Semantic Structural Decomposition for Neural Machine Translation

Elior Sulem, Omri Abend and Ari Rappoport

*SEM 2020

Penn
Cognitive Computation Group
האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

Semantic Structural Decomposition for Neural Machine Translation



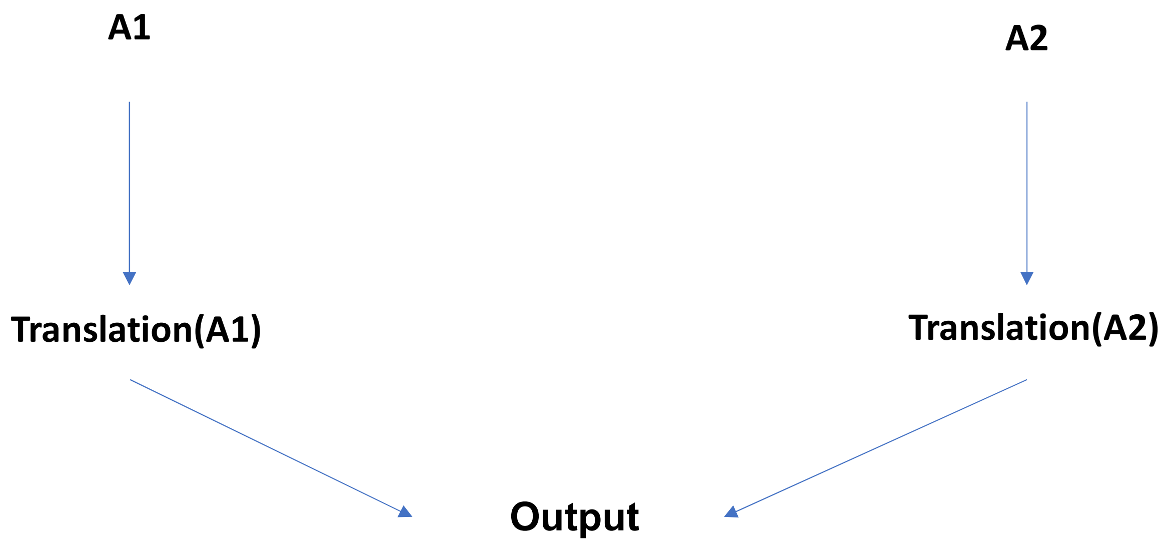
Elior Sulem   Omri Abend   Ari Rappoport

Penn
Cognitive Computation Group

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

Semantic Structural Decomposition for Neural Machine Translation

Input: Douglas Kidd of the National Association of Airline Passengers said he believes interference from the devices is genuine even if the risk is minimal.



Motivations

1. Improving the translation quality for long and complex sentences.

2. Allowing one-to-many translation, since structural simplicity can be an important component of the translation quality (Li and Nenkova, 2015).

Challenges

1. What is the right decomposition?

Input: Douglas Kidd of the National Association of Airline Passengers said he believes interference from the devices is genuine even if the risk is minimal.

2. What is the effect of the decomposition on the different aspects of the translation?

In particular: adequacy, fluency, coherence, simplicity.

In this talk

- We experiment with a Transformer model in English-to-French translation.
- We evaluate using large-scale crowdsourcing experiments.
- We show that the new approach improves the fluency when less data (5M sentence pairs) is available but it degrades the adequacy.
- Focusing on long sentences, we show that our approach significantly improves the fluency, while maintaining comparable adequacy.

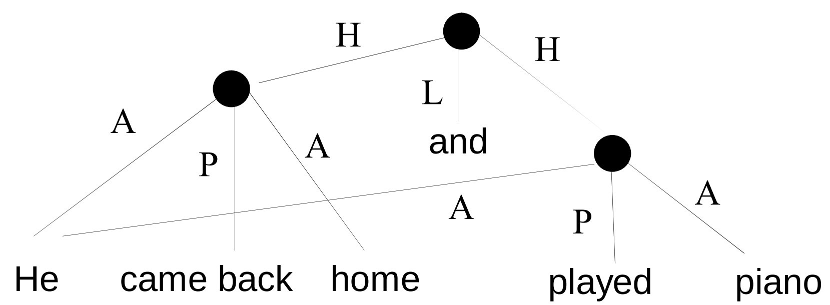
Outline

1. The UCCA Semantic Annotation
2. Direct Semantic Splitting
3. The Translation Pipeline
4. Experiments and Crowdsourcing Evaluation
5. Results
6. Comparison with Other Splitting Methods
7. Manual Analysis
8. Conclusion

