Information Theory in Translation and Interpreting Studies

Elke Teich[a], Heike Przybyl[a], and Ekaterina Lapshinova-Koltunski[b], [a] Saarland University, Department of Language Science and Technology, Saarbrücken, Germany; and [b] University of Hildesheim, Department of Language and Information Science, Hildesheim, Germany

Introduction	**2**
Information: Surprisal, Entropy, Relative Entropy	**2**
Information-Based Models for Translation Analysis	**3**
Measuring Variation in Translation With Relative Entropy	3
Assessing Translationese Effects With Surprisal	4
Exploring Translation Difficulty by Entropy	5
Conclusion and Outlook	**6**
Acknowledgments	**6**
References	**7**

Key Points

- Information theory provides a formal model for describing and explaining language use in context, including mediated communication such as translation and interpreting.
- Translationese effects, such as simplification or explicitation, can be formally modeled with information theory (surprisal, entropy).
- Information theory provides a link between linguistic choice and cognitive processing.

Glossary

Bit (short for binary digit) the standard measure of the amount of information transmitted through a communication channel.
Surprisal the predictability of a linguistic unit in context (measured in bits of information).
Entropy the uncertainty about the outcome of an event (measured in bits of information).
Relative entropy the divergence between two distributional models A and B in bits of information.
Kullback-Leibler Divergence (KLD) the asymmetric variant of relative entropy.
Translation entropy the complexity of the translation search space for a given source language item (number and distribution of possible translations).
Translationese the statistically measurable linguistic effects of the process of translation in the translation/interpreting product.

Nomenclature

AvS average surprisal
KLD Kullback-Leibler Divergence
LM language model
SL source language
TL target language

Abstract

This article deals with the application of information theory in empirical, quantitative translation and interpreting studies. In particular, we focus on translationese, i.e., the specific linguistic choices in translation compared to original, non-mediated communication. We show how Shannon's information theory can be used to (i) explore translationese effects in corpora,

(ii) test specific translationese hypotheses (e.g., simplification, explicitation), and (iii) link up with explanations from rational communication, such as processing efficiency and cognitive resource limitations.

Introduction

Information theory plays an important role in modeling language use and variation within the wider framework of rational communication. According to rational communication, speakers modulate the level of information in their utterances by their linguistic choices, in order to convey their intended message successfully while keeping their cognitive effort at a reasonable level (cf. Crocker et al. (2016)).

In this article, we explain how information theory can be fruitfully applied to characterize translation as a specific type of language use within the wider conceptual framework of rational communication. Translations have been shown to exhibit specific properties compared to original texts, commonly referred to as *translationese* (cf. Baker (1995)). The terms *translationese* (for written translation) and *interpretese* (for spoken translation) describe the empirically observable effects of the cross-lingual mediation process in the translation/interpreting product, such as over-emphasis of target language norms or shining-through of the source language (Halverson, 2003; Olohan & Baker, 2000; Teich, 2003). Two basic, interrelated types of effects can be distinguished: one relating to the process of translation from a source language expression to a target language expression, the other relating to the translation product when compared to original texts in the same language as the target language (cf. Chesterman (2004)'s S- and T-Universals). The former are detectable in parallel corpora (source language texts aligned with their target language translations), the latter are analyzed with monolingually comparable corpora (target language translations and comparable original target language texts). Apart from source language shining-through and over-normalization (cf. Teich (2003)), other prominent translationese effects are:

- simplification, i.e., the language of translations appearing simplified compared to original target language texts (cf. Baker (1993) observing e.g., less varied lexis due to repeated choice of more general words);
- explicitation/implicitation, i.e., linguistic material that is optional such as complementizers, relativizers or discourse connectives, may be omitted or added in the translation process (Blum-Kulka, 1986) or more specific items are used in the TL text (Klaudy & Károly, 2005).

