

Four explanatory variables in an empirical model of translation

Tatiana Serbina
RWTH Aachen University
Mario Bisiada
Universitat Pompeu Fabra
Stella Neumann
RWTH Aachen University

TRICKLET Conference 2022, 20.05.2022

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:
 - language

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:
 - language
 - register

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:
 - language
 - register
 - translation status

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:
 - language
 - register
 - translation status
 - editorial intervention

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:
 - language
 - register
 - translation status
 - editorial intervention
- Research question: what is the role of editorial intervention in this combination of factors?

Aim of the study

- Assessment of four explanatory variables for an empirical model of translation:
 - language
 - register
 - translation status
 - editorial intervention
- Research question: what is the role of editorial intervention in this combination of factors?
- Method: Geometric Multivariate Analysis, GMA (Diwersy et al. 2014; Evert & Neumann 2017)

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing
 - socio-cultural and political factors

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing
 - socio-cultural and political factors
 - the translation status of a text

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing
 - socio-cultural and political factors
 - the translation status of a text
 - the text-revision process (editing, proof-reading, etc.)

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing
 - socio-cultural and political factors
 - the translation status of a text
 - the text-revision process (editing, proof-reading, etc.)
- Using a general linguistic theory for testing such models is essential (e.g. Neumann 2013; Halverson & Kotze 2022)

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing
 - socio-cultural and political factors
 - the translation status of a text
 - the text-revision process (editing, proof-reading, etc.)
- Using a general linguistic theory for testing such models is essential (e.g. Neumann 2013; Halverson & Kotze 2022)
 - to expand the explanatory potential of empirical studies

Translation models

Proposals in the literature

- An empirical model of translation should take into account a variety of different factors (Steiner 2001; Hansen-Schirra et al. 2017; Carl & Schaeffer 2017), such as:
 - the language pair
 - register differences within each language
 - cognitive processing
 - socio-cultural and political factors
 - the translation status of a text
 - the text-revision process (editing, proof-reading, etc.)
- Using a general linguistic theory for testing such models is essential (e.g. Neumann 2013; Halverson & Kotze 2022)
 - to expand the explanatory potential of empirical studies
 - to ensure a coherent framework of analysis, for instance while combining different data types

Translation models

Cognitive Linguistics (Langacker 2008; Goldberg 2006)

- linguistic features – form-meaning pairings, also referred to as constructions, varying in size and levels of abstractness

Translation models

Cognitive Linguistics (Langacker 2008; Goldberg 2006)

- linguistic features – form-meaning pairings, also referred to as constructions, varying in size and levels of abstractness
- characterized by different degrees of entrenchment, i.e. the degree of automaticity associated with production and comprehension of constructions

Translation models

Cognitive Linguistics (Langacker 2008; Goldberg 2006)

- linguistic features – form-meaning pairings, also referred to as constructions, varying in size and levels of abstractness
- characterized by different degrees of entrenchment, i.e. the degree of automaticity associated with production and comprehension of constructions
- higher frequency, attested through corpus studies → higher level of cognitive entrenchment

Translation models

Cognitive Linguistics (Langacker 2008; Goldberg 2006)

- linguistic features – form-meaning pairings, also referred to as constructions, varying in size and levels of abstractness
- characterized by different degrees of entrenchment, i.e. the degree of automaticity associated with production and comprehension of constructions
- higher frequency, attested through corpus studies → higher level of cognitive entrenchment
- every additional encounter of a linguistic unit strengthens its cognitive representation → subsequent production in a given context is more likely

Explanatory factors

Language

- the lexico-grammatical features of texts are determined by the corresponding language

Explanatory factors

Language

- the lexico-grammatical features of texts are determined by the corresponding language
- numerous contrastive studies, such as König & Gast (2018) or, from the perspective of Construction Grammar, Boas (2010)

Explanatory factors

Language

- the lexico-grammatical features of texts are determined by the corresponding language
- numerous contrastive studies, such as König & Gast (2018) or, from the perspective of Construction Grammar, Boas (2010)
- corresponding constructions (Leino 2010) → form-function pairings in different languages with similarities on both formal and functional poles

Explanatory factors

Language

- the lexico-grammatical features of texts are determined by the corresponding language
- numerous contrastive studies, such as König & Gast (2018) or, from the perspective of Construction Grammar, Boas (2010)
- corresponding constructions (Leino 2010) → form-function pairings in different languages with similarities on both formal and functional poles
- corresponding constructions could be entrenched to a different degree → more or less likely to be selected in a given language

Explanatory factors

Register

- linguistic features vary in their frequency depending on registers, i.e. situational contexts

