Corpus-based Translation Studies

Human Translation Quality Estimation (HTQE)

Maria Kunilovskaya

HS Empirical Linguistics and Translatology
MA Translation Science and Technology

May 16, 2024

Outline

1. Translated language as a language variety
   – overview of the research field
   – methodology and features
   – translation detection results

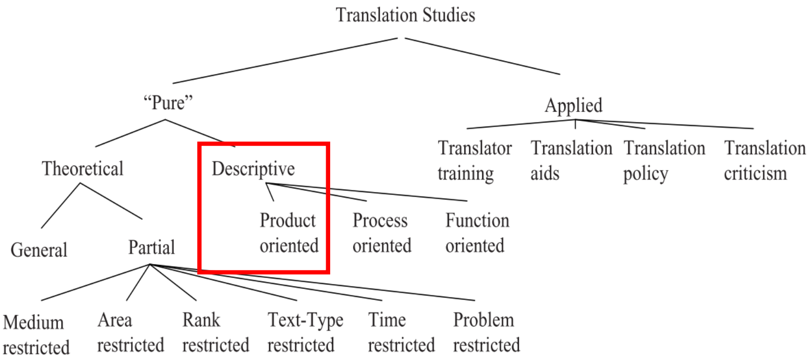
2. Translation quality
   – quality aspects and benchmarking
   – MT: quality evaluation and estimation tasks

3. Is translationese related to quality?
   – debate
   – experimental evidence

4. References

Translated language as the object of study

The map of Translation Studies
by Chesterman(2009)[13] based on Holmes (1988)[21]



Baker (1993)[4]: translations as texts in their own right,
not 'deformation' of sources (Berman, 1985)[7]

Area of research: Empirical Translation Studies

Empirical (Corpus-based) Translation Studies (CBTS) seeks to explain linguistic choices in translations vs. non-translations by language-pair internal or external factors.

Translationese and translated language

Translationese :

properties that make translations statistically different from comparable non-translations[19]

- reflects linguistic, cognitive and sociolinguistic aspects of cross-cultural communication and translation process,
- describes and explains linguistic specificity of translations, 'the property of being a translation', which makes translated language a variety of the target language (TL),
- is revealed through comparison of translations and TL non-translations in the same register as the source texts,
- exists at document level (document length $>$ 450 tokens).

Suggested translational tendencies[12]

S-universals

*** how the translators process the source language (SL) ***

1. interference/transfer, ‘shining through’ effect[50]

   translations follow ST rather than TL patterns, e.g frequency calques, strange strings

2. explicitation

   spelling things out rather than leave them implicit

   - more frequent use of conjunctions, connectives,
   - more re-phrasing, comments, elaboration in brackets
   - ST non-finite clauses > TT finite clauses
   - ST pro-forms (this, they) and ST ellipsis > TT full NPs

3. levelling-out (aka Standardisation/Convergence)

   higher level of homogeneity of translations against sources;

4. lengthening

Suggested translational tendencies, cont.

T-universals
*** how the translators process the TL ***

1. simplification

   lexical (lower TTR=less varied voc, lower lexical density), syntactic (higher readability scores), stylistic (less figurative)

2. normalisation

   "tendency to exaggerate features of the target language and to conform to its typical patterns"[5]

3. unique items hypothesis

   TL specific items are underrepresented (e.g German passive-like constructions: sein+zu, lassen+sich)

NB! Matching trends and specific translationese indicators is tricky.

Translationese varieties

Factors which give rise to different types of translationese:

register and genre :

- a explicitation tendency in German popular scientific translation is weaker in economic texts;
- varying degree of tolerance to literal translation[37];
- same tendency is manifested through different indicators[29]

source language (SL) :
(Evert, 2017; Nikolaev, 2020)[18;40],

competence level : how does professional translation differs from a layman/novice rendition?[39;30;8],

method of translation : MT vs humans vs post-editing[51].