The article is organized as follows. We first provide a formal definition of information and explain three interrelated information-based notions that are widely used in the language sciences: surprisal, entropy and relative entropy (**Information: Surprisal, Entropy, Relative Entropy** section). Here, we emphasize that surprisal and related measures are indexed by a broad range of behavioral and neurophysiological indicators of cognitive activity, and correlate with linguistic encoding choices across linguistic levels. Thus, information theory provides a linking theory between language use and cognition. Furthermore, informationbased measures can be used as an analytic tool to capture linguistic variation, including translationese. In **Information-Based Models for Translation Analysis** section, we explain the utility of surprisal, entropy and relative entropy for corpus-based, quantitative translation and interpreting studies, specifically for the analysis and explanation of translationese, using examples from the domain of European Parliament translation and interpreting. **Conclusion and Outlook** section concludes with an appraisal of information theory for the analysis of translation as product and process and some open questions for a comprehensive account of translationese in the framework of rational communication.

Information: Surprisal, Entropy, Relative Entropy

It is commonly acknowledged that human language processing is to a substantial degree based on expectation (Crocker et al., 2016) and that language users rely on predictability in context (situational, linguistic) for efficient comprehension and production (Levy, 2008; Levy & Jaeger, 2007; Smith & Levy, 2013).

Predictability in context is formalized by the notion of information (Shannon (1948)), also known as *surprisal*. Surprisal measures the information content of a linguistic unit (e.g., a word) in context (e.g., *n* preceding words) as the number of bits needed for encoding that unit so that it can be successfully transmitted through the communication channel **Eq. (1)**.

$$Surprisal = -log_2\, p(unit|context) \quad [1]$$

For illustration, consider the following examples:

Example 1. Lily went to the library to get a book.

Example 2. Lily went to town to get a book.

In Example 1, the item "library" in the preceding context of "book" makes "book" highly predictable, so "book" has low surprisal (few bits of information). In Example 2 in contrast, "town to get a" does not strongly license a particular continuation, so surprisal on "book" is relatively high (many bits of information). Importantly, surprisal is proportional to cognitive effort in comprehension, as shown in behavioral experiments (e.g., reading times, fixation durations), or in brain activity (e.g., event-related potentials (ERP) such as the N400 measured in EEG studies) (cf. Frank et al., 2015). This means that a highly informational word has higher cognitive costs than a low information word.

While surprisal measures the information content or unexpectedness of a single specific event, *entropy* calculates the average surprisal across all possible events, representing the overall uncertainty of a system of choices. Consider the following example. There are three bowls with four apples in each of them: in bowl A there are four green apples, bowl B contains one red and three green apples and bowl C has two green and two red apples. With bowl A we can be 100% certain that we get a green apple, with bowl B there is a 75% chance that the apple we get is green and a 25% chance that it is red, and with bowl C green and red are equally probable (50%). Comparing the three cases, in A the outcome is absolutely certain—therefore entropy is zero. C has the highest uncertainty with both outcomes being equally likely (entropy is 1), and entropy in B is 0.81.

Going back to Examples 1 and 2, all other things being equal, the entropy H(*book*) in Example 1 is lower than in Example 2 because it is a very likely choice in the given context and not many other options are available, while in Example 2 there are more options and no clearly preferred one. If we imagine a guessing game giving these two contexts in which people are asked to complete the sentences, the target "book" will be produced faster (and with less cognitive effort) in 1 compared to any target in the context of 2 (see e.g., Jaeger (2010) for a study on the use of complementizers modeled with entropy).

Surprisal and entropy have been extensively used to investigate human language processing. Their common core is predictability in context (cf. **Eq. (1)**: $p(unit|context)$). If we want to model language use on this basis, we need to obtain probability estimations based on usage data. While in experimentally oriented studies, a common method are cloze probabilities obtained from sentence completion tasks, computational language models (LMs), which are based on large text collections, are increasingly being used to estimate surprisal (Michaelov et al., 2023). LMs range from statistical to neuro-computational models with varying degrees of learning (see Hedderich et al. (2021) or Oh and Schuler (2023) for useful overviews of the present state-of-the-art).