Explanatory factors

Register

- linguistic features vary in their frequency depending on registers, i.e. situational contexts
- some empirical findings on register variation in translation (for a review see Neumann (2021)):

Explanatory factors

Register

- linguistic features vary in their frequency depending on registers, i.e. situational contexts
- some empirical findings on register variation in translation (for a review see Neumann (2021)):
 - Neumann (2013) has shown that for culturally comparable registers there are only minor differences in the feature distribution found in translations and the comparable originals

Explanatory factors

Register

- linguistic features vary in their frequency depending on registers, i.e. situational contexts
- some empirical findings on register variation in translation (for a review see Neumann (2021)):
 - Neumann (2013) has shown that for culturally comparable registers there are only minor differences in the feature distribution found in translations and the comparable originals
 - Delaere (2015) suggests that register plays an important role in norm conformity in translated language

Explanatory factors

Translation status

- the uniqueness of translation as a linguistic activity should not be overstated →

Explanatory factors

Translation status

- the uniqueness of translation as a linguistic activity should not be overstated →
 - the same cognitive mechanisms as in language production in general are at play (Halverson & Kotze 2022)

Explanatory factors

Translation status

- the uniqueness of translation as a linguistic activity should not be overstated →
 - the same cognitive mechanisms as in language production in general are at play (Halverson & Kotze 2022)
 - translation as a type of constrained communication (Kotze 2022)

Explanatory factors

Translation status

- the uniqueness of translation as a linguistic activity should not be overstated →
 - the same cognitive mechanisms as in language production in general are at play (Halverson & Kotze 2022)
 - translation as a type of constrained communication (Kotze 2022)
- linguistic profiles depend on language- and register-specific entrenchment but some distributions specific to translations → research on translation properties (overview in de Sutter & Lefer 2020) and translationese (Volansky et al. 2015)

Explanatory factors

Translation status

- the uniqueness of translation as a linguistic activity should not be overstated →
 - the same cognitive mechanisms as in language production in general are at play (Halverson & Kotze 2022)
 - translation as a type of constrained communication (Kotze 2022)
- linguistic profiles depend on language- and register-specific entrenchment but some distributions specific to translations → research on translation properties (overview in de Sutter & Lefer 2020) and translationese (Volansky et al. 2015)
- cognitive representation and linguistic behavior shaped by the 'practice of translating texts, of particular kinds, for particular purposes and for particular clients' (Halverson & Kotze 2022: 72)

Explanatory factors

Editorial intervention

- published texts are often edited

Explanatory factors

Editorial intervention

- published texts are often edited
- editorial intervention could be seen as normal and 'not something that the linguist should wish away as noise or change imposed on "authentic data"' (Kruger & van Rooy 2018: 220)

Explanatory factors

Editorial intervention

- published texts are often edited
- editorial intervention could be seen as normal and 'not something that the linguist should wish away as noise or change imposed on "authentic data"' (Kruger & van Rooy 2018: 220)
- → linguistic features present in the final published version contribute to entrenchment of these features (Kruger & van Rooy 2018: 220)

Explanatory factors

Editorial intervention

- published texts are often edited
- editorial intervention could be seen as normal and 'not something that the linguist should wish away as noise or change imposed on "authentic data"' (Kruger & van Rooy 2018: 220)
- → linguistic features present in the final published version contribute to entrenchment of these features (Kruger & van Rooy 2018: 220)
- still essential to acknowledge and assess the additional workflow stages, particularly for a successful integration of product and process research (Serbina & Neumann 2022: 142)

Interaction between editorial intervention & translation status

Kruger (2012): translation universals in 'mediated discourse'

- normalisation, explicitation & simplification in 'mediated' (translated, edited) and 'unmediated' (unedited) text

Interaction between editorial intervention & translation status

Kruger (2012): translation universals in 'mediated discourse'

- normalisation, explicitation & simplification in 'mediated' (translated, edited) and 'unmediated' (unedited) text
- 1.2 million word corpus
 - translations Afrikaans–English
 - edited English texts
 - unedited English texts

Interaction between editorial intervention & translation status

Kruger (2012): translation universals in 'mediated discourse'

- normalisation, explicitation & simplification in 'mediated' (translated, edited) and 'unmediated' (unedited) text
- 1.2 million word corpus
 - translations Afrikaans–English
 - edited English texts
 - unedited English texts
- academic, instructional, popular and reportage texts

Interaction between editorial intervention & translation status

Kruger (2012): translation universals in 'mediated discourse'

