# Assessing Genre and Method Variation in Translation Using Computational Techniques

*Ekaterina Lapshinova-Koltunski and Marcos Zampieri*

Paris

16 January 2015

# Overview

1. Aims and Motivation
2. Related Work and Theory
   - Register
   - Translation method
   - Our previous work
   - Text Classification
3. Methods and Data
   - Methods
   - Data
4. Experiment Results
   - BoW
   - Bigrams

# Motivation

- variation in translation can include several parameters or dimensions, e.g. language, method, register, etc.
- different types of translations distinguished by these dimensions $\Rightarrow$ **translation varieties**, see [Lapshinova-Koltunski, 2015].
- interaction of these dimensions is reflected in the translation product, i.e. in its linguistic features
- dimensions are "recognisable" via feature profiles formed by distributions of these features
- Features: "known" and "unknown"
- classification with "known" features deliver average results (previous work)
- What about "unknown" features?

# Aims and Goals

UNIVERSITÄT DES SAARLANDES

- use automatic text classification techniques to analyse variation in English-German translations

Main goals:

- discriminate between
  - different registers
  - different translation methods
- to level out discriminative features in this classification task

(!) text classification methods can level out features of different subcorpora including those not implied by existing theories

$\Rightarrow$ “unknown” features

- investigate in more detail the properties of each of them

# Register and Genre in Translation

UNIVERSITÄT DES SAARLANDES

- human translation: analysis of register and genre settings, see [House, 1997]/[House, 2014], [Steiner, 1996], [Steiner, 2004], [Hansen-Schirra et al., 2012], [Sutter et al., 2012], [Delaere and Sutter, 2013] and [Neumann, 2013]
- machine translation: **?**
- some examples: errors in translation of new domains in [Irvine et al., 2013]
- However: lexical level only, as the authors operate solely with the notion of domain (field of discourse) and not register (which includes more parameters)
- further examples: application of in-domain comparable corpora, see [Laranjeira et al., 2014, Irvine and Callison-Burch, 2014]

# Register and Genre Theory

UNIVERSITÄT DES SAARLANDES

- contextual variation of languages:
  languages vary according to their context or situation of use, see [Quirk et al., 1985], [Halliday and Hasan, 1989] or [Biber, 1995]
- contexts influence the distribution of particular lexico-grammatical patterns which manifest language registers
- parameters of variation: variables of *field, tenor* and *mode* in SFL, cf. [Halliday and Hasan, 1989] and [Halliday, 2004]
- in language:
  - field: term patterns or functional verb classes (e.g. , activity, communication, etc.)
  - tenor: modality (expressed e.g. by modal verbs) or stance expressions
  - mode: information structure and textual cohesion (e.g. personal and demonstrative reference).

# Register and Genre Theory

UNIVERSITÄT DES SAARLANDES

- ⇒ differences between registers can be identified through the analysis of distributions of lexico-grammatical features in these registers, e.g. [Biber, 1988, Biber, 1995] or [Biber et al., 1999]

- Multilingual context (linguistic variation across languages):
    - [Biber, 1995] on English, Nukulaelae Tuvaluan, Korean and Somali
    - [Hansen-Schirra et al., 2012] and [Neumann, 2013] on English and German (including translation)
    - register and translation also in [House, 1997], [House, 2014], [Steiner, 1996], [Steiner, 2004], [Sutter et al., 2012], [Delaere and Sutter, 2013]
    - However: no distributions, individual texts, individual features

# Translation Method

- studies addressing both human and machine translations: [White, 1994], [Papineni et al., 2002], [Babych et al., 2004], [Popović and Burchardt, 2011], [Popovic and Ney, 2011]
- all focus solely on translation error analysis, using human translation as a reference
- studies operating with linguistically-motivated categories: [Popović and Burchardt, 2011], [Popovic and Ney, 2011] or [Fishel et al., 2012]
- However: none of them provides a comprehensive analysis of specific linguistically motivated features of different registers and translation methods

# Translation Method

- works on differentiation between human and machine translation: (1) [Volansky et al., 2011] and (2) [El-Haj et al., 2014]:

(1)
- analysis of human and machine translations, and comparable non-translated texts
- a range of features based on the theory of translationese, see [Gellerstam, 1986]
- claim that the features specific for human translations can be used to identify MT
- coinciding and diversifying features