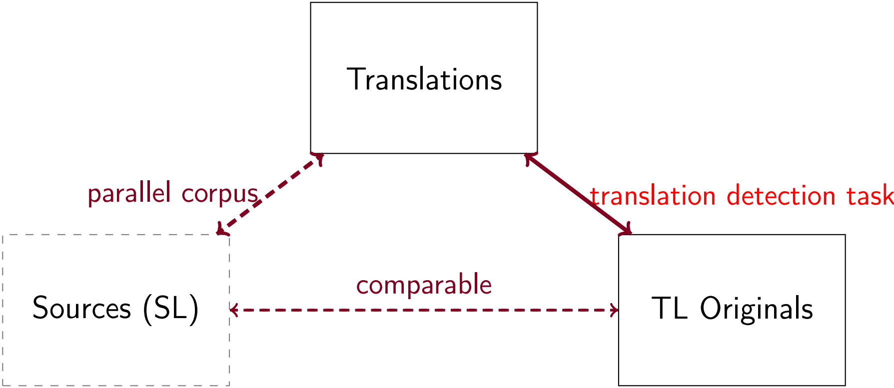
Methodology of a translationese study

related tasks
- translation detection (translations vs non-translations),
- SL detection,
- translation direction detection

required corpora
- translations vs non-translation
- ideally: sent-aligned documents and register-comparable non-translations

methods
- univariate statistical analysis[55;16]
- feature selection
- text classification (single feature[52;23], multivariate[30])
- mildly-supervised methods (LDA, PCA)[18;17] and exploratory clustering[44;34]

Typical resources for a translationese study



Numerical representation: hand-crafted features

(1) Count-based features:

- frequencies of individual items/patterns (e.g. relative *that*)
- cumulative frequencies of listed items (connectives, pronouns)
- frequencies of PoS tags, syntactic dependencies (and combination)
- character[43] or word ngrams (inc. on 'mixed' representations[6])

(2) Calculated metrics and scores:

- lexical variety, density, TTR
- average of senses/syllables per word
- sentence depth as parse tree depth, mean dependency distance
- ratios of N/V, 1st frequency quartile bigrams, neologisms
- Flesch-Kincaid Reading Ease score[46]
- LM entropy scores

Numerical representation: feature-learning approaches

(3) Embedding spaces learnt from delexicalised corpus versions:

- sequences of PoS tags, semantic tags[15;14]

(4) Embeddings from embedding models

- static (fasttext, word2vec)
- contextualised (BERT, XLM-Roberta)[2;9]

Interpretable translationese indicators

Desired properties:

- well-motivated (by contrastive studies, variational/register analysis, prior TS);
- content-independent;
- reasonably frequent;
- reliably extractable;
- language-independent or shared by SL and TL

Lexicogrammar and discourse features fit the bill best.

32 features from Vered Volansky (2015)[52] are used as a benchmark.

Expected deviations and exploratory setups

Structural delexicalised features from UD annotations

well-known indicators and expectations for translations:

- lower lexical variety, TTR,
- lower lexical density,
- overuse of discourse markers,
- higher sentence length,
- overuse of pronouns.

patterns expected from English-to-Russian studies:

- higher hierarchical distance,
- underuse of nsubj:pass (ex. 'resheno prodlit'), negative particles, deverbal nouns,
- overuse of connectives and modal predicates.

Abstract lexical features

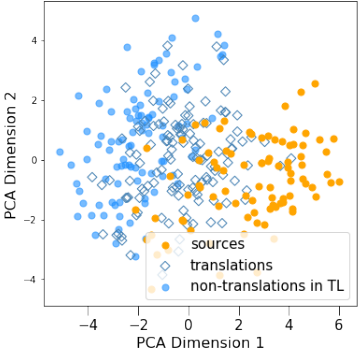
- ratios of 1-2-3-grams from top/bottom freq quartiles,
- mean and σ for sentence perplexity scores.
- ratios of highly- and negatively-collocated phrases,
- using NPMI and Tscore association metrics.