The UCCA Semantic Annotation

Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013)

- Explicitly annotates semantic distinctions, abstracting away from syntax
- Semantic units are directly anchored in the text



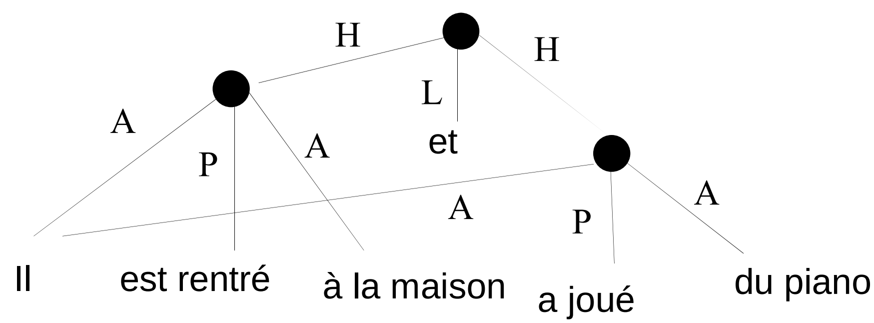
Parallel Scene (H)   Linker (L)

Participant (A)   Process (P)

The UCCA Semantic Annotation

Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013)

- Based on typological and cognitive theories (Dixon, 2010, 2012; Langacker, 2008)
- Stable across translations (Sulem, Abend and Rappoport, 2015)



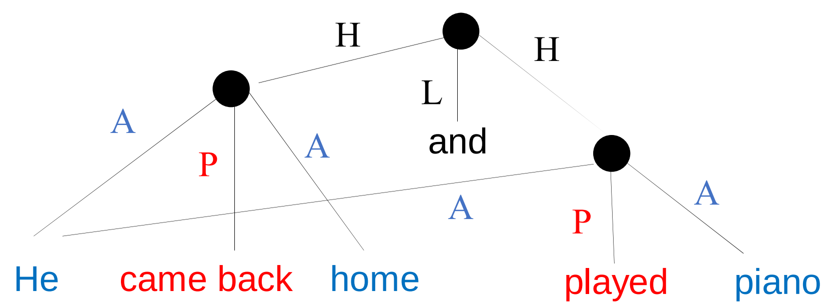
Parallel Scene (H) Linker (L)

Participant (A) Process (P)

The UCCA Semantic Annotation

Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013)

- Scenes are evoked by a Main Relation (Process/State)
- A Scene may contain one or several Participants.



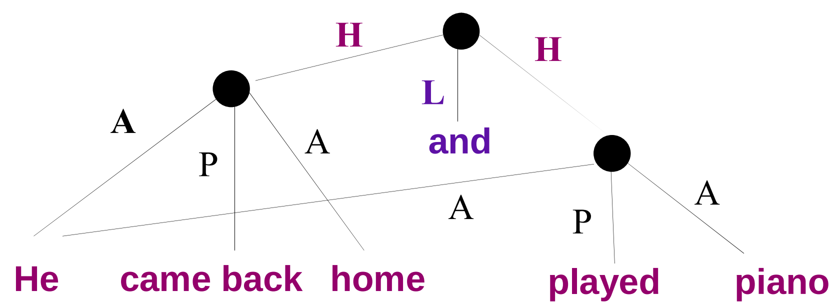
Parallel Scene (H)   Linker (L)

Participant (A)   Process (P)

The UCCA Semantic Annotation

Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013)

- A Scene can provide additional information on an established entity: Elaborator Scenes.
- A Scene may also be a Participant in another Scene: Participant Scene
- The default case (non-embedded): Parallel Scenes (H) that can be linked by a Linker.



