

Cognitive insights into analysis of translationese combining product and process data

Ekaterina Lapshinova-Koltunski

November 6, 2024, Göteborg



Acknowledgement

Project B7: Translation as Rational Communication

SFB 1102 – Information Density and Linguistic Encoding (IDeAL)
funded by the Deutsche Forschungsgemeinschaft, Project ID 232722074

Elke
Teich



Christina
Pollkläsener



Heike
Przybyl



Maria
Kunilovksaya



Frances
Yung



Yuri
Bizzoni



Seed Funding Project: Data Triangulation in Translation Process Research

Experimental and corpus-based study of cognitive processing
funded by the University of Hildesheim

Silvana Deilen



Anna Lafrenz



Michael Carl



Overview

1 Introduction: Translationese

2 Methodology

3 Analyses

- Discourse connectives
- (Compound) nouns

4 Conclusion and Discussion

INTRODUCTION: TRANSLATIONESE

Translationese

- **Typical linguistic features** of translation (or interpreting) output compared to original written (or spoken) productions
[Baker, 1993, Toury, 1995, Halverson, 2003, Teich, 2003]
- Related to
 - the **process of translation** from a source language expression to a target language expression and
 - to the **translation product** when compared to originals in the same language as the target language [Chesterman, 2004]

Translationese Variation

- Register-, producer-, method- and mode-dependent
[Neumann, 2013, Lapshinova-Koltunski, 2022]
- e.g. differences across **translation and interpreting**
[Shlesinger and Ordan, 2012, Przybyl et al., 2022a, Kunilovskaya et al., 2023]:
 - Interpreting reinforces features of oral production (parataxis, general words, low surprisal verbs)
 - Translation reinforces features of written production (nominal categories and prepositions)
 - Less variation in interpreting regarding the word choice
[Yung et al., 2023, Przybyl et al., 2022a]
- Interpreting is an **extremely cognitively demanding** process
- Interpreting tends to exhibit **less explication and more simplification** effects

Explication and simplification

Explication/ Implicitation

According to [Klaudy and Károly, 2005] observed when

- a source language unit with a more general meaning is replaced by a translation unit with a more specific meaning
- a source text unit is unpacked and rendered as several units or
- the translator adds new meaningful elements into the target

Simplification

translations appearing linguistically simpler compared to original target language products, e.g. lower type-token ratio (less varied lexically), simpler syntax, etc. [Blum-Kulka and Levenston, 1983, Laviosa, 1998]

Translation vs. Interpreting

Aber → However = explication

DE: Original

Aber ich glaube, in einer Hinsicht gibt es Einigkeit: Der Reformelan in der Türkei scheint erlahmt zu sein...

EN: Translation

However, I believe that in one respect there is consensus: the pace of reform in Turkey seems to have slackened;

Aber → but ≠ explication

DE: Original

aber ich glaube in einer Hinsicht gibt Einigkeit der Reformelan in der Türkei scheint erlahmt zu sein...

EN: Interpreting

but euh one thing we agree on it seems that euh the impetus has gone out of Turkish reform processes

⇒ **Interpreting/translation vs. source language originals**

Explication and Implicitation

EN: Some of the most vulnerable countries of the world have contributed the least to climate change, **but** are bearing the brunt of it.

DE: Einige der Länder, die weltweit am wenigsten zum Klimawandel beigetragen, tragen **jedoch** die Hauptlast.
"Some of the countries that have contributed the least to climate change worldwide are **however** bearing the brunt."

Explication

Explication and Implicitation

EN: Some of the most vulnerable countries of the world have contributed the least to climate change, **but** are bearing the brunt of it.

DE: Einige der Länder, die weltweit am wenigsten zum Klimawandel beigetragen, tragen **jedoch** die Hauptlast.

"Some of the countries that have contributed the least to climate change worldwide are **however** bearing the brunt."

Explication

DE: Einige der am meisten gefährdeten Länder der Welt haben am wenigsten zum Klimawandel beigetragen, leiden **aber** dessen Folgen. "Some of the world's most vulnerable countries have contributed the least to climate change, **but** are suffering its consequences."

Equivalence

Explication and Implicitation

EN: Some of the most vulnerable countries of the world have contributed the least to climate change, **but** are bearing the brunt of it.

DE: Einige der Länder, die weltweit am wenigsten zum Klimawandel beigetragen, tragen **jedoch** die Hauptlast.

"Some of the countries that have contributed the least to climate change worldwide are **however** bearing the brunt."

Explication

DE: Einige der am meisten gefährdeten Länder der Welt haben am wenigsten zum Klimawandel beigetragen, leiden **aber** dessen Folgen. "Some of the world's most vulnerable countries have contributed the least to climate change, **but** are suffering its consequences."

Equivalence

DE: Einige der Länder, die den Klimawandel am härtesten zu spüren bekommen, haben nur sehr wenig dazu beigetragen. "Some of the countries that are feeling the effects of climate change the hardest **⊗** have contributed very little."

Implicitation

Explication and Simplification

Preference for **general over more specific** words

EN: Interpreting

we wanted to look at micro entities and that means **entities** which are really **very small** with very few **people** working for them minimum turnover EUH minimum profit **amount** which **are** very locally **active** ...

EN: Translation

We wanted to free micro-entities – and here we are talking about **companies** that are **particularly small**, with few **employees**, minimum turnover and profit **figures** and which effectively only **operate** in a regional, local area...

Explication and Simplification

Preference for **general over more specific** words

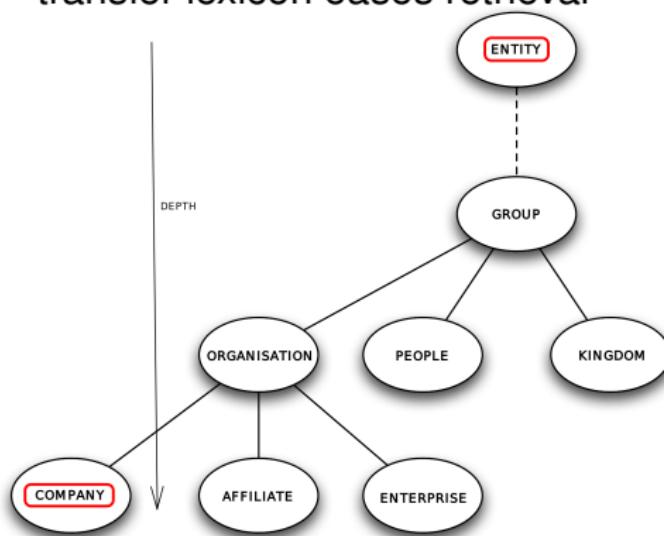
EN: Interpreting

we wanted to look at micro entities and that means **entities** which are really **very small** with very few **people** working for them minimum turnover EUH minimum profit **amount** which **are** very locally **active** ...