Document-level binary SVM classifier results

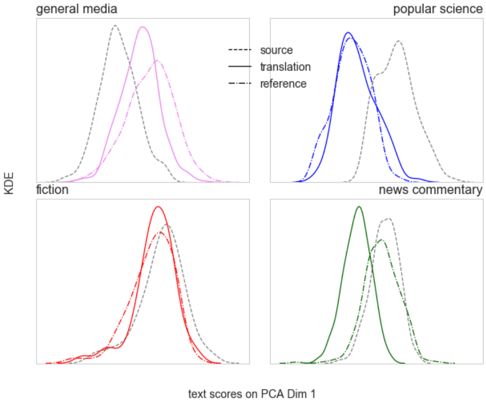
direction	register	F1 (%)	reference
FR-EN	Europarl	83.6	Rabinovich (2016)[44]
DE-EN	Europarl	80.0	Kunilovskaya (2024)Kunilovskaya et al.[28]
EN-DE	Europarl	88.8	
EN-DE	mass media	79.0	Kunilovskaya,Lapshinova (2020)[31]
EN-DE DE-EN	multiregister	75-77	Evert (2017)[18]
EN-ES	Europarl	96.2	Poltorak (2022)[42]
EN-ES	fiction	77.3	
EN-ES	technical	97.6	Ilisei (2010)[24]
EN-RU	mass media	90.2	Kunilovskaya (2023)[27]
EN-RU	fiction	75.6	Kunilovskaya (2021)[32]
DE-RU	fiction	84.4	
ES-RU	fiction	74.4	
SV-RU	fiction	71.1	
UK-RU	fiction	59.4	

Translation detection results on lexicogrammatical and discursive features (in a multivariate setup) across translation directions and registers (on professional published translations)

Visualisations



The specificity of EN-to-DE translations captured by the shining-through indicators



Distribution of texts by category across registers (EN-to-RU)

What's a good translation?

How good is this translation?

Adequacy	usefulness, fitness for communicative purpose, acceptability[41;22]
Accuracy	semantic similarity: *How much of the meaning expressed in the source is also expressed in the target*
Fluency	readability, compliance with TL norms from *Flawless English* to *Incomprehensible*

Undifferentiated approach:

How much do you agree that the translation adequately expresses the meaning of the source?

Quality labels/scores: Human assessment (of HT or MT)

(1) Real-life quality judgments:
education, certification, industrial quality control

(2) Experimental setups

Assessment purpose: quantitative or diagnostic

- summative vs formative
- holistic vs analytical

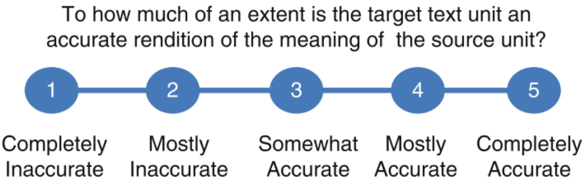
Methods:

1. direct assessment,
2. (analytical) rubrics,
3. error annotation.

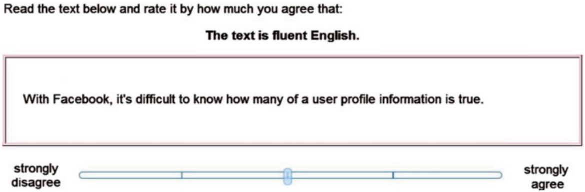
+ in MT: post-editing time/effort (not discussed)

Granularity: document-, sentence-, word-level

Assessment method 1: Direct Assessments (DA)



from Moorkens (2018)[38]



from Graham (2015)[20]

DA: recommendations for producing MT benchmarks

from Läubli et al. (2020)[35]

- use language professionals as annotators,
- evaluate documents, not sentences; or sentences in context,
- evaluate fluency in a monolingual setup, separately from accuracy,
- avoid reference translations $\rightarrow$ use bilingual setups for accuracy,
- use original source texts.

Assessment method 2: Rubrics

Diploma in Translation (DipTrans, UK certification)

1. comprehension, accuracy and register (max 50);
2. grammar (morphology, syntax, etc.), cohesion, coherence and organisation of work (max 35);
3. technical aspects: punctuation, spelling, accentuation, names, dates, figures, etc (max 15).