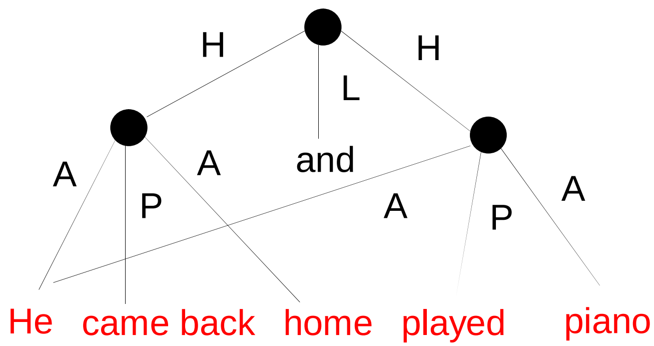
Parallel Scene (H) Linker (L)

Participant (A) Process (P)

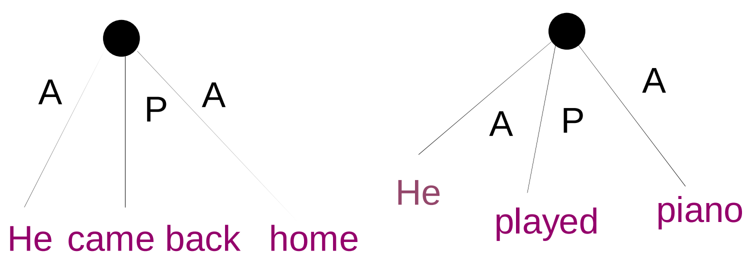
Direct Semantic Splitting

Parallel Scenes

DSS for Sentence Simplification (Sulem, Abend and Rappoport, 2018)



He came back home and played piano.

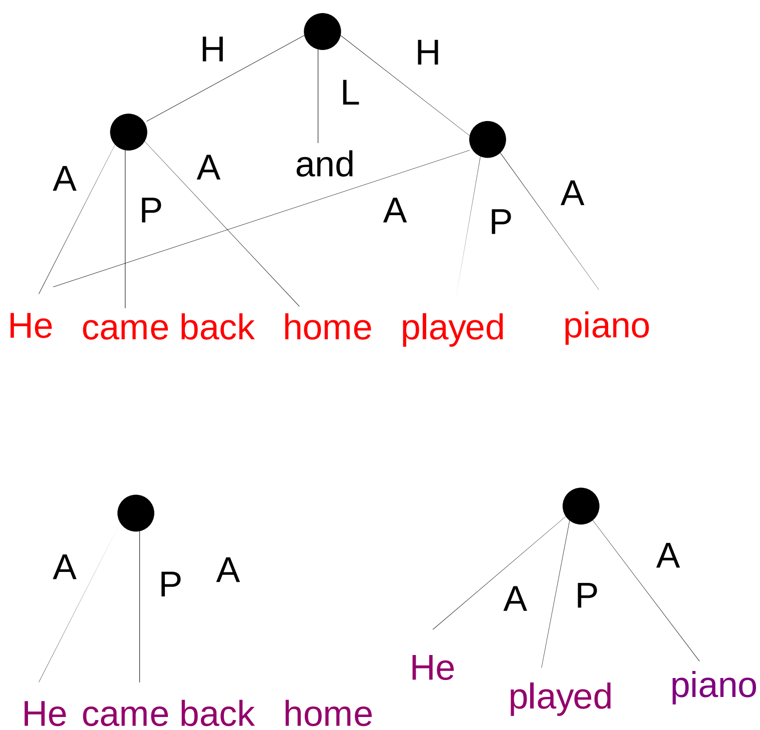


He came back home. He played piano.