EN: Translation

We wanted to free micro-entities – and here we are talking about **companies** that are **particularly small**, with few **employees**, minimum turnover and profit **figures** and which effectively only **operate** in a regional, local area...

A smaller and more general transfer lexicon eases retrieval



Explication and Simplification

Example: Compound translation

EN: *Epilepsy syndrome*

DE: *Syndrom, Epilepsiesyndrom, Syndrom mit epileptischen Anfällen*

- **Implicitation:** *Syndrom* – "syndrome"
- **Equivalence:** *Epilepsiesyndrom* – "epilepsy syndrome"
- **Explication:** *Syndrom mit epileptischen Anfällen* – "syndrome with epileptic seizures"

Reserch Agenda

- Account of **translationese effects** in various translation products
- Discover **driving forces** behind these translationese effects

Hypotheses/Assumptions:

- ① **Explication** is a strategy of audience design as it helps to shape the content for the recipient and thus facilitates rational communication
- ② **Implicitation** and **Equivalence** also facilitate rational communication but for the sake of the producer (translator or interpreter) as they may reduce the effort on the translator's side
- ③ We expect differences in translationese effects across translation products (translation, interpreting) because of the differences in **producer conditions** (interpreting is cognitively extremely demanding) and also depending on the **target audience**

METHODOLOGY

Methods of Analysis

., 2009) or a combination of both (Green and Abutalebi, 2013). Similar language control areas have not yet been fully elucidated, but some studies have shown differences between the brain regions involved in language control in individuals with brain damage. For example, in the inferior parietal lobe (Abutalebi et al., 2004), in the left putamen (Abutalebi et al., 2005), and in the left insular cortex (Abutalebi et al., 2006).



Corpus-based Analysis

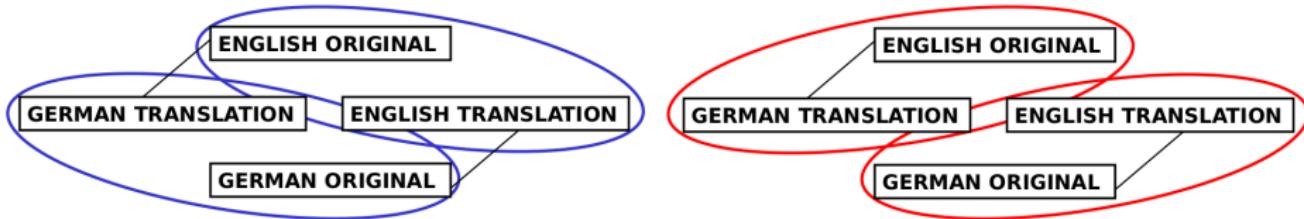
Analyse **actual translation relation** to explain translationese effects

Comparable corpora

- General translationese effects
- Link between translationese and efficient language use

Parallel corpora

- Source language triggers
- Source-language dependent vs. -independent translationese



Comparable and Parallel Corpora



Europarl-UdS

[Karakanta et al., 2018]

TR	EN-DE	3,994,453
	DE-EN	6,260,869
	ES-EN	3,162,915
ORG	DE	8,954,825
	EN	8,693,135
	ES	6,140,211

EPIC-UdS

[Przybyl et al., 2022b]

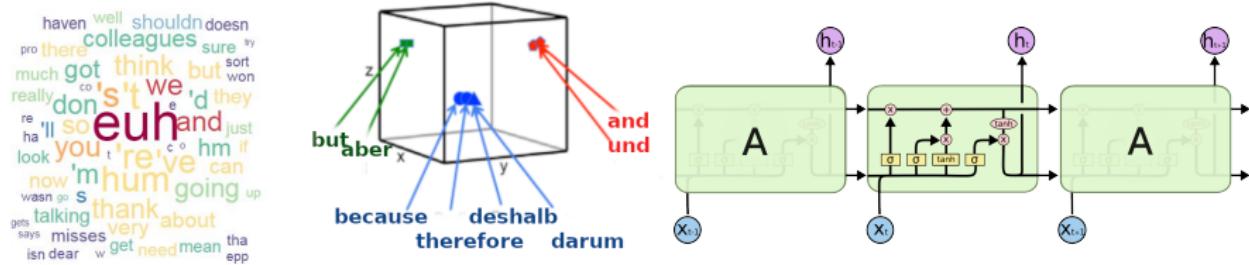
SI	EN-DE	57,622
	DE-EN	56,789
	ES-EN	52,737
ORG	DE	56,251
	EN	66,226
	ES	54,336



Corpora: Probabilistic Measures

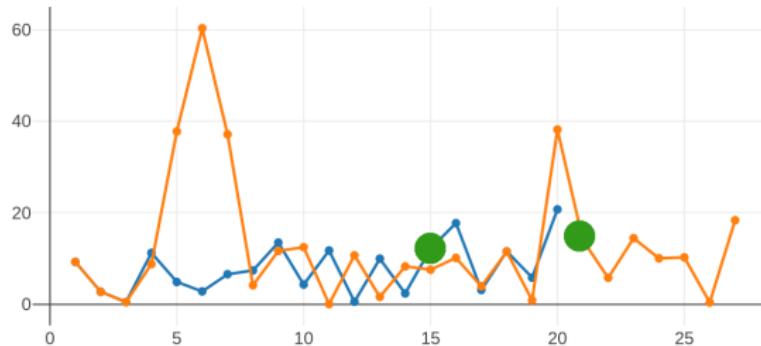
Information Theory [Shannon, 1948]

- Surprisal or (un)-predictability in context as a measure of cognitive effort [Hale, 2001, p. 4] and cognitive load [Teich et al., 2020] for translationese analysis
[Kunilovskaya et al., 2023, Lapshinova-Koltunski et al., 2022]
- Also: relative entropy (KLD) [Przybyl et al., 2022a], similarity measures [Lapshinova-Koltunski et al., 2021], perplexity [Bizzoni and Lapshinova-Koltunski, 2021]
- translation entropy [Wei, 2022, Schaeffer et al., 2016]



Corpora: Probabilistic Measures

- 1 Some of the countries that have contributed the least to climate change worldwide are **however** bearing the brunt.
- 2 Some of the world's most vulnerable countries have contributed the least to climate change, **but** are suffering its consequences.



