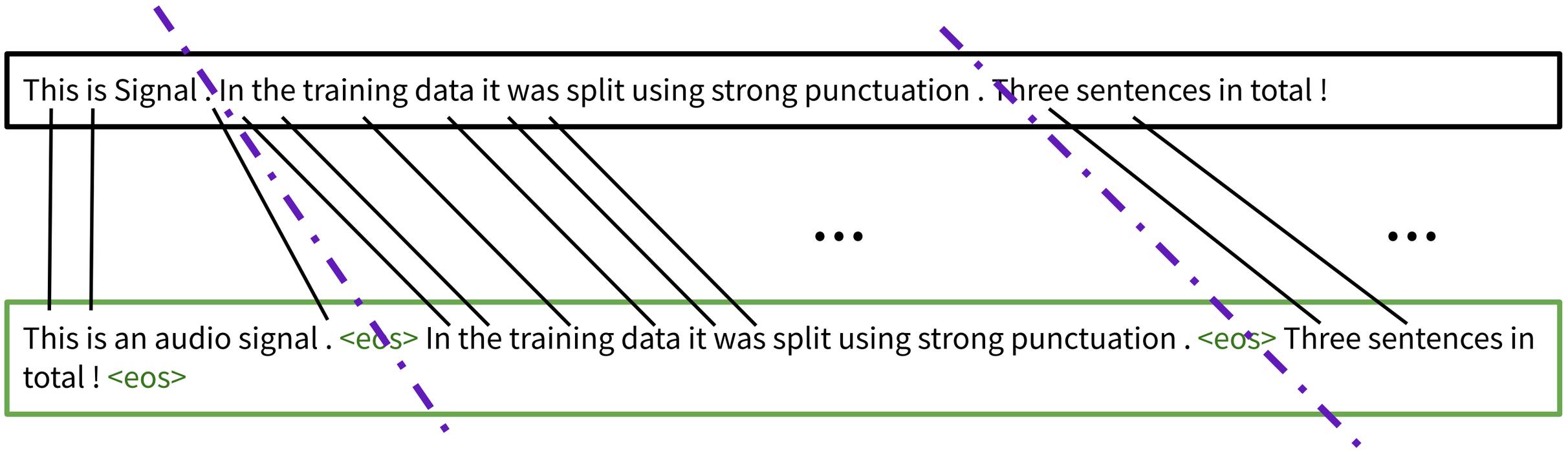
Automatic re-segmentation algorithm



Automatic re-segmentation algorithm



Automatic re-segmentation algorithm



Based on the word alignments and <eos>, the SLT output and reference are segmented.

Alignment and segmentation in one step using the Levenshtein distance (Matuzov et al., 2015).

New segments used to compute the automatic metrics.