In the work presented in this article, we focus on count-based n-gram language models, i.e., models that estimate probabilities of words in the context of their (typically three) preceding words. Importantly, we can use LMs to model language varieties, including translation/interpreting and apply them to detect translationese effects. One method that has been used extensively for exploring linguistic variation in general (e.g., register variation) is *relative entropy*. Relative entropy measures the number of additional bits that are needed to model a language variety when a non-optimal model is used. For example, when we use a model A of news texts and apply it to a corpus of scientific texts B, clearly A will not exactly match B. Relative entropy formalizes the mismatch or divergence between two models A and B in bits of information. The greater the divergence between A and B, the more linguistically different the two varieties are (as represented by the models). The asymmetric variant of relative entropy, Kullback-Leibler Divergence (short: KLD; Kullback and Leibler (1951)) captures the two perspectives inherent in a contrast, i.e., we obtain both the typical features of A when modeled on the basis of B and of B when modeled on the basis of A. Applied to translation, we can thus explore the specific features of translation when modeled on the basis of originally produced texts as well as the typical features of originally produced texts when modeled on the basis of translation.

Information-Based Models for Translation Analysis

We now illustrate the application of the three information-based measures described above to the analysis of translation/interpreting with a special focus on translationese/interpretese. The underlying data for all examples shown are taken from European Parliament sessions (Karakanta et al., 2018; Przybyl, Lapshinova-Koltunski et al., 2022).

Measuring Variation in Translation With Relative Entropy

As explained above, KLD is an efficient method for exploring differences between linguistic varieties, as originally proposed by Fankhauser et al. (2014) accounting for variation in the English Brown corpus. Similarly, KLD can be efficiently employed to analyze variation in translation. For instance, it can be used to detect typical characteristics of interpreting (vs. translation) compared to original productions in the same language as the target language (see e.g., Karakanta et al. (2021) and Przybyl, Karakanta, et al. (2022))

For instance, consider **Fig. 1**, a visualization of words' relative frequency (color: red = high relative frequency, blue = low relative frequency) and distinctivity by KLD (size: big = high KLD, small = low KLD) comparing interpreting versus spoken target language originals (left) and translation versus written target language originals (right). Only words that are significant are shown. Overall, interpreting is found to be characterized mainly by spoken features (discourse markers, fillers, intensifiers etc.) and general verbs whereas typical features of translations are nominal categories and prepositions. This confirms previous frequency-based results from e.g., Shlesinger and Ordan (2012) stating that interpreting reinforces features of oral production and that translation emphasizes features of written production.

The advantage of the KLD approach over purely frequencybased approaches commonly pursued in corpus-based translation studies, is that it is data-driven: linguistic features do not have to be pre-determined and tested post-hoc for significance but they are dynamically "discovered", where only significant features are inherently considered.

Fig. 1 Visualization of variation in translation output (source language: German). Left: interpreting (vs. target language originals), right: translation (vs. target language originals). Color of items denotes relative frequency, size represents pointwise KLD.

Assessing Translationese Effects With Surprisal

Surprisal can be straightforwardly used to investigate the simplification hypothesis by comparing the average surprisal (AvS) of segments, i.e., aligned source language units and their target language equivalents, where a lower AvS in the TL is taken as an indicator of simplification. This was demonstrated in a corpus analysis based on English and German original and interpreting data with English and German as source as well as target language by Kunilovskaya et al. (2023). A simplification effect was found for interpreting for both English and German as target languages, but no consistent effect for translation (simplification into English, no simplification into German). While the linear regression analysis indicates a positive correlation between sources and targets in terms of AvS in general, for the same level of AvS in the source, interpreters produce lower surprisal outputs (on average) than translators (see Fig. 2). Thus, interpreting output appears more simplified compared to translation output, i.e., interpreters choose translation variants that are more expected and thus easier to produce.