Corpora: Probabilistic Measures

May help to explain Simplification effects

- Interpreters tend to produce more expected (low surprisal) lexical verbs and nouns than comparable original speakers
[Przybyl et al., 2023]
- For the same information content (measured in surprisal) in the source, interpreters produce lower surprisal output than translators [Kunilovskaya et al., 2023]

May help to explain Explicitation effects

[Pollkläsener et al., 2024] [Yung et al., 2023]

[Lapshinova-Koltunski et al., 2022]

Translation Process Analysis

Analyse translation process data to explain translationese effects in terms of cognitive effort:

- Typing **pauses** preceding the production of a translation (i.e., lag of time between last keystroke of preceding word and first keystroke of the current TL word)
- Total **reading time** of the translation unit – the sum total of all fixation durations on a particular area of interest (e.g., token) irrespective of when these occurred during the session:
 - Total reading time in the source
 - Total reading time in the target
- **logDur**: Time needed to type the translation – time from the first to the last keystroke).

Translation Process Data

- * Multilingual database CRITT TPR-DB [Carl et al., 2016]

- Recorded with Translog [Carl et al., 2015]
- English-German study [Nitzke, 2018]:
 - 6 EN sources
 - 24 translators (professional and students)
 - no participant worked with the same text sequence
 - no time/resource restrictions

Part	STseg	Stoken	Lemma	SGid	SGroup	SGx	SGnbr	STime	Cur	Ins	Del	Pause	Dur	Sdur	Prob1	Prob2	PoS	UpoS	TGroup	TC
P01	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P02	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P03	6	Because	—	97	Because	1	1	0	561	38	29	25801	10749	0	-3.1711	-7.7901	IN	SCONJ	Aufgrund	1
P04	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P05	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P06	6	Because	—	97	Because	1	1	0	561	3	0	3541	578	0	-3.1711	-7.7901	IN	SCONJ	Da	1
P07	6	Because	—	97	Because	1	1	0	561	0	35	4649	1575	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P08	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P09	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P10	6	Because	—	97	Because	0	1	0	561	0	0	0	0	0	-3.1711	-7.7901	IN	SCONJ	Well	1
P11	6	Because	—	97	Because	1	1	0	561	6	0	515	858	0	-3.1711	-7.7901	IN	SCONJ	Wegen	1
All																				

[Nitzke, 2018, p. 103]

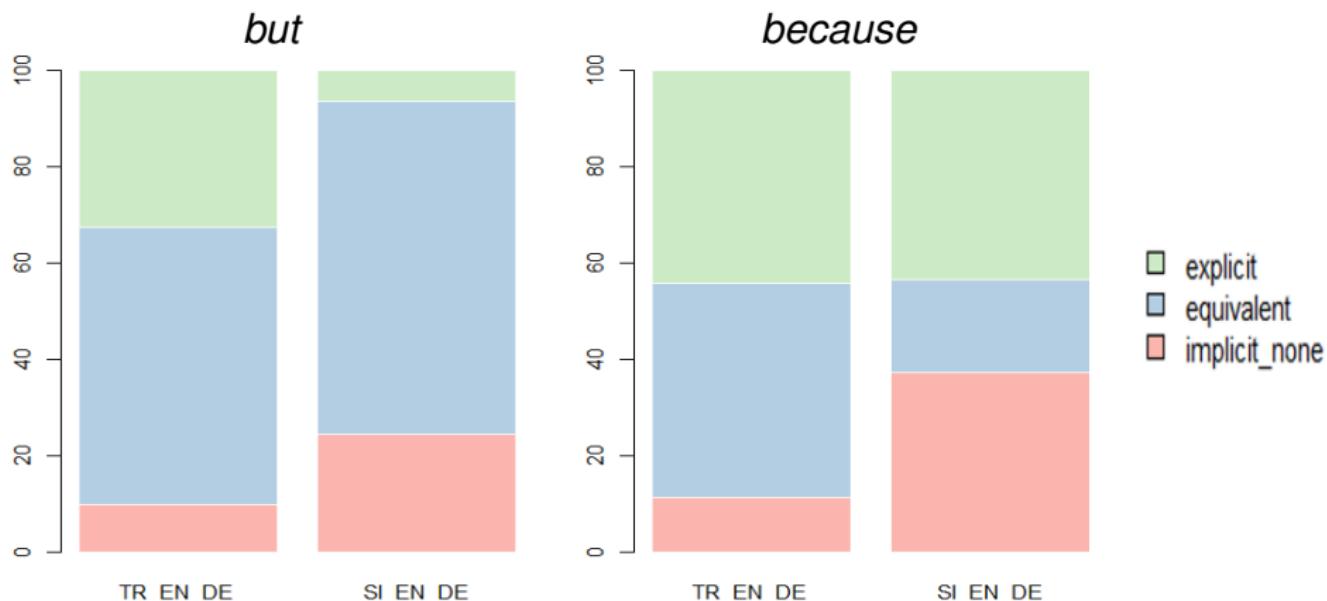
- * Additional experiments with student translators

ANALYSES

DISCOURSE CONNECTIVES

Parallel Corpora

Translation patterns: explicitation, equivalence, implication



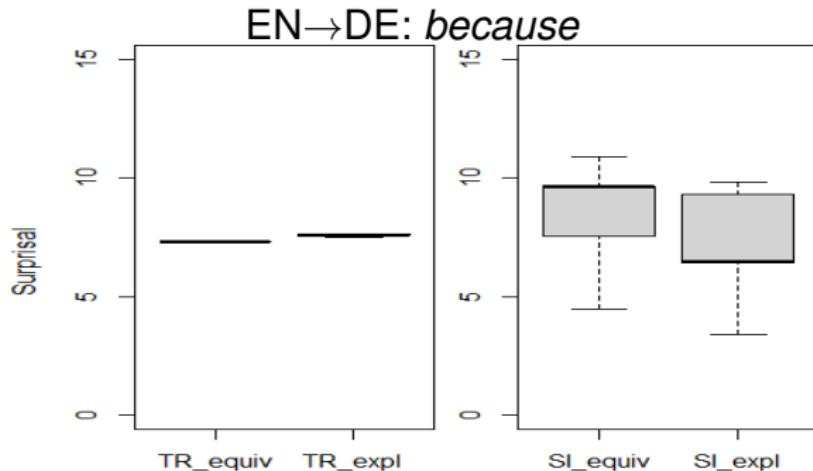
- While equivalence and implication prevail in interpreting,
- we do observe explicitation for some connectives