- normalisation, explicitation & simplification in 'mediated' (translated, edited) and 'unmediated' (unedited) text
- 1.2 million word corpus
 - translations Afrikaans–English
 - edited English texts
 - unedited English texts
- academic, instructional, popular and reportage texts
- no evidence of shared 'mediation effect'
 - translators favour 'explicit and standardised language'
 - editors 'introduce collocational variety'

Interaction between editorial intervention & translation status

Kruger (2012): translation universals in 'mediated discourse'

- normalisation, explicitation & simplification in 'mediated' (translated, edited) and 'unmediated' (unedited) text
- 1.2 million word corpus
 - translations Afrikaans–English
 - edited English texts
 - unedited English texts
- academic, instructional, popular and reportage texts
- no evidence of shared 'mediation effect'
 - translators favour 'explicit and standardised language'
 - editors 'introduce collocational variety'

~> drawback: no edited translations studied

Interaction between editorial intervention & translation status

Bisiada (2017): replicating Kruger's study for German

1 Explicitation

- Frequency of use of *dass* ('that')
- Frequency of linking & pronominal adverbs
- Conjunction vs preposition ratio
- ...

Interaction between editorial intervention & translation status

Bisiada (2017): replicating Kruger's study for German

1 Explicitation

- Frequency of use of *dass* ('that')
- Frequency of linking & pronominal adverbs
- Conjunction vs preposition ratio
- ...

2 Normalisation/conservatism

- Degree of unconventional language use
- ...

Interaction between editorial intervention & translation status

Bisiada (2017): replicating Kruger's study for German

1 Explicitation

- Frequency of use of *dass* ('that')
- Frequency of linking & pronominal adverbs
- Conjunction vs preposition ratio
- ...

2 Normalisation/conservatism

- Degree of unconventional language use
- ...

3 Simplification

- Lexical diversity
- Mean word and sentence length

Interaction between editorial intervention & translation status

Findings (Bisiada 2017)

- 1 Explicitation
No difference between corpora

Interaction between editorial intervention & translation status

Findings (Bisiada 2017)

1 Explicitation

No difference between corpora

2 Normalisation/conservatism

Difference between translated and non-translated language

↔ non-translated texts make more use of unconventional/creative language

Interaction between editorial intervention & translation status

Findings (Bisiada 2017)

1 Explicitation

No difference between corpora

2 Normalisation/conservatism

Difference between translated and non-translated language

↔ non-translated texts make more use of unconventional/creative language

3 Simplification

Difference between manuscripts and published texts

↔ editors' influence strongest in this aspect

Interaction between editorial intervention & translation status

Findings (Bisiada 2017)

1 Explicitation

No difference between corpora

2 Normalisation/conservatism

Difference between translated and non-translated language

↔ non-translated texts make more use of unconventional/creative language

3 Simplification

Difference between manuscripts and published texts

↔ editors' influence strongest in this aspect

⇒ some evidence for normalisation universal, little evidence for “mediation universals”

Editorial intervention: individual features and multivariate analysis

- some phenomena like sentence splitting are caused by both translators and editors (Bisiada 2016; 2018b)

Editorial intervention: individual features and multivariate analysis

- some phenomena like sentence splitting are caused by both translators and editors (Bisiada 2016; 2018b)
- translators and editors are guided by different goals

Editorial intervention: individual features and multivariate analysis

- some phenomena like sentence splitting are caused by both translators and editors (Bisiada 2016; 2018b)
- translators and editors are guided by different goals
- they both make extensive changes to nominalisations (Bisiada 2018a,c)

Editorial intervention: individual features and multivariate analysis

- some phenomena like sentence splitting are caused by both translators and editors (Bisiada 2016; 2018b)
- translators and editors are guided by different goals
- they both make extensive changes to nominalisations (Bisiada 2018a,c)
- editors eliminate passive constructions from translations, especially when the verb is in the past tense (Bisiada 2019)

Editorial intervention: individual features and multivariate analysis

- some phenomena like sentence splitting are caused by both translators and editors (Bisiada 2016; 2018b)
- translators and editors are guided by different goals
- they both make extensive changes to nominalisations (Bisiada 2018a,c)
- editors eliminate passive constructions from translations, especially when the verb is in the past tense (Bisiada 2019)
- previous multivariate analysis considering only German originals and translations did not indicate a profound effect of editorial intervention (Serbina et al. 2021) – calling for more extensive analysis across languages

Data: Overview of the data sample

Corpora: Harvard Business Corpus, HBC (Bisiada 2018a) and CroCo Corpus (Hansen-Schirra et al. 2012)

Corpus	Translation