Direct Semantic Splitting

Parallel Scenes

DSS for Sentence Simplification (Sulem, Abend and Rappoport, 2018)

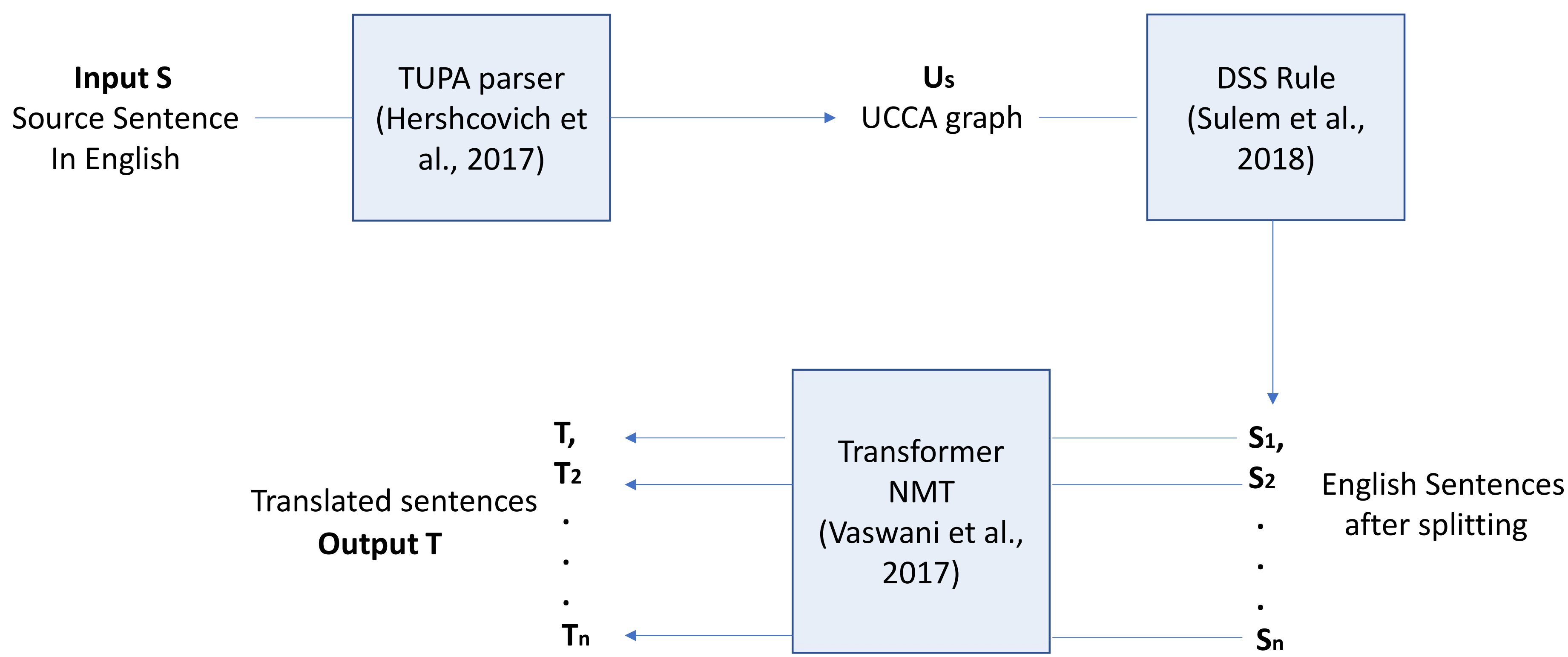


$$S \rightarrow Sc_1 \mid Sc_2 \mid \cdots \mid Sc_n$$

Input sentence

Input Scenes

SemSplit Transformer Pipeline



Experiments

- English-to French Translation
- Transformer system using OpenNMT-py implementation
- 2 settings: - FullTrain setting – original WMT training data (39 M sentence pairs)

  - LessTrain data (5M sentence pairs)
- Dev set: Newstest2013 - 3000 sentences
- Test set: Newstest2014 – 3003 sentences
- Comparison: SemSplit vs. Baseline (without splitting)

Evaluation

- BLEU is not correlated with meaning preservation when sentence splitting is involved (Sulem, Abend, Rappoport, 2018b)
- We use crowdsourcing evaluation with Amazon Mechanical Turk, following the protocol of Graham et al., (2016) for system-level comparison.
- Adequacy is evaluated my comparing the output to the reference sentence in French.
- Fluency is evaluated given the output, in a different evaluation experiment.
- Repetitions, good and bad translations are used to ensure the quality of the evaluation.