Surprisal may also be used to accoount for explicitation. For instance, Lapshinova-Koltunski et al. (2022) describe various strategies of translating discourse connectives (English-German, German-English) linking them to cognitive effort as manifested by surprisal. A comparison of surprisal values for the same connectives in translation and interpreting, as well as in comparable

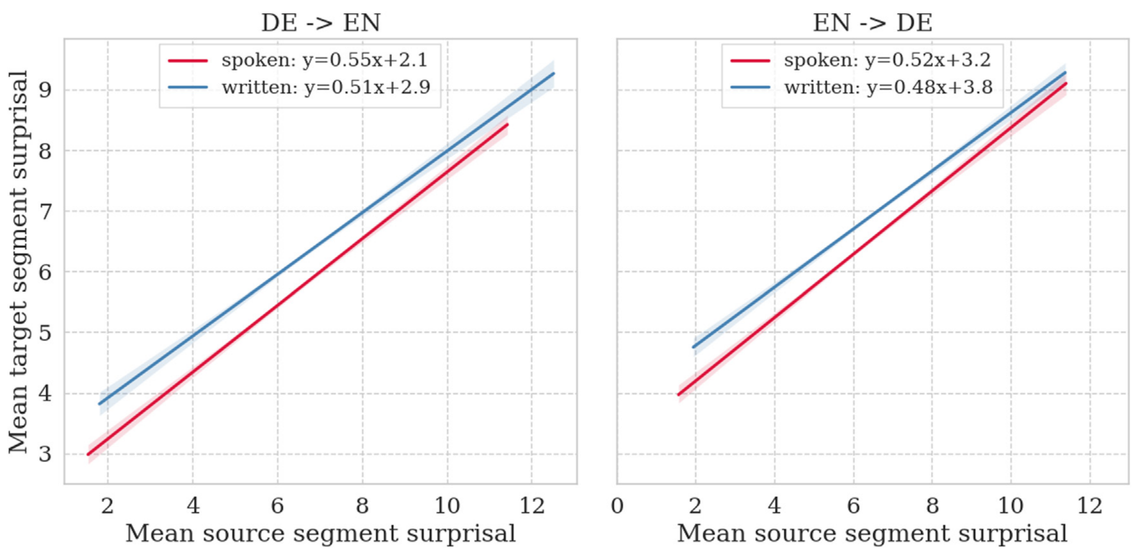

Fig. 2 Linear regression based on AvS of aligned source and target segments by translation direction (DE-EN, EN-DE) and crosslingual mediation type (spoken/interpreting, written/translation). Kunilovskaya et al., 2023, p. 612.

originals in both English and German, reveals that implicitation (a connective is omitted or a more general one is used) as well as direct equivalence frequently occurs in interpreting while translation shows a tendency toward explicitation. See the translation in Example 4 with "however" emphasizing the contrast more than the original German "aber" (Example 3), while in interpreting (Example 5), the more generic "but" is used. Other things being equal, more general words (like "but") are overall less surprising, i.e., incur a lower processing cost, than more specific words (like "however") which have a higher processing cost.

Example 3. **Aber** ich glaube, in einer Hinsicht gibt es Einigkeit ...

Example 4. **However** I believe that in one respect there is consensus ... (translation)

Example 5. **but** euh one thing we agree on ... (interpreting)

Choosing a generic variant over a more specific one (here: "but" vs. "however") seems to be a producer-centric mechanism, facilitating processing for interpreters, while choosing a more specific variant over a more generic one may be a means of audience design, potentially making a message easier for the recipient.

Similar trends emerge in other studies looking at other kinds of linguistic features. For example, focusing on spoken data (originals vs. interpreting), again in the language pair English-German, Przybyl, Lapshinova-Koltunski, and Teich (2023) show that interpreters tend to produce more expected (low surprizal) lexical verbs than comparable original speakers as well as more expected nouns in English Again, this may reflect a mechanism to reduce the cognitive load for the interpreter.

Surprisal can also be used to show the impact of specific production conditions in interpreting, such as source speech delivery type (imprompu vs. read-out) or source speech delivery rate, which may influence interpreters' ability to fully transfer information contained in the source to the target, as shown in Kunilovskaya et al. (2023) on German and English data (both translation directions). While no effect regarding delivery type was found, high delivery rate indeed seems to lead to lower surprisal in interpreting on average. Once again, this is a highly plausible effect of the specific coping mechanisms of interpreters under high cognitive load.