(2)
- compare translation style and consistency in human and machine translations of Camus' novel "The Stranger" (French-English and French-Arabic)
- measure: readability as a proxy for style
- evaluative and not descriptive character

- However: one register only

# Translationese

- [Gellerstam, 1986], [Baker, 1993] and [Baker, 1995]
- fine-grained classification:
  - explicitation: a tendency to spell things out rather than leave them implicit
  - simplification: a tendency to simplify the language used in translation
  - normalisation: a tendency to exaggerate features of the target language and to conform to its typical patterns
  - convergence: a relatively higher level of homogeneity of translated texts with regard to their own scores of lexical density, sentence length, etc.
  - shining through: features of the source texts observed in translations

# Our Previous Work

UNIVERSITÄT DES SAARLANDES

1. [Lapshinova-Koltunski, 2015]: clustering (HCA)
2. [Lapshinova-Koltunski and Vela, tted]: classification with K-nearest-neighbour (KNN)

- a set of features derived from:
  - studies on register
  - studies on translationese
- lexico-grammatical patterns of more abstract concepts expressed via certain syntactic constructions
- Requirements:
  - reflect linguistic characteristics of all texts under analysis
  - content-independent (do not contain terminology or keywords)
  - easy to interpret

# Our Previous Work: Features

UNIVERSITÄT DES SAARLANDES

| | patterns | register | translationese |
|---|---|---|---|
| 1 | content vs. grammatical words | mode | simplification |
| 2 | nominal vs. verbal word classes and phrases | field | normalisation / shining through |
| 3 | *ung*-nominalisation | field | normalisation / shining through |
| 4 | nominal vs. pronominal and demonstrative vs. personal | mode | explicitation, normalisation / shining through |
| 5 | abstract or general nouns vs. all other nouns | fiels | explicitation |
| 6 | logico-semantic relations: additive, adversative, causal, temporal, modal | mode | explicitation |
| 7 | modal meanings: obligation, permission, volition | tenor | normalisation / shining through |
| 8 | evaluative patterns | tenor | normalisation / shining through |

# Our Previous Work: Results

- variation is greater along register, not translation method
- machine translations are less diverse than human ones
- intratranslational variation is similar across different translation methods

- Influencing factors:
  - register settings of EO and GO
  - the nature of features

- We need further features, e.g. new patterns which can be provided by the output of a text classification based on bags of words

# Text Classification

- Text classification is an important area of research in NLP and it has been applied to a wide range of tasks such as spam detection, language identification and temporal text classification .
- In recent works, text classification operates with linguistically motivated features to investigate language variation across corpora [Diwersy et al., 2014]
- [Corston-Oliver et al., 2001] present a method to evaluate the fluency of machine translation output by training a classifier to distinguish between human translations and MT (using linguistically-motivated features extracted from a Spanish-English corpus)
- [Ilisei et al., 2010] apply machine learning classifiers to distinguish between translated and non-translated texts (using simplification features and an English-Spanish corpus)

# Algorithms: Naive Bayes

UNIVERSITÄT DES SAARLANDES

Naive Bayes (NB) classifier, based on Bayes theory and probability represented by the following equation:

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)} \quad (1)$$

As described in [Kibriya et al., 2004], NB applied to text classification computes class probabilities for a given document and the set of classes is represented by $C$. NB assigns a text document $t_i$ to the class with the highest probability $P(c|t_i)$ given by the equation below for $c \in C$:

$$P(c|t_i) = \frac{P(t_i|c)P(c)}{P(t_i)} \quad (2)$$

# Algorithms: Likelihood Estimation

Likelihood function calculated over smoothed language models. Models can contain characters and words or linguistic motivated features such as POS categories [Zampieri et al., 2013], morphological categories or (semi-)delexicalized models (described here).

$$P(L|text) = \arg\max_{L} \sum_{i=1}^{N} \log P(n_i|L) + \log P(L) \tag{3}$$

$N$ is the number of n-grams in the test text, $n_i$ is the ith n-gram and $L$ stands for the language models. Given a test text, we calculate the probability for each of the language models. The language model with highest probability determines the identified class for each particular text.