BANDS: distinction, merit, pass, fail with numeric marks

American Translators Association (ATA)

1. usefulness/transfer (max 35);
2. terminology/style (max 25);
3. idiomatic writing (max 25)
4. target mechanics (max 15)

BANDS: standard, strong, acceptable, deficient and minimal (Williams, 2013; Yuan, 2018)[54;56]

Assessment method 3: harmonised DQF-MQM error taxonomy[1]

a standard but adjustable way to categorise and measure translation quality

Top-level categories (with some subcategories)

- accuracy (addition/omission, improper exact TM match, mistranslation, untranslated)
- fluency (grammar, spelling, character encoding)
- locale convention (address/currency format, shortcut key)
- style (awkward, company style, unidiomatic)
- terminology (inconsistent with termbase)
- verity (culture-specific references)

[1] https://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf

Digression 1: Quality-related NLP tasks in MT

Quality Evaluation

measure distance from the candidate translation to another translation (aka reference), usually a human translation

Most used metrics:

- BLEU family
- HTER
- COMET[47]
- ...

in HT this means punishing creativity and variety

Quality Estimation

devise a way (supervised or unsupervised) to predict quality labels without references

Approach:

- feature-engineering (QuEst++[49])
- using embeddings (deepQuest[25], TransQuest[45])

Granularity:

- sentence-level
- word-level (predicting errors)

Digression 2: Humans vs machines in translation

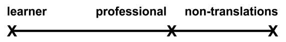
How human translation (HT) differs from MT:

1. HT is essentially document-level $\rightarrow$
   sentence-level representations less adequate
2. HT is more varied, less literal $\rightarrow$
   reference-based approaches not good
   higher granularity of quality analysis required
3. in HT publishable quality is expected
4. lack of reliable quality labels / available datasets $\rightarrow$
   same as in MT: k 0.2-0.4 (Graham, 2015)
5. HT allows no direct access to internal processes $\rightarrow$
   no 'glassbox' features
6. HT and SOTA NMT might need focus on different aspects of quality: fluency and accuracy respectively

Is translationese indicative of quality?

Yes:
more translated, lower quality

- the link is suggested or explored in previous work: for HT[48;17;36], for MT[1;3],
- professionals are less deviant than students (based on univariate analysis)[33]:



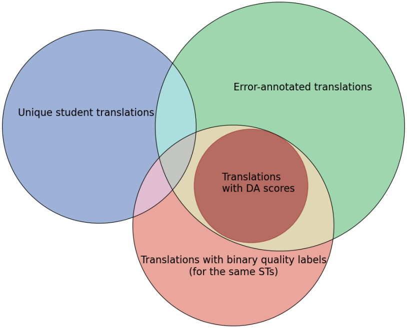
- quality in HT is mostly about fluency[11],
- human annotators have difficulties differentiating between the three aspects of quality[10]

No:
it is a subtle inherent property

- absence of translationese can signal reduced accuracy or another type of cross-cultural communication (transcreation, adaptation)
- Wein (2023)[53]: human annotators cannot identify translations (also see Baroni (2006)[6].
- Jimenez-Crespo (2023)[26]: pejorative implications of translationese are unethical towards translators.

Research subcorpora

1. subsets from *Russian Learner Translator corpus* of various sizes by type of quality annotation



2. comparable professional translations: 404 parallel docs, 384 K words (BBC Russian Service, InoSMi, RNC);
3. comparable non-translations: 497 docs, 523 K words (RNC)

Four types of quality labels/scores

Operational definitions of quality

- Holistic judgments: agreed assessment of competition jury/exam board in real life; top and bottom grades converted to 'bad', 'good' labels, verified in an additional annotation experiment ($\alpha = 0.524$, accuracy 91%).
- Scores from error annotation used as part of feedback to students in a real-life practical translation course, which implemented accuracy-fluency distinction (top-level category agreement: 80.5% of errors in the same location, $\alpha = 0.535$).
- Direct assessment: perceived quality for sentences presented in the context on a 100-point scale (documents: $\alpha = 0.541$, sentences: $\alpha = 0.463$)

+ Known ontological status of translations produced by defined subjects (students, professionals).

Results for HTQE on translationese indicators using SVM

All results are at document level on the best-performing features.

Translationese classification on T-features: F1 = 90.2% (professional), 88.96% (students)

Binary labels, SVM, F1-score

- best-worst: 68.9%
- students-prof: 73.3%

Continuous scores, SVR, Pearson r

error-annotation scores: 0.43
direct assessment: 0.23

HTQE previous research results (alternative approaches)

Pearson correlation coefficient (r) between predicted and true scores

- Yuan (2018, 2020)[56]
  - setting: English-to-Chinese, 458 student translations to 6 sources (sic!),
    4 continuous scores (ATA rubrics),
    360 hand-crafted language-independent features
  - best result
    - document-level (features): $r = 0.62$-0.76
      (cf. MTQE WMT20 $r = 0.53$)
- Zhou (2019)[57]
  - setting: Japanese-to-English, unsupervised approach: correlation between ST/TT similarity/distance measures based on word vectors and overall quality graded by humans for 130 sentence pairs from camera manuals
  - result: $r = 0.53$

Summary

1. Translated language is a subsystem of the TL.
2. Machines pick up statistical deviations better than humans.
3. The relevance of translationese for quality estimation is low (especially if quality is assessed at sentence level).