Exploring Translation Difficulty by Entropy

Technically, entropy measures the size of the search space of possible values of a random variable and its associated probabilities (Manning & Schütze, 2003, p. 63). Applied to translation, we arë interested in translation entropy (Schaeffer et al. (2016); Martínez Martínez and Teich (2017); Teich et al. (2020)), i.e., the search space of possible translations for a given source language item (e.g., a word), as an indicator of translation difficulty **Eq. (2)**.

$$H(T) = -\sum_{t \in T} p(t) \log_2 p(t) \qquad [2]$$

where T stands for the translation space, i.e., the set of all possible translations t for a given source text unit as found in a parallel corpus.

In case of no default equivalent fitting a source item, various translation options for that item compete for selection during translation production (Hantsch et al., 2005). From a processing perspective, translation entropy correlates with cognitive effort, i.e., the higher the translation entropy, the more cognitive effort is required for selection (cf. Wei (2022)). In translation studies, this effort is commonly estimated experimentally by fixation durations during reading, where first fixations are connected to retrieval. Short first fixation indicates easy retrieval, while a longer fixation points to higher processing effort or difficulty (Carl & Schaeffer, 2017; Schaeffer et al., 2016).

Estimated on a parallel corpus, translation entropy can give us an idea of the overall translation difficulty of particular texts, text types and registers. For example, it may turn out that Parliament speeches exhibit a lower overall translation entropy than newspaper texts and should therefore be easier to translate. Translation entropy can also be used as an indicator of translation difficulty in a specific context, e.g., the translation difficulty of a specific word. Consider the examples shown in **Fig. 3**. The x-axis shows the individual translation options, the y-axis plots relative frequency. The examples show a fairly low entropy H of 0.157 for translation of "council" due to few alternative choices and one clearly dominant choice ("Rat"). In contrast, while "Verletzung" is the dominant translational choice for "breach", there are many other options that are fairly evenly distributed, together resulting in a fairly high entropy of 0.635.

Thus, we would predict from the examples in **Fig. 3** that the translation of "council" is easier than the translation of "breach". This makes much sense intuitively since "council" in the context of the European Parliament has a very specific meaning and translators/interpreters will be primed for the common translation as "Rat", while "breach" is a more general word which can be vague or ambiguous, its meaning (and translation) depending much more on the (linguistic) context and therefore requiring more instantaneous effort.

Translation entropy is also linked to explicitation. Returning to the example of connectives, the more translation equivalents a connective has, the more ambiguous and less specific and therefore less explicit it is (see Klaudy & Károly, 2005, p. 15). Comparing translation and interpreting, Pollklasener et al. (2024) show that translation entropy of the same discourse connectives differs in translated and interpreted corpus data. Interpreters reduce their cognitive effort by using fewer different target equivalents indicated by a lower translation entropy. The same connectives (especially those signaling more complex relations) have a higher translation entropy in translation than in interpreting. Here, explicitation with respect to connectives is measured by the entropy of the sense

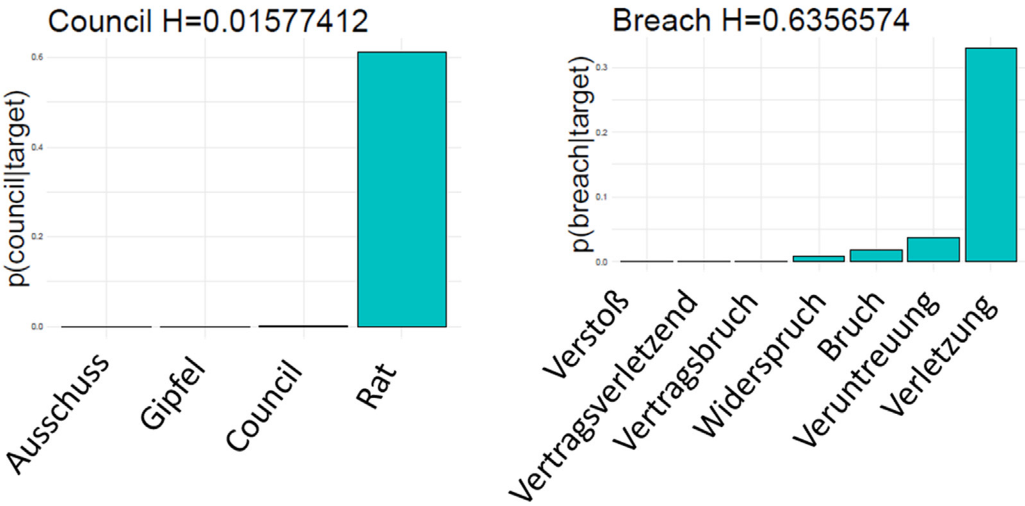

Fig. 3 Translation entropy from English into German for "council" (left) and "breach" (right).

distribution of a connective in relation to the entropy of all explicitly signaled relations in the corresponding corpus (Yung et al. (2023)).