[Lapshinova-Koltunski et al., 2022]

Comparable Corpora

Surprisal from comparable corpora explains explicitation effects

Explicitation provides a bonus in cognitive processing effort for the recipient for certain connectives



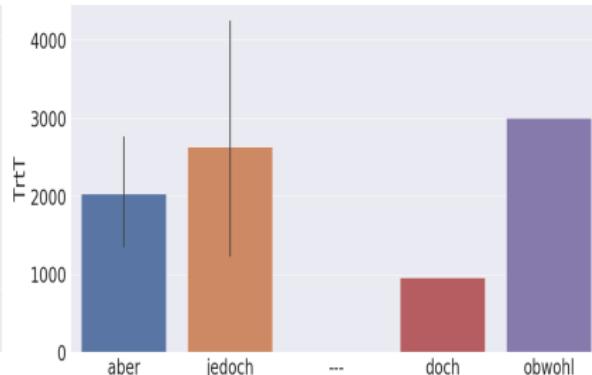
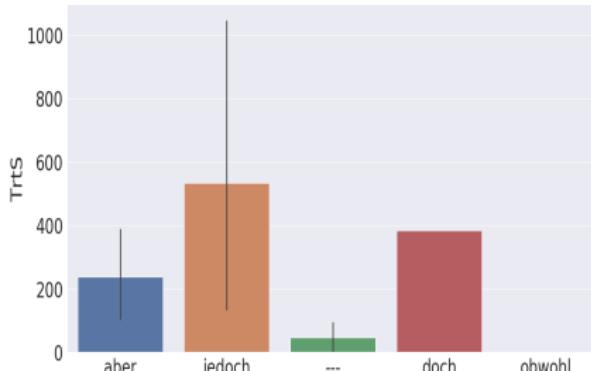
[Lapshinova-Koltunski et al., 2022]

Translation Process Data

Total reading time from translation process data

shows that explication costs more cognitive processing effort for the producer (translator)

Translation of connective *but* into DE in CRITT TPR-DB



Trt = sum of all fixation durations on a particular connective in the source (TrtS) or the target (TrtT)

[Lapshinova-Koltunski and Carl, 2022]

Parallel Corpora

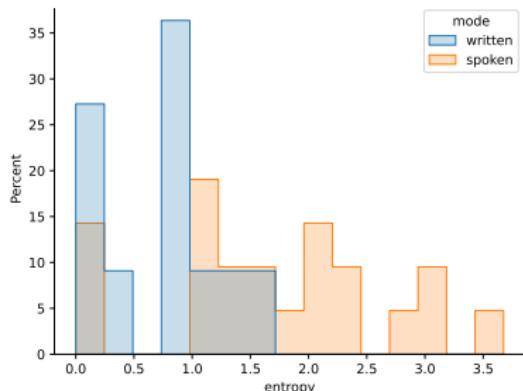
Translation entropy from parallel corpora helps to understand differences in explicitation strategies in translation and interpreting: interpreters use a smaller range of connectives to translate a wide range of source connectives than translators, especially in case of cognitively harder relations such as comparison

	spoken	written
<i>and</i>	2.74	1.49
<i>so</i>	2.82	3.39
<i>therefore</i>	2.47	2.37
<i>however</i>	1.30	2.24
<i>but</i>	1.77	2.32
<i>because</i>	2.92	1.61
<i>also</i>	1.58	1.86
<i>if</i>	1.79	1.99
<i>as</i>	1.28	2.64
<i>yet</i>	1.88	2.61

Parallel Corpora

Translation entropy from parallel corpora helps to understand differences in explicitation strategies in translation and interpreting: interpreters use a smaller range of connectives to translate a wide range of source connectives than translators, especially in case of cognitively harder relations such as comparison

	spoken	written
and	2.74	1.49
so	2.82	3.39
therefore	2.47	2.37
however	1.30	2.24
but	1.77	2.32
because	2.92	1.61
also	1.58	1.86
if	1.79	1.99
as	1.28	2.64
yet	1.88	2.61



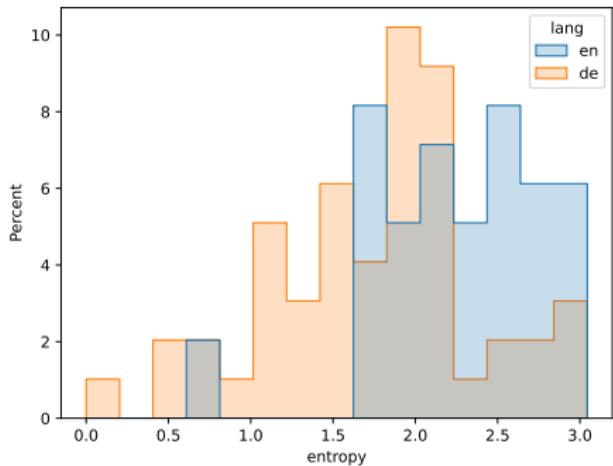
⇒ Interpreters reduce translation entropy

[Pollkläsener et al., 2024]

Parallel Corpora

Translation entropy from parallel corpora helps to detect cross-lingual differences in explicitation: English connectives correspond to a wider range of German ones, while German connectives more often have one dominating English translation

Entropy of the distribution of alignments of each connective in source:



- normal distribution for DE: some DE connective have more correspondences in EN, some less
- more varied for EN: correspond to a wide range of DE connectives

[Yung et al., 2023]