The end

Thank you!

Questions?

References I

[1] Aharoni, R. (2015). Automatic Detection of Machine Translated Text and Translation Quality Estimation. Master's thesis, Bar-Ilan University.

[2] Amponsah-Kaakyire, K., Pylypenko, D., van Genabith, J., & España-Bonet, C. (2022). Explaining Translationese: why are Neural Classifiers Better and what do they Learn? BlackboxNLP 2022 - BlackboxNLP Analyzing and Interpreting Neural Networks for NLP, Proceedings of the Workshop, (pp. 281–296).

[3] Aranberri, N. (2020). Can translationese features help users select an MT system for post-editing? Procesamiento de Lenguaje Natural, 64(1), 93–100.

[4] 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. 232–250). Amsterdam: John Benjamins Publishing Company.

References II

[5] Baker, M. (1996). Corpus-based translation studies: the challenges that lie ahead. Terminology, LSP and translation: studies in language engineering, in honour of Juan C. Sager., 18, 175–186.

[6] Baroni, M. & Bernardini, S. (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text. Literary and Linguistic Computing, 21(3), 259–274.

[7] Berman, A. (1985). Translation and the trials of the foreign. In L. Venuti (Ed.), The Translation Studies Reader (pp. 284–297). London: Routledge. Reprinted from La Traduction comme épreuve de l'étranger, Texte 4:67–81. Translated by L.Venuti.

References III

[8] Bizzoni, Y. & Lapshinova-Koltunski, E. (2021). Measuring Translationese across Levels of Expertise: Are Professionals more Surprising than Students? In Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa) (pp. 53–63). Reykjavik: Linköping University Electronic Press.

[9] Borah, A., Pylypenko, D., & Espa, C. (2023). Measuring Spurious Correlation in Classification : " Clever Hans " in Translationese. In Proceedings of the Recent Advances in Natural Language Processing, number 2021 (pp. 196–206).

[10] Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., & Schroeder, J. (2007). (Meta-) Evaluation of Machine Translation. In Proceedings of the Second Workshop on Statistical Machine Translation, ACL (pp. 136–158).

References IV

[11] Carl, M. & Buch-Kromann, M. (2010). Correlating translation product and translation process data of professional and student translators. 14 Annual Conference of the European Association for Machine Translation, Saint-Raphaël, France, (May).

[12] Chesterman, A. (2004). Hypotheses about translation universals. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), Claims, Changes and Challenges in Translation Studies (pp. 1–14). Amsterdam and Philadelphia:: Benjamins.

[13] Chesterman, A. (2009). The Name and Nature of Translator Studies. Hermes, 42(42), 13–22.

[14] Chowdhury, K. D., Espã na Bonet, C., van Genabith, J., & Genabith, J. (2020). Understanding Translationese in Multi-view Embedding Spaces. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 6056–6062). Barcelona, Spain: Online.

References V

[15] Chowdhury, K. D., Espa˜na-Bonet, C., & van Genabith, J. (2021). Tracing Source Language Interference in Translation with Graph-Isomorphism Measures. In RANLP-2021 (pp. 380–390).

[16] Dai, G. & Xiao, R. (2011). "SL Shining Through" in Translational Language: A Corpus-based Study of Chinese Translation of English Passives. Translation Quarterly, 62, 85–108.

[17] De Sutter, G., Cappelle, B., De Clercq, O., Loock, R., & Plevoets, K. (2017). Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations. Linguistica Antverpiensia, 16, 25–39.

References VI

[18] Evert, S. & Neumann, S. (2017). The impact of translation direction on characteristics of translated texts : A multivariate analysis for English and German. In G. De Sutter, M. Lefer, & I. Delaere (Eds.), Empirical Translation Studies: New Methodological and Theoretical Traditions, volume 300 (pp. 47—-80). Walter de Gruyter GmbH & Co KG.