Conclusion and Outlook

Information theory enjoys two properties relevant for application in the language sciences. First, it provides a linking theory between language use and cognition; and second, it opens the way for a unified treatment of otherwise mostly disjoint enterprises (e.g., research on comprehension vs. production, language learning, language evolution, translation) in the wider theoretical framework of rational communication (cf. **Information: Surprisal, Entropy, Relative Entropy** section above).

In rational communication, we assume that interlocutors strive for accurate message transmission while optimizing their cognitive effort. This is true for cross-lingually mediated communication, too: while translators want to deliver a loss-less output, they have to be efficient in their task (especially in simultaneous interpreting). Adopting the conceptual framework of rational communication with information theory as its core allows us to connect insights from product-oriented, corpus-based studies on translationese (such as Olohan and Baker (2000); Teich (2003); Bernardini et al. (2016); Lapshinova-Koltunski (2021)) with results from process-oriented, corpus-based and experimental studies on the specific cognitive factors at play in translation (e.g., Plevoets and Defrancq (2016), Plevoets et al. (2020); Chmiel et al. (2021, 2023); Hodzik (2023)).

While we have focused on predictability in this article, there are other cognitive factors involved in language processing. One of them is working memory. Recent studies in language typology suggest that certain properties of language (e.g., word order) are optimized for a trade-off between memory and surprisal (see e.g., Hahn et al. (2020, 2021)). Clearly, working memory is a crucial factor for translators, too, especially when under time pressure as in simultaneous interpreting. Thus, managing memory capacity may well explain particular aspects of interpretese, such as omissions, filled pauses, or specific lexical or grammatical choices (see e.g., Pollklasener et al. (2025)). An information-theoretic framework can be extended by a memory component fairly organically, using the same currency—information in bits—to model the interplay of memory and surprisal. While a promising avenue toward more comprehensive explanations of cross-lingually mediated communication, there are a number of yet untackled methodological issues, including the co-modeling of overlapping comprehension (source language) and production (target language).

Finally, if the product of translation/interpreting is so different from original text/speech, what are the effects of simplification, explicitation etc. on the recipients of cross-lingual mediation? Are translations perhaps easier to comprehend than originals? Is interpreting output easier to understand than translation output? Investigating effects on the recipients of translation and interpreting output would provide an interesting complementary perspective on language comprehension and could be straightforwardly approached in an information-theoretic framework of language use as the one presented in this article.

Acknowledgments

The reported research by the authors has been funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) − Project-ID 232722074 − SFB 1102.

References

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233—250). Amsterdam/Philadelphia: John Benjamins.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*(2), 223—245.

Bernardini, S., Ferraresi, A., & Miličević, M. (2016). From EPIC to EPTIC — Exploring simplification in interpreting and translation from an intermodal perspective. *Target, 28*(1), 61—86.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House, & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication. Discourse and cognition in translation and second language acquisition studies* (pp. 17—35). Tübingen: Narr.

Carl, M., & Schaeffer, M. (2017). Why translation is difficult: A corpus-based study of non-literality in post-editing and from scratch translation. *HERMES - Journal of Language and Communication in Business*, (56), 43—57. https://doi.org/10.7146/hjlcb.v0i56.97201

Chesterman, A. (2004). Beyond the particular. In A. Mauranen, & P. Kujamäki (Eds.), *Translation universals: Do they exist?* (pp. 33—49). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.48.04che

Chmiel, A., Kajzer-Wietrzny, M., Koržinek, D., & Janikowski, P. (2021). Cross-linguistic similarities in lexis: Examining cognate activation through temporal and accuracy data from the Polish interpreting corpus (PINC). In *Using corpora in contrastive and translation studies conference (UCCTS)* (6th ed.) Bertinoro (Italy).