# Corpus

UNIVERSITÄT DES SAARLANDES

**VARTRA-SMALL,** cf. Lapshinova (2013)

contains:

- variants of translation from English into German
  = translation varieties produced by:
  (1) human professional translators (PT1)
  (2) human inexperienced translators (PT2)
  (3) a rule-based MT system (RBMT)
  (4) 2 statistical MT systems (SMT1 and SMT2)

  TOTAL number of tokens in translations ca. 600,000

# Corpus

- PT1 – CroCo, [Hansen-Schirra et al., 2012]
- PT2 – trained translators (over BA) with no/little experience
- RBMT – SYSTRAN
- SMT1 – Google Translate (big undefined data)
- SMT2 – Moses system (small known data)

Each translation covers 7 registers:

- political essays – ESSAY
- fictional texts– FICTION
- instruction manuals– INSTR
- popular-scientific articles– POPSCI
- letters of share-holders– SHARE
- prepared political speeches– SPEECH
- touristic leaflets – TOU

# Data Pre-processing

- The corpus was split into sentences and classification is therefore performed on sentence level.
- A total number of 6200 instances.
- Splitting: training set (80%) vs. testing set (20%).
- Previous studies show that named entities influence classification $\Rightarrow$ we use a semi-delexicalised representation (placeholders instead of nouns).
- This is done to minimize topic variation

# Features Used

- Bag-of-words (BoW).
- Semi-delixicalized BoW.
- Word bigrams and word trigrams (both semi-delixicalized) using an n-gram language model with add one smoothing.

$$P_{lap}(w_1 \ldots w_n) = \frac{C(w_1 \ldots w_n) + 1}{N + B} \qquad (4)$$

$C$ is the count of the frequency of $w_1$ to $w_n$ in the training data, $N$ is the total number of n-grams and $B$ is the number of distinct n-grams in the training data.

# Classification: Registers and Methods

UNIVERSITÄT DES SAARLANDES

- use bag-of-words (including lexical information) to distinguish:
  1. translation methods: PT1 vs. PT2 vs. RBMT vs. SMT1 vs SMT2
  2. registers: ESSAY vs. FICTION vs. INSTR vs. POPSCI vs. SHARE vs. SPEECH vs. TOU

| Type | Classes | Precision | Recall | F-Measure | Baseline |
|---|---|---|---|---|---|
| method | 5 | 35.9% | 36.2% | 35.3% | 20.0% |
| register | 7 | 57.4% | 57.8% | 57.3% | 14.2% |

- registers are better distinguishable than translation method
- similar tendencies in our previous work
- differences between method-based translation varieties less prominent ⇒ convergence?
- performance might be influenced by domain-specific items?
- ⇒ domain-independent features (placeholders) in the next steps

# Method of Translation

- use domain-independent bag-of-words to distinguish:
  1. PT1 vs. PT2 vs. RBMT vs. SMT1 vs. SMT2
  2. PT1 vs. PT2vs. RBMT vs. SMT

| Classes | Precision | Recall | F-Measure | Baseline |
|---|---|---|---|---|
| (1) | 35.1% | 35.9% | 34.9% | 20.0% |
| (2) | 43.2% | 44.9% | 43.1% | 25.0% |

- achieve a better performance for set (2)
- differences in translation methods are less fine-grained
- differences between method-based translation varieties less prominent?

# Register

- use domain-independent bag-of-words to distinguish:
    - seven classes: ESSAY vs. FICTION vs. INSTR vs. POPSCI vs. SHARE vs. SPEECH vs. TOU

| Classes | Precision | Recall | F-Measure | Baseline |
|---|---|---|---|---|
| register | 45.5% | 46.1% | 45.4% | 14.2% |

- performance for register distinction decreases with domain-independent features
- $\Leftarrow$ domain represent one of the parameters of register and reflects what a text is about, i.e. its topic
- more about text than register

# Consistency in Register Variation

- use domain-independent bag-of-words to distinguish:
  - seven classes: ESSAY vs. FICTION vs. INSTR vs. POPSCI vs. SHARE vs. SPEECH vs. TOU within one translation method

| Method | ESS | FIC | INS | POP | TOU | SPE | SHA | Baseline |
|---|---|---|---|---|---|---|---|---|
| PT1 | 0.314 | 0.606 | 0.664 | 0.456 | 0.425 | 0.371 | 0.507 | 0.142 |
| PT2 | 0.399 | 0.533 | 0.595 | 0.372 | 0.421 | 0.346 | 0.536 | 0.142 |
| RBMT | 0.397 | 0.536 | 0.632 | 0.411 | 0.440 | 0.320 | 0.515 | 0.142 |
| SMT | 0.394 | 0.503 | 0.630 | 0.455 | 0.460 | 0.408 | 0.505 | 0.142 |