[19] Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin & H. Lindquist (Eds.), Translation Studies in Scandinavia (pp. 88–95). Lund: CWK Gleerup.

[20] Graham, Yvette., Baldwin, Timothy, Moffat, Alistair and Zobel, Justin (2015). Can machine translation systems be evaluated by the crowd alone. Natural Language Engineering, 23(1), 3–30.

References VII

[21] Holmes, J. S. (1988). The name and nature of translation studies. In L. Venuti (Ed.), The Translation Studies Reader (pp. 172–185). London: Routledge. Reprinted from Translated! Papers on Literary Translation and Translation Studies by J.S. Holmes (pp. 67-80). Amsterdam: Rodopi.

[22] House, J. (2001). Translation Quality Assessment: Linguistic Description versus Social Evaluation. Meta: Journal des traducteurs, 46(2), 243.

[23] Hu, H. & Kuebler, S. (2021). Investigating Translated Chinese and Its Variants Using Machine Learning. Natural Language Engineering, 5, 339–372.

References VIII

[24] Ilisei, I., Inkpen, D., Pastor, G. C., & Mitkov, R. (2010). Identification of Translationese : A Machine Learning Approach. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 503–511). Berlin Heidelberg: Springer.

[25] Ive, J., Blain, F., & Specia, L. (2018). deepQuest: A Framework for Neural-based Quality Estimation. In Proceedings ofthe 27th International Conference on Computational Linguistics (pp. 3146–3157).

[26] Jiménez-Crespo, M. A. (2023). "Translationese" (and "post-editese"?) no more: on importing fuzzy conceptual tools from Translation Studies in MT research. Proceedings of the 24th Annual Conference of the European Association for Machine Translation, EAMT 2023, (June), 261–268.

References IX

[27] Kunilovskaya, M. (2023).
Translationese indicators for human translation quality estimation (base
PhD thesis, University of Wolverhamptopn.

[28] Kunilovskaya, M., Chowdhury, K. D., Przybyl, H., España i Bonet, C., & Van Genabith, J. (2024). Mitigating Translationese with GPT-4: Strategies and Performance. In Proceedings of the 25th Annual conference of the European Association for Machine Translation (pp. in print). Sheffield, UK: Association for Computational Linguistics.

[29] Kunilovskaya, M. & Corpas Pastor, G. (2021). Translationese and register variation in English-to-Russian professional translation. In L. Lim, D. Li, & V. Wang (Eds.), New Perspectives on Corpus Translation Studies, New Frontiers in Translation Studies. Springer.

References X

[30] Kunilovskaya, M. & Lapshinova-Koltunski, E. (2020a). Lexicogrammatic Translationese across Two Targets and Competence Levels. In N. Calzolari, F. Bechet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, & A. Others (Eds.), Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020) (pp. 4102–4112).: The European Language Resources Association (ELRA).

[31] Kunilovskaya, M. & Lapshinova-Koltunski, E. (2020b). Lexicogrammatic translationese across two targets and competence levels. In Proceedings of The 12th Language Resources and Evaluation Conference (pp. 4102–4112). Marseille, France: European Language Resources Association.

References XI

[32] Kunilovskaya, M., Lapshinova-koltunski, E., & Mitkov, R. (2021). Translationese in Russian literary texts. In Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (pp. 101–112).: Association for Computational Linguistics.

[33] Kunilovskaya, M., Morgoun, N., & Pariy, A. (2018). Learner vs. professional translations into Russian: Lexical profiles. Translation and Interpreting, 10(1).

[34] Lapshinova-Koltunski, E. (2017). Exploratory analysis of dimensions influencing variation in translation. The case of text register and translation method. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), Empirical Translation Studies. New Theoretical and Methodological Traditions. . Vol. 300 (pp. 207–234). Walter de Gruyter GmbH & Co KG.

References XII

[35] Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., & Toral, A. (2020). A set of recommendations for assessing human-machine parity in language translation. Journal of Artificial Intelligence Research, 67, 653–672.

[36] Loock, R. (2016). La traductologie de corpus. Villeneuve d'Ascq: Presses universitaires du Septentrion.