Chmiel, A., Janikowski, P., Koržinek, D., Lijewska, A., Kajzer-Wietrzny, M., Jakubowski, D., et al. (2023). Lexical frequency modulates current cognitive load, but triggers no spillover effect in interpreting. *Perspectives, 32*(5), 905—923.

Crocker, M. W., Demberg, V., & Teich, E. (2016). Information density and linguistic encoding (ideal). *KI, 30*(1), 77—81. https://doi.org/10.1007/s13218-015-0391-y

Fankhauser, P., Knappen, J., & Teich, E. (2014). Exploring and visualizing variation in language resources. In *Proceedings of the ninth international conference on language resources and evaluation (LREC'14)* (pp. 4125—4128). Reykjavik, Iceland: European Language Resources Association (ELRA). Retrieved from https://aclanthology.org/L14-1191/.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language, 140*, 1—11. https://doi.org/10.1016/j.bandl.2014.10.006

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America, 117*, 2347—2353.

Hahn, M., Degen, J., & Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Psychological Review, 128*, 726—756. https://doi.org/10.1037/rev0000269

Halverson, S. (2003). The cognitive basis of translation universals. *Target, 15*(2), 197—241.

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2005). Semantic competition between hierarchically related words during speech planning. *Memory & Cognition, 33*, 984—1000. https://doi.org/10.3758/BF03193207

Hedderich, M. A., Lange, L., Adel, H., Strötgen, J., & Klakow, D. (2021). A survey on recent approaches for natural language processing in low-resource scenarios. In *Proceedings of the 2021 conference of the north American chapter of the association for computational linguistics: Human language technologies* (pp. 2545—2568). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.201

Hodzik, E. (2023). Predictive processes in interpreters: Existing findings and future directions in interpreting process research. *Translation, Cognition & Behavior, 6*(2), 141—163. https://doi.org/10.1075/tcb.00083.hod

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23—62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Karakanta, A., Vela, M., & Teich, E. (2018). EuroParl-UdS: Preserving and extending metadata in parliamentary debates. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC)* (pp. 33—38). Miyazaki, Japan: European Language Resources Association (ELRA).

Karakanta, A., Przybyl, H., & Teich, E. (2021). Exploring variation in translation with probabilistic language models. In J. Lavid-López, C. Maíz-Arevalo, & J. R. Zamorano-Mansilla (Eds.), *Corpora in translation and contrastive research in the digital age: Recent advances and explorations* (pp. 308—323). Amsterdam: Benjamins. https://doi.org/10.1075/btl.158.12kar

Klaudy, K., & Károly, K. (2005). Implicitation in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures, 6*(1), 13—28. https://doi.org/10.1556/Acr.6.2005.1.2

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79—86.

Kunilovskaya, M., Przybyl, H., Lapshinova-Koltunski, E., & Teich, E. (2023). Simultaneous interpreting as a noisy channel: How much information gets through. In *Proceedings of the 14th international conference on recent advances in natural language processing* (pp. 608—618). Retrieved from https://aclanthology.org/2023.ranlp-1.66/.

Lapshinova-Koltunski, E., Pollkläsener, C., & Przybyl, H. (2022). Exploring explicitation and implicitation in parallel interpreting and translation corpora. *The Prague Bulletin of Mathematical Linguistics*, (119), 5—22. https://doi.org/10.14712/00326585.020

Lapshinova-Koltunski, E. (2021). Analysing the dimension of mode in translation. In M. Bisiada (Ed.), *Empirical studies in translation and discourse* (pp. 223—243). Berlin: Language Science Press. https://doi.org/10.5281/zenodo.4450014

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19, pp. 849—856). Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126—1177. https://doi.org/10.1016/j.cognition.2007.05.006

Manning, C. D., & Schütze, H. (2003). *Foundations of statistical natural language processing* (6th ed.). Cambridge, MA: MIT Press.