Results

System	Adequacy	Fluency
Baseline	**48.8**	**57.1**
SemSplit	40.0	43.5

Raw scores for adequacy and accuracy in the **FullTrain setting** (39M), considering **sentences of every length.**

System	Adequacy	Fluency
Baseline	**47.5**	42.5
SemSplit	39.8	**52.5**

Raw scores for adequacy and accuracy in the **LessTrain setting** (5M), considering **sentences of every length.**

- In the LessTrain setting, SemSplit significantly outperforms the baseline in terms of fluency ($p < 10^{-4}$).
- It is significantly surpassed by the baseline in terms of adequacy ($p < 10^{-4}$).

Results – Long Sentences

System	Adequacy	Fluency
Baseline	**48.0**	**49.4**
Our System	28.7	37.1

Raw scores for adequacy and accuracy in the **FullTrain setting** (39M), considering **sentences with more than 30 words.**

System	Adequacy	Fluency
Baseline	41.7	39.6
Our system	40.1	**52.1**

Raw scores for adequacy and accuracy in the **LessTrain setting** (5M), considering **sentences with more than 30 words.**

- In the LessTrain setting, SemSplit significantly outperforms the baseline in terms of fluency (52.1 vs. 39.6; p = 0.02) without significantly degrading adequacy (41.7 vs. 40.1; p = 0.46).

Other Splitting Methods

More UCCA-based rules:

- Adding a UCCA rule also separating Elaborator and Participant Scenes (embedded Scenes) did not improve fluency.

Neural-Based Sentence Splitting:

- We use Split and Rephrase models (Aharoni and Goldberg, 2018) trained on WEB-SPLIT (Narayan et al., 2017) or Wiki-Split (Botha et al., 2018)

- We obtain low quality splitting when transferring to the English WMT data, leading to low translation scores and supporting the use of corpus-independent sentence-splitting.

System	Adequacy	Fluency
Neural Wiki-Split	12.9	6.0
Neural WEB-SPLIT	4.1	5.1

Manual Analysis

We focus on LessTrain and on a sample of 150 sentences preserving the proportion of sentences of various length categories.

Scene preservation

- 298 out of the 450 annotated Scenes (66.22 %) are equally preserved by the SemSplit and Baseline.

- 20.89 % of the Scenes are better preserved by the Baseline
  10.67 % of the Scenes are better preserved by SemSplit

Sentence Cohesion

- 36% Baseline is better
- 59% Equally good
- 5% SemSplit is better

Example

Baseline:

Input: Douglas Kidd of the National Association of Airline Passengers said he believes interference from the devices is genuine even if the risk is minimal.

Output: Douglas., de l'Association nationale des compagnies aériennes, a déclaré qu' il considérait que l'ingérence avec les appareils était réelle, même si le risque était minimal.

Wrong translation

Literal Translation: Douglas., from the Association national of the companies airline, claimed that he believed the intervention with the devices was genuine, even if the risk was minimal.

Example

SemSplit:

Input: Douglas Kidd of the National Association of Airline Passengers said he believes interference from the devices is genuine even if the risk is minimal.

Output: Douglas., de l’Association nationale des compagnies aériennes, a déclaré qu’ il estimait que l’interférence avec les appareils était réelle.

Le risque est minimal.

Literal Translation: Douglas., from the Association national of the companies airline, claimed that he believed the interference with the devices was genuine.

The risk is minimal.

Conclusion

- We investigated the application of Semantic Structural Decomposition to NMT.
- An intermediary way between sentence segmentation used in MT and Text Simplification preprocessing (Štajner and Popović, 2018).
- Tradeoff between adequacy and fluency for a 5M English-French setting.
- Improvement of the fluency, while maintaining a comparable adequacy in the case of long sentences.

Future Work:

- Experimenting on additional language pairs
- Addressing sentence cohesion by inserting the linkage between the translated sentences.

Thank you

Elior Sulem

Data and code: https://github.com/eliorsulem/Semantic-Structural-Decomposition-for-NMT

eliors@seas.upenn.edu

https://www.cis.upenn.edu/~eliors/