Summary on Connectives

Back to Hypotheses/Assumptions

- ① **Explication** is a strategy of audience design as it helps to shape the content for the recipient and thus facilitates rational communication:
 - ✓ but for certain connectives (relations) only
 - requires the highest cognitive effort on the producer's side
- ② **Implicitation** and **Equivalence** also facilitate rational communication but for the sake of the producer (translator or interpreter) as they may reduce the effort on the translator's side:
 - ✓ especially equivalence
- ③ We expect differences in translationese effects across translation products (translation, interpreting) because of the differences in **producer conditions** (interpreting is cognitively extremely demanding) and also depending on the **target audience**
 - ✓ interpreters reduce translation entropy by using a limited range of connective equivalents

(COMPOUND) NOUNS

Explication and Implicitation

Preference for **general over more specific** words

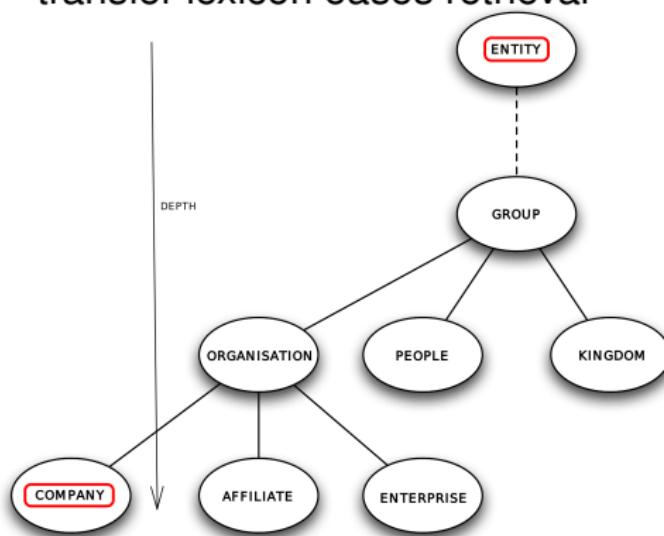
EN: Interpreting

we wanted to look at micro entities and that means **entities** which are really **very small** with very few **people** working for them minimum turnover EUH minimum profit **amount** which **are** very locally **active** ...

EN: Translation

We wanted to free micro-entities – and here we are talking about **companies** that are **particularly small**, with few **employees**, minimum turnover and profit **figures** and which effectively only **operate** in a regional, local area...

A smaller and more general transfer lexicon eases retrieval



Comparable Corpora

Preference for **general over more specific** words

EN: Interpreting

we wanted to look at micro entities and that means **entities** which are really **very small** with very few **people** working for them minimum turnover EUH minimum profit **amount** which **are** very locally **active** ...

EN: Translation

We wanted to free micro-entities – and here we are talking about **companies** that are **particularly small**, with few **employees**, minimum turnover and profit **figures** and which effectively only **operate** in a regional, local area...

Comparable Corpora

Preference for **general over more specific** words

EN: Interpreting

we wanted to look at micro entities and that means **entities** which are really **very small** with very few **people** working for them minimum turnover EUH minimum profit **amount** which **are** very locally **active** ...

EN: Translation

We wanted to free micro-entities – and here we are talking about **companies** that are **particularly small**, with few **employees**, minimum turnover and profit **figures** and which effectively only **operate** in a regional, local area...

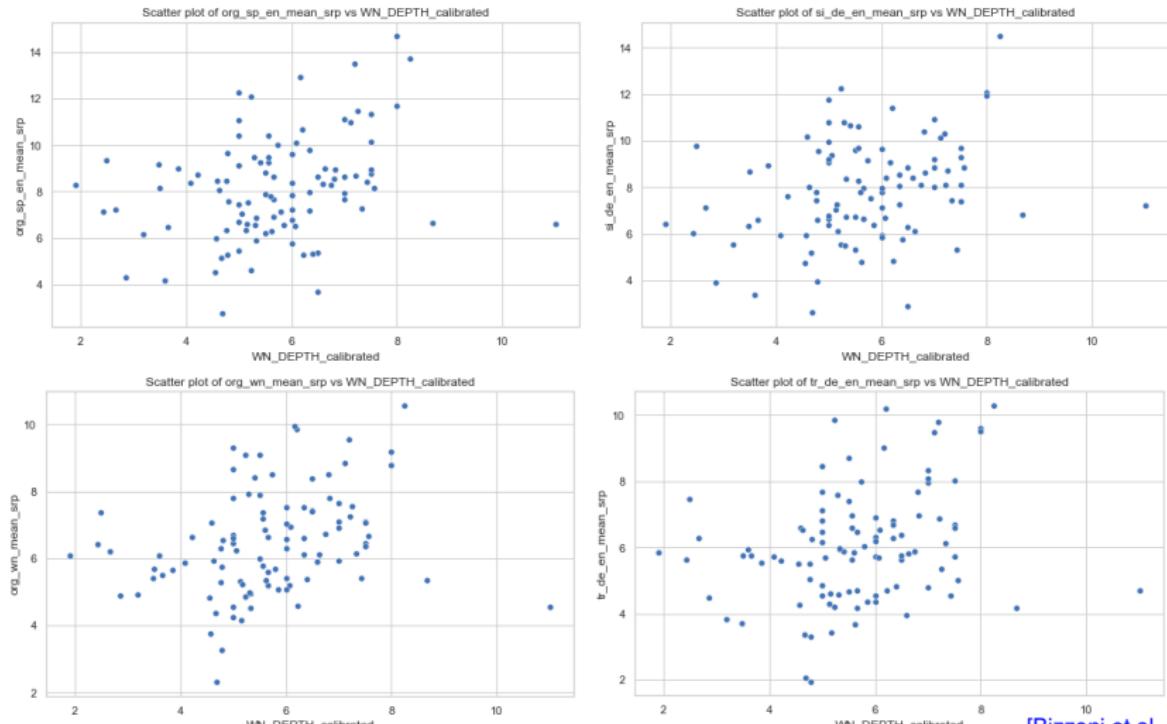
WordNet depth for 100 most frequent nouns in EN:

Spoken ORG	5.18
Interpreting	5.12
Written ORG	5.33
Translation	5.32

- Higher WordNet depth score in written than spoken
 - Written originals and translation employ more specific nouns
- ⇒ **Explication**
[Bizzoni et al., subm]

Comparable Corpora

Surprisal of words correlates with WordNet dept: the higher the score, the more cognitive effort is required to process



[Bizzoni et al., subm]