- the results are similar over all translation methods
- our classification is robust

# More Complex Features

- use semi-delexicalised bi-/trigrams

- differences in translation methods are less fine-grained
  ⇒ reduce the dataset to two classes: human vs. machine

| method | precision | recall | F-measure |
|---|---|---|---|
| human | 0.53 | 0.58 | 0.55 |
| machine | 0.54 | 0.49 | 0.51 |

- two classes of register as an example: ESSAY vs. FICTION

| register | precision | recall | F-measure |
|---|---|---|---|
| ESSAY | 0.54 | 1.00 | 0.70 |
| FICTION | 1.00 | 0.14 | 0.25 |

# Method of Translation: Features

- human:
    1. *Ein PLH* ⇒ full NP (with an indef.modif)
    2. *Wir sind* ⇒ personal reference (1st pers. plural)
    3. *Dies ist* ⇒ extended reference (demonst.)
    4. *Bei der* ⇒ prepositional phrase with local meaning
    5. *Auf dem* ⇒ prepositional phrase with local meaning
    6. *Zu den* ⇒ prepositional phrase with local meaning
    7. *Und wenn* ⇒ ⇒ conditional conj. relation (with a multi-word conj)
    8. *Durch das* ⇒ prepositional phrase with local meaning
    9. *Die PLHSA* ⇒ full NP (with a def.modif)
    10. *Bei PLH* ⇒ prepositional phrase with local meaning
    11. *Auf PLH* ⇒ prepositional phrase with local meaning
    12. *Dies wird* ⇒ extended reference (demonst.)
    13. *' Und* ⇒ additive conjunctive relation
    14. *Wenn sie* ⇒ conjunctive relations
    15. *Die PLHU* ⇒ full NP

# Method of Translation: Features

- machine
  1. *Der PLH* ⇒ full NP (with a def.modif)
  2. *Diese PLH* ⇒ full NP (with a def.modif)
  3. *Wenn die* ⇒ conditional conj. relation
  4. *In PLH* ⇒ prepositional phrase with local meaning
  5. *Aber wir* ⇒ adversative conj. relation
  6. *Aber die* ⇒ adversative conj. relation
  7. *Mit PLH* ⇒ prepositional phrase
  8. *Ich habe* ⇒ personal reference (1st pers. sg)
  9. *Zum PLH* ⇒ prepositional phrase
  10. *Und es* ⇒ additive conj. relation and extended reference (pers)
  11. *Es war* ⇒ extended reference (pers)
  12. *A PLH* ⇒ full NP (with an indef.modif)
  13. *Unser PLH* ⇒ full NP (with a poss.modif)
  14. *Aber es* ⇒ adversative conj. relation
  15. *Mit der* ⇒ prepositional phrase

# Method of Translation: Features

UNIVERSITÄT DES SAARLANDES

## Summary for human and machine

| **human** | **machine** |
|---|---|
| full NP<br>(with def./indef. modif.) | full NP<br>(with def./indef./poss. modif.) |
| personal reference<br>(1st pers. plural) | personal reference<br>(1st pers. sg) |
| extended reference (demonst.) | extended reference (pers.) |
| prepositional phrase<br>with local meaning | prepositional phrase<br>with different meanings |
| additive and conditional conj. relations<br>(often with a multi-word conj) | adversative and conditional conj. relations |

# Register: Features

- ESSAY
  1. *Und - im* ⇒ additive conj. relation
  2. *und/ oder technische* ⇒ additive conj. relation
  3. *Ich möchte absolut* ⇒ modal meaning of volition
  4. *dass wir haben* ⇒ additive conj. relation, that-clause
  5. *in PLH gezahlt.* ⇒ passive
  6. *2003 verkündete PLHäsident* ⇒ passive
  7. *dieses PLH gelegt.* ⇒ demonstrative reference, passive
  8. *weniger befestigt zu* ⇒ passive
  9. *zu erfüllen hat.* ⇒ to-infinitive
  10. *nicht fürchten, sondern* ⇒ adversative conj. relation
  11. *auf langgehaltenen PLH* ⇒ prepositional phrase with local meaning
  12. *letzten PLH verzerrt.* ⇒ passive
  13. *PLH haben sollten,* ⇒ modal meaning of obligation
  14. *zu liberalisieren und* ⇒ to-infinitive
  15. *dass sie weder* ⇒ additive conj. relation, that-clause