[37] Mauranen, A. (2004). Corpora, universals and interference. In Translation universals: Do they exist (pp. 65–82). Amsterdam: J. Benjamins.

[38] Moorkens, J., Castilho, S., Gaspari, F., & Doherty, S. (2018). Translation Quality Assessment From Principles to Practice Machine Tr
Springer International Publishing AG.

References XIII

[39] Nakamura, S. (2007). Comparison of Features of Texts Translated by Professional and Learner Translators. In Proceedings of the 4th Corpus Linguistics conference: University of Birmingham.

[40] Nikolaev, D., Karidi, T., Kenneth, N., Mitnik, V., Saeboe, L., & Abend, O. (2020). Morphosyntactic Predictability of Translationese. Linguistics Vanguard, 6(1).

[41] Nord, C. (2006). Translating as a purposeful activity: a prospective approach. TEFLIN Journal: A publication on the teaching and . . . , 17(2), 131–143.

[42] Poltorak, K. (2022). Computational Approaches to Register as a Factor in English-to-Spanish Translation. In New Trends in Translation and Technology (pp. 82–90).

References XIV

[43] Popescu, M. (2011). Studying Translationese at the Character Level. Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, (September), 634–639.

[44] Rabinovich, E. & Wintner, S. (2016). Unsupervised Identification of Translationese. Transactions of the Association for Computational Linguistics, 3, 419—-432.

[45] Ranasinghe, T., Orasan, C., & Mitkov, R. (2020). TransQuest at WMT2020: Sentence-Level Direct Assessment. In Proceedings ofthe 5th Conference on Machine Translation (WMT) (pp. 1049–1055).

[46] Redelinghuys, K. & Kruger, H. (2015). Using the features of translated language to investigate translation expertise A corpus-based study*. International Journal of Corpus Linguistics, 20(3), 293–325.

References XV

[47] Rei, R., De Souza, J. G., Alves, D., Zerva, C., Farinha, A. C., Glushkova, T., Lavie, A., Coheur, L., & Martins, A. F. (2022). COMET-22: Unbabel-IST 2022 submission for the metrics shared task. In Proceedings of the Seventh Conference on Machine Translation (WMT) (pp. 578–585).

[48] Scarpa, F. (2006). Corpus-based quality-assessment of specialist translation: A study using parallel and comparable corpora in English and Italian. In M. Gotti, S. Sarcevic, & S. Šarčević (Eds.), Insights into specialized translation (pp. 155–172). Bern: Peter Lang.

References XVI

[49] Specia, L., Paetzold, G. H., & Scarton, C. (2015). Multi-level translation quality prediction with QUEST++. In ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Proceedings of System Demonstrations (pp. 115–120).: Association for Computational Linguistics.

[50] Teich, E. (2003). Cross-linguistic variation in system and text. A methodology for the investigation of translations and comparable texts. Berlin/Boston: Walter de Gruyter.

[51] Vela, M. & Lapshinova-Koltunski, E. (2015). Register-Based Machine Translation Evaluation with Text Classification Techniques. In Y. Al-Onaizan & W. Lewis (Eds.), Proceedings of the 15th Machine Translation Summit. Vol. 1: MT Researchers' Track (pp. 215–228).: Association for Machine Translation in the Americas.

References XVII

[52] Volansky, V., Ordan, N., & Wintner, S. (2015). On the features of translationese. Digital Scholarship in the Humanities, 30(1), 98–118.

[53] Wein, S. (2023). Human raters cannot distinguish English translations from original English texts. In H. Bouamor, J. Pino, & K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 12266–12272). Singapore: Association for Computational Linguistics.

[54] Williams, M. (2013). A holistic-componential model for assessing translation student performance and competency. Mutatis Mutandis, 6(2), 419–443.

[55] Xiao, R. (2010). How different is translated Chinese from native Chinese? International Journal of Corpus Linguistics, (15(1)), 5–35.

References XVIII

[56] Yuan, Y. (2018). Human Translation Quality Estimation: Feature-based and Deep Learning-based. PhD thesis, University of Leeds.

[57] Zhou, Y. & Bollegala, D. (2019). Unsupervised evaluation of human translation quality. In Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management: SciTePress-Science and Technology Publications.