Comparable Corpora

Combined measure of surprisal and WordNet depth confirms observations on translationese: while written translation exaggerates features of written language production, interpreting exaggerates features of spoken language production in terms of semantic and contextual specificity

	SP	$\ln(SP)$
Spoken ORG	1 317	7.18
Interpreting	7	2.08
Written ORG	53 642	10.89
Translation	92 455	11.43

SP: multiply of the WordNet average depth and surprisal

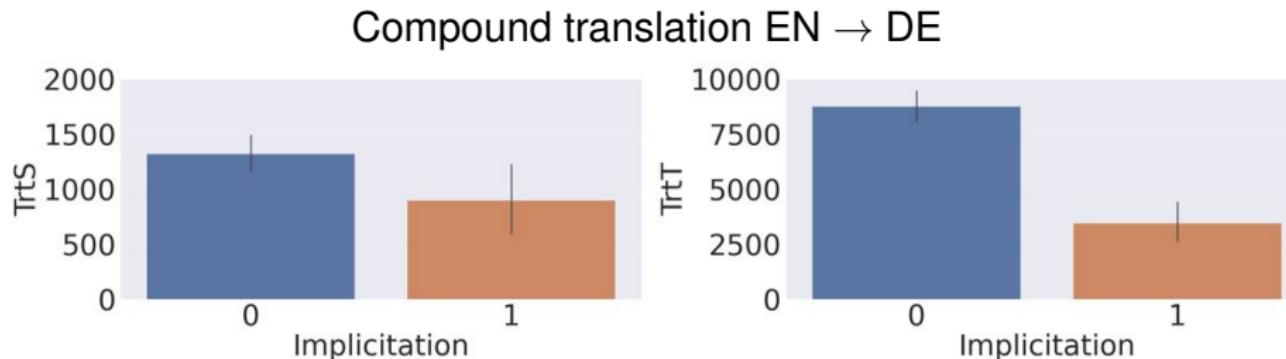
- Higher level of specificity for written (compared to spoken)
- Highest level of specificity for translation
- Lowest level of specificity for interpreting

[Bizzoni et al., subm]

Translation Process

Total reading time from translation process data

shows that implicitation provides a bonus in processing for the producer (translator)



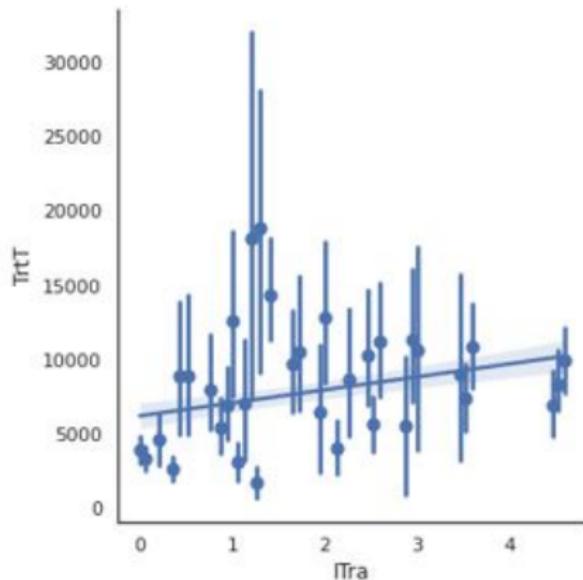
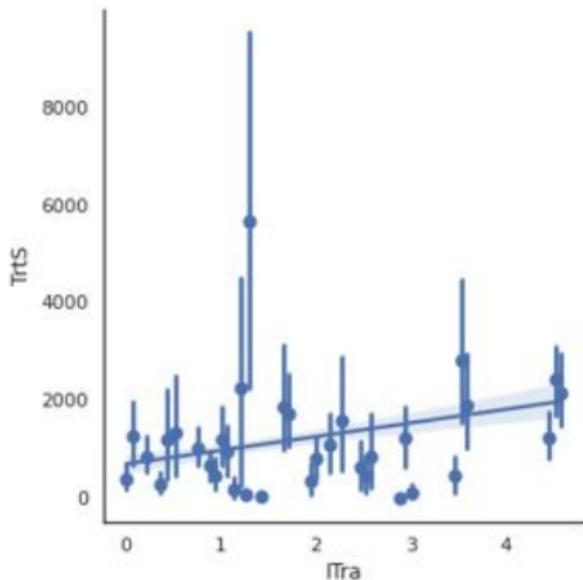
Example: *Epilepsy syndrome*

- 1. Implicitation:** *Syndrom* (syndrome)
- 0. Equivalence:** *Epilepsiesyndrom* (epilepsy syndrome)
- 0. Explicitation:** *Syndrom mit epileptischen Anfällen*
(syndrome with epileptic seizures)

[Deilen et al., 2023]

Translation Process

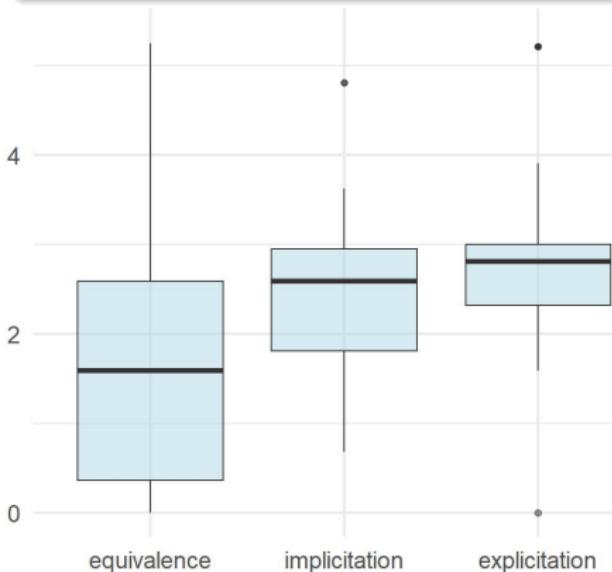
Total reading time (also further measures) correlates with word translation information: the higher the score, the more cognitive effort is required to process



[Deilen et al., 2023]

Translation Process

Word translation information indicates
that equivalence provides a processing bonus for the translator