# Register: Features

- FICTION
  1. *' Die PLH* ⇒ full NP with a def. modifier
  2. *' ' Aber* ⇒ adversative conj. relation
  3. *PLH. Ich bin* ⇒ personal reference (1st pers. sg.)
  4. *nett. Kein PLH,* ⇒ adjective, negation
  5. *PLH. Nicht lyrisch,* ⇒ adjective, negation
  6. *der großen merkwürdigen* ⇒ adjectives
  7. *trug ein weißes* ⇒ active verb, adjective
  8. *wissen, ist sie* ⇒ active verb
  9. *versuchte, sie an* ⇒ active verb
  10. *würden sie mich* ⇒ subjunctive
  11. *getan. Ich respektiere* ⇒ active verb
  12. *innen, selben schimmern,* ⇒ active verb
  13. *stabil und ein* ⇒ adjective
  14. *eine billige PLH,* ⇒ adjective, full NP
  15. *das PLH, aber* ⇒ full NP, adversative conj. relation

# Register: Features

UNIVERSITÄT DES SAARLANDES

## Summary for ESSAY and FICTION

| ESSAY | FICTION |
|---|---|
| passive constructions | active verbs |
| modal verbs with the meaning of volition and obligation | |
| to-infinitives | |
| prepositional phrase | adjectives and adj. phrases |
| demonstrative reference | personal reference (1st pers. sg.) |
| additive conj. relations | adversative conj. relations |

# Summary and Discussion

UNIVERSITÄT DES SAARLANDES

- experiment: use automatic text classification techniques to analyse variation in English-German translations
- discriminate between different registers and different translation methods
- classification performs better on register ⇒ dimension of register is stronger
- level out discriminative features (“unknown” features)
- top features for register classification differ from those for method classification
- need for more detailed interpretation
- further algorithms?
- more data?

# Thank you!

Questions? Comments? Suggestions?

**e.lapshinova@mx.uni-saarland.de**
marcos.zampieri@uni-saarland.de

Babych, B., Hartley, A., and Sharoff, S. (2004).
Modelling legitimate translation variation for automatic evaluation of mt quality.
In *Proceedings of LREC-2004*, volume Vol. 3.

Baker, M. (1993).
Corpus linguistics and translation studies: Implications and applications.
In Baker M., G. F. and Tognini-Bonelli, E., editors, *Text and Technology: in Honour of John Sinclair*, pages 233–250. Benjamins, Amsterdam.

Baker, M. (1995).
Corpora in translation studies: An overview and some suggestions for future research.
*Target*, 7(2):223–243.

Biber, D. (1988).
*Variation Across Speech and Writing*.
Cambridge University Press, Cambridge.

Biber, D. (1995).
*Dimensions of Register Variation. A Cross Linguistic Comparison*.
Cambridge University Press, Cambridge.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999).
*Longman Grammar of Spoken and Written English*.
Longman, Harlow.

Corston-Oliver, S., Gamon, M., and Brockett, C. (2001).
A machine learning approach to the automatic evaluation of machine translation.
In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 148–155. Association for Computational Linguistics.

Delaere, I. and Sutter, G. D. (2013).
Applying a multidimensional, register-sensitive approach to visualize normalization in translated and non-translated Dutch.

Diwersy, S., Evert, S., and Neumann, S. (2014).

A semi-supervised multivariate approach to the study of language variation.
*Linguistic Variation in Text and Speech, within and across Languages.*

El-Haj, M., Rayson, P., and Hall, D. (2014).
Language independent evaluation of translation style and consistency: Comparing human and machine translations of camus' novel "the stranger".

Fishel, M., Sennrich, R., Popovic, M., and Bojar, O. (2012).
Terrorcat: a translation error categorization-based mt quality metric.
In *7th Workshop on Statistical Machine Translation*.

Gellerstam, M. (1986).
Translationese in Swedish novels translated from English.
In Wollin, L. and Lindquist, H., editors, *Translation Studies in Scandinavia*, pages 88–95. CWK Gleerup, Lund.

Halliday, M. (2004).
*An Introduction to Functional Grammar*.
Arnold, London.