Martínez Martínez, J. M., & Teich, E. (2017). Modeling routine in translation with entropy and surprisal: A comparison of learner and professional translations. In L. Cercel, M. Agnetta, & M. T. Amido Lozano (Eds.), *Kreativität und Hermeneutik in der Translation* (pp. 103—126). Tübingen: Narr.

Michaelov, J. A., Coulson, S., & Bergen, B. K. (2023). So cloze yet so far: N400 amplitude is better predicted by distributional information than human predictability judgements. *IEEE Transactions on Cognitive and Developmental Systems, 15*(3), 1033—1042. https://doi.org/10.1109/TCDS.2022.3176783

Oh, B.-D., & Schuler, W. (2023). Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times? *Transactions of the Association for Computational Linguistics, 11*, 336—350. https://doi.org/10.1162/tacl_a_00548

Olohan, M., & Baker, M. (2000). Reporting "that" in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures, 1*(2), 141—158.

Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting. A corpus-based regression analysis. *Translation and Interpreting Studies, 11*(2). https://doi.org/10.1075/tis.11.2.04ple

Plevoets, K., & Defrancq, B. (2020). Imported load in simultaneous interpreting: An assessment. In R. Muñoz Martín, & S. L. Halverson (Eds.), *Multilingual mediated communication and cognition* (pp. 18—43). London: Routledge.

Pollkläsener, C., Yung, F., & Lapshinova-Koltunski, E. (2024). Capturing variation of discourse relations in English parallel data through automatic annotation and alignment. *Across Languages and Cultures, 25*(2), 288—309.

Pollkläsener, C., Kunilovskaya, M., & Teich, E. (2025). Surprisal explains the occurrence of filler particles in simultaneous interpreting. *SKASE Journal of Translation and Interpreting, Special Issue: Empirical Translation and Interpreting Studies, 18*(2), 55—78.

Przybyl, H., Karakanta, A., Menzel, K., & Teich, E. (2022). Exploring linguistic variation in mediated discourse: Translation vs. interpreting. In M. Kajzer-Wietrzny, S. Bernardini, A. Ferraresi, & I. Ivaska (Eds.), *Empirical investigations into the forms of mediated discourse at the European parliament* (pp. 191–218). Berlin: Language Science Press. https://doi.org/10.5281/zenodo.6977050

Przybyl, H., Lapshinova-Koltunski, E., Menzel, K., Fischer, S., & Teich, E. (2022). EPIC UdS - Creation and applications of a simultaneous interpreting corpus. In *Proceedings of the 13th conference on language resources and evaluation (LREC 2022)* (pp. 1193–1200). Marseille, France: European Language Resources Association (ELRA).

Przybyl, H., Lapshinova-Koltunski, E., & Teich, E. (2023). Investigating surprisal in simultaneous interpreting. In *Using corpora in contrastive and translation studies conference (UCCTS)* (7th ed.) Poznan (Poland).

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., & Carl, M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New directions in empirical translation process research* (pp. 183–210). Cham: Springer International Publishing.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423.

Shlesinger, M., & Ordan, N. (2012). More spoken or more translated?: Exploring a known unknown of simultaneous interpreting. *Target, 24*(1), 43–60.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, (128), 302–319.

Teich, E. (2003). *Cross-linguistic variation in system und text. a methodology for the investigation of translations and comparable texts*. Berlin: Mouton de Gruyter.

Teich, E., Martínez, J. M., & Karakanta, A. (2020). Translation, information theory and cognition. In F. Alves, & A. L. Jakobsen (Eds.), *The routledge handbook of translation and cognition* (pp. 360–375). London, UK: Routledge.

Wei, Y. (2022). Entropy as a measurement of cognitive load in translation. In *Proceedings of the 15th biennial conference of the association for machine translation in the americas (workshop 1: Empirical translation process research)* (pp. 75–86). Orlando, USA: Association for Machine Translation in the Americas.

Yung, F., Scholman, M., Lapshinova-Koltunski, E., Pollkläsener, C., & Demberg, V. (2023). Investigating explicitation of discourse connectives in translation using automatic annotations. In *Proceedings of the 24th annual meeting of the special interest group on discourse and dialogue (SIGDIAL)* (pp. 21–30). Prague, Czechia: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.sigdial-1.2