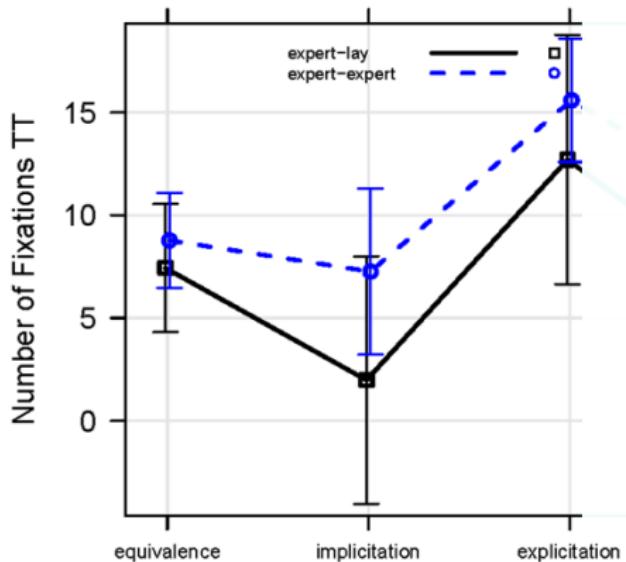
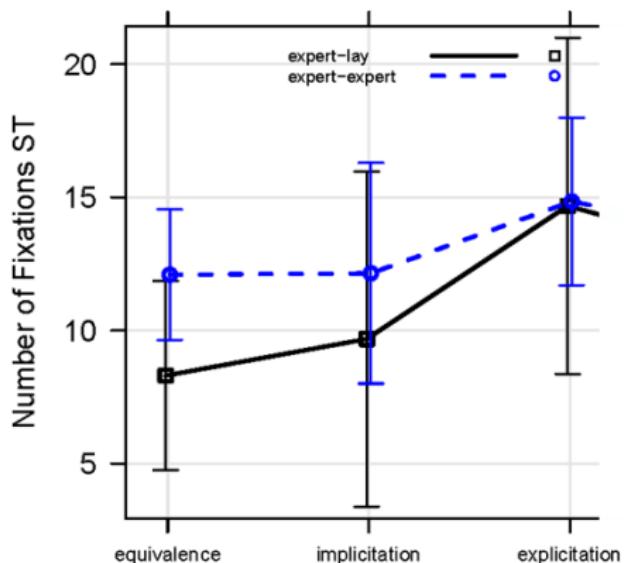
- Equivalence: lowest translation entropy
- Explicitation: highest translation entropy

[Deilen et al., prep]

Translation Process

Total reading time indicates

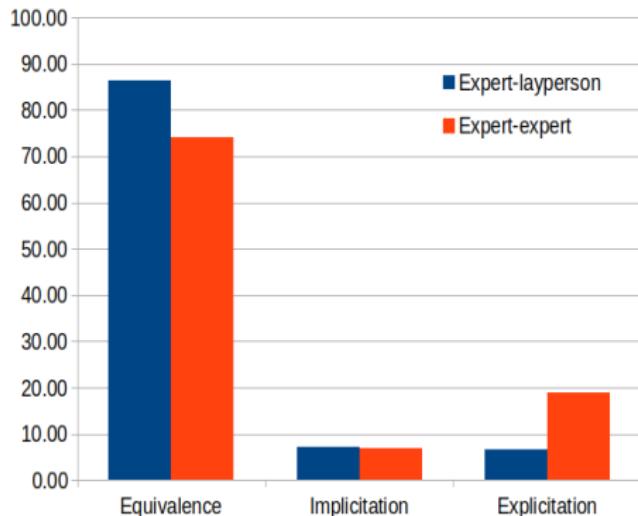
that explicitation causes a higher processing effort, however,
depending on the target audience of the text



[Deilen et al., prep]

Translation Process

Nevertheless explicitation occurs more frequently in translation of expert-to-expert texts



- Equivalence: more in expert-to-layperson than in expert-to-expert communication
- Explicitation: more in expert-to-expert than in expert-to-layperson communication

[Deilen et al., prep]

Summary on (Compound) Nouns

Back to Hypotheses/Assumptions

- ① **Explicitation** is a strategy of audience design as it helps to shape the content for the recipient and thus facilitates rational communication:
 - ✓ however depending on the target audience
 - still requires high cognitive effort on the translator's side
- ② **Implicitation** and **Equivalence** also facilitate rational communication but for the sake of the producer (translator or interpreter) as they may reduce the effort on the translator's side:
 - ✓ especially equivalence
- ③ We expect differences in translationese effects across translation products (translation, interpreting) because of the differences in **producer conditions** (interpreting is cognitively extremely demanding) and also depending on the **target audience**
 - ✓ implicitation is most evident in interpreting
 - ✓ explicitation depends on target audience too

Conclusion and Discussion

language control are not yet fully elucidated, Green and Abutalebi, 2013). Similar differences between the brain regions involved in language control have shown differences between the brain regions involved in language control (Abutalebi et al., 2004), in the left putamen (Abutalebi et al., 2004), and in the left insular cortex (Abutalebi et al., 2004).



Thank you!



Baker, M. (1993).

Corpus linguistics and translation studies: Implications and applications.

In Baker, M., Francis, G., and Tognini-Bonelli, E., editors, *Text and Technology: In honour of John Sinclair*, pages 233–250. John Benjamins, Amsterdam.



Bizzoni, Y., Fischer, S., Lapshinova-Koltunski, E., Pollkläsener, C., Przybyl, H., and Teich, E. (subm.).

Hierarchical semantic relations in translation and simultaneous interpreting.

Linguistics: An Interdisciplinary Journal of the Language Sciences.

Empirical Approaches to Effects of Context on Language Production and Comprehension.



Bizzoni, Y. and Lapshinova-Koltunski, E. (2021).

Measuring translationese across levels of expertise: Are professionals more surprising than students?

In Dobnik, S. and Øvreliid, L., editors, *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 53–63, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.



Blum-Kulka, S. and Levenston, E. A. (1983).

Universals of lexical simplification, pages 119–139.

Longman, London and New York.



Carl, M., Gutermuth, S., and Hansen-Schirra, S. (2015).

Post-editing machine translation: Efficiency, strategies, and revision processes in professional translation settings.

In Ferreira, A. and Schwieter, J. W., editors, *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting*, volume 115 of *Benjamins Translation Library*. John Benjamins.



Carl, M., Schaeffer, M., and Bangalore, S. (2016).

The CRITT translation process research database.