Halliday, M. and Hasan, R. (1989).
*Language, context and text: Aspects of language in a social-semiotic perspective*.
Oxford University Press, Oxford.

Hansen-Schirra, S., Neumann, S., and Steiner, E. (2012).
*Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*.
de Gruyter, Berlin, New York.

House, J. (1997).
*Translation Quality Assessment. A Model Revisited*.
Günther Narr, Tübingen.

House, J. (2014).
*Translation Quality Assessment. Past and Present*.
Routledge.

Ilisei, I., Inkpen, D., Pastor, G. C., and Mitkov, R. (2010).
Identification of translationese: A machine learning approach.
In *Computational Linguistics and Intelligent Text Processing*, pages 503–511. Springer.

Irvine, A. and Callison-Burch, C. (2014).
Using comparable corpora to adapt MT models to new domains.
In *Proceedings of the ACL Workshop on Statistical Machine Translation (WMT)*.

Irvine, A., Morgan, J., Carpuat, M., III, H. D., and Munteanu, D. S. (2013).
Measuring machine translation errors in new domains.
*TACL*, 1:429–440.

Kibriya, A., Frank, E., Pfahringer, B., and Holmes, G. (2004).
Multinomial naive bayes for text categorization revisited.
In *Proceedings of the Australian Conference on Artificial Intelligence*, pages 488–499.

Lapshinova-Koltunski, E. (to appear 2015).
Linguistic features in translation varieties: Corpus-based analysis.
In De Sutter, G., Delaere, I., and Lefer, M.-A., editors, *New Ways of Analysing Translational Behaviour in Corpus-Based Translation Studies*, TILSM. Mouton de Gruyter.

Lapshinova-Koltunski, E. and Vela, M. (submitted).
Comparable corpora as a measure for 'registerness' of translations.
*Natural Language Engineering*.
Special Issue on Machine Translation Using Comparable Corpora".

Laranjeira, B., Moreira, V., Villavicencio, A., Ramisch, C., and Finatto, M. J. (2014).
Comparing the quality of focused crawlers and of the translation resources obtained from them.
In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).

Neumann, S. (2013).
*Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German.*

De Gruyter Mouton, Berlin, Boston.

Papineni, K., Roukus, S., Ward, T., and Zhu, W.-J. (2002).
BLEU: a method for automatic evaluation of machine translation.
In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318.

Popović, M. and Burchardt, A. (2011).
From human to automatic error classification for machine translation output.
In *15th International Conference of the European Association for Machine Translation (EAMT-2011)*, Leuven, Belgium. European Association for Machine Translation.

Popovic, M. and Ney, H. (2011).
Towards automatic error analysis of machine translation output.
*Computational Linguistics*, 37(4):657–688.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985).
*A Comprehensive Grammar of the English Language*.
Longman, London.

Steiner, E. (1996).
An extended register analysis as a form of text analysis for translation.
In Wotjak, G. and Schmidt, H., editors, *Modelle der Translation – Models of Translation*, pages 235–256. Leipziger Schriften zur Kultur-, Literatur-, Sprach- und Übersetzungswissenschaft, Leipzig.

Steiner, E. (2004).
*Translated Texts. Properties, Variants, Evaluations*.
Peter Lang Verlag, Frankfurt/M.

Sutter, G. D., Delaere, I., and Plevoets, K. (2012).
Lexical lectometry in corpus-based translation studies: Combining profile-based correspondence analysis and logistic regression modeling.
In Oakes, M. P. and Meng, J., editors, *Quantitative Methods in Corpus-based Translation Studies: a Practical Guide to Descriptive Translation Research*, volume 51, pages 325–345. John Benjamins Publishing Company, Amsterdam, The Netherlands.

Volansky, V., Ordan, N., and Wintner, S. (2011).
More human or more translated? original texts vs. human and machine translations.
In *Proceedings of the 11th Bar-Ilan Symposium on the Foundations of AI With ISCOL (Israeli Seminar on Computational Linguistics)*.

White, J. S. (1994).
The ARPA MT evaluation methodologies: Evolution, lessons, and further approaches.
In *Proceedings of the 1994 Conference of the Association for Machine Translation in the Americas*, pages 193–205.

Zampieri, M., Gebre, B. G., and Diwersy, S. (2013).
N-gram language models and POS distribution for the identification of Spanish varieties.
In *Proceedings of TALN2013*, pages 580–587, Sable d'Olonne, France.