In Carl, M., Bangalore, S., and Schaeffer, M., editors, *New Directions in Empirical Translation Process Research*, New Frontiers in Translation Studies, pages 13–54. Springer Science+Business Media, Germany.



Chesterman, A. (2004).

Beyond the particular.

In Mauranen, A. and Kujamäki, P., editors, *Translation Universals: Do they exist?*, pages 33–49. John Benjamins, Amsterdam/Philadelphia.



Deilen, S., Lapshinova-Koltunski, E., and Carl, M. (2023).

Cognitive aspects of compound translation: Insights into the relation between implicitation and cognitive effort from a translation process perspective.

Ampersand, 11:100156.



Deilen, S., Lapshinova-Koltunski, E., and Lafrenz, A. (inprep).

Translation of compound terms: Cognitive insights.

Lebende Sprachen.



Hale, J. (2001).

A probabilistic Earley parser as a psycholinguistic model.

In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 159–166, Stroudsburg, PA. Association for Computational Linguistics.



Halverson, S. (2003).

The cognitive basis of translation universals.

Target, 15(2):197–241.



Karakanta, A., Vela, M., and Teich, E. (2018).

Europarl-UdS: Preserving metadata from parliamentary debates.

In Fišer, D., Eskevich, M., and de Jong, F., editors, *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).



Klaudy, K. and Károly, K. (2005).

Implication in translation: Empirical evidence for operational asymmetry in translation.

Across Languages and Cultures, 6:13–28.



Kunilovskaya, M., Przybyl, H., Lapshinova-Koltunski, E., and Teich, E. (2023).

Simultaneous interpreting as a noisy channel: How much information gets through.

In Mitkov, R. and Angelova, G., editors, *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 608–618, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.



Lapshinova-Koltunski, E. (2022).

Tracing normalisation and shining through in novice and professional translations with text classification.
In Granger, S. and Lefer, M.-A., editors, *Extending the Scope of Corpus-Based Translation Studies*, Bloomsbury Advances in Translation, pages 182—206. Bloomsbury Publishing, London, New York, Sydney, Delhi.



Lapshinova-Koltunski, E. and Carl, M. (2022).

Using translation process data to explore explicitation and implicitation through discourse connectives.

In *Proceedings of the 3rd Workshop on Computational Approaches to Discourse*, pages 42–47, Gyeongju, Republic of Korea and Online. International Conference on Computational Linguistics.



Lapshinova-Koltunski, E., Pollkläsener, C., and Przybyl, H. (2022).

Exploring Explicitation and Implicitation in Parallel Interpreting and Translation Corpora.

The Prague Bulletin of Mathematical Linguistics, 119:5–22.



Lapshinova-Koltunski, E., Przybyl, H., and Bizzoni, Y. (2021).

Tracing variation in discourse connectives in translation and interpreting through neural semantic spaces.

In *Proceedings of CODI at EMNLP-2021*, pages 134–142, Punta Cana and Online. ACL.



Laviosa, S. (1998).

Core patterns of lexical use in a comparable corpus of English narrative prose.

Meta, 43(4):557–570.



Neumann, S. (2013).

Contrastive register variation. A quantitative approach to the comparison of English and German.

Mouton de Gruyter, Berlin, Boston.



Nitzke, J. (2018).

Problem solving activities in post-editing and translation from scratch.

Number 12 in *Translation and Multilingual Natural Language Processing*. Language Science Press, Berlin.



Pollkläsener, C., Yung, F., and Lapshinova-Koltunski, E. (2024).

Capturing variation of discourse relations in english parallel data through automatic annotation and alignment.

Across Languages and Cultures.

special issue.



Przybyl, H., Karakanta, A., Menzel, K., and Teich, E. (2022a).

Exploring linguistic variation in mediated discourse: translation vs. interpreting.

In Kajzer-Wietrzny, M., Bernardini, S., Ferraresi, A., and Ivaska, I., editors, *Empirical investigations into the forms of mediated discourse at the European Parliament*, Translation and Multilingual Natural Language Processing, pages 191–218. Language Science Press, Berlin.



Przybyl, H., Lapshinova-Koltunski, E., Menzel, K., Fischer, S., and Teich, E. (2022b).

EPIC UdS - Creation and Applications of a Simultaneous Interpreting Corpus.

In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 1193–1200, Marseille, France. ELDA.



Przybyl, H., Lapshinova-Koltunski, E., and Teich, E. (2023).

Investigating surprisal in simultaneous interpreting.

In Kajzer-Wietrzny, M. and Chmiel, A., editors, *Using Corpora in Contrastive and Translation Studies Conference (7th Edition)*.



Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., and Carl, M. (2016).

Word Translation Entropy: Evidence of Early Target Language Activation During Reading for Translation.

In Carl, M., Bangalore, S., and Schaeffer, M., editors, *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*, pages 183–210. Springer, Cham.



Shannon, C. E. (1948).

A mathematical theory of communication.

Bell System Technical Journal, 27:379–423, 623–656.



Shlesinger, M. and Ordan, N. (2012).

More spoken or more translated?: Exploring a known unknown of simultaneous interpreting.

Target, 24(1):43–60.



Teich, E. (2003).

Cross-Linguistic Variation in System und Text. A Methodology for the Investigation of Translations and Comparable Texts. Mouton de Gruyter, Berlin.



Teich, E., Martínez Martínez, J., and Karakanta, A. (2020).

Translation, information theory and cognition.

In Alves, F. and Jakobsen, A. L., editors, *The Routledge Handbook of Translation and Cognition*, chapter 20. Routledge, London.



Toury, G. (1995).

Descriptive Translation Studies – and Beyond.

John Benjamins, Amsterdam.



Wei, Y. (2022).

Entropy as a measurement of cognitive load in translation.

In Carl, M., Yamada, M., and Zou, L., editors, *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Workshop 1: Empirical Translation Process Research)*, pages 75–86. Association for Machine Translation in the Americas.



Yung, F., Scholman, M., Lapshinova-Koltunski, E., Pollklaesener, C., and Demberg, V. (2023).

Investigating explicitation of discourse connectives in translation using automatic annotations.

In Stoyanchev, S., Joty, S., Schlangen, D., Dusek, O., Kennington, C., and Alikhani, M., editors, *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 21–30, Prague, Czechia. Association for Computational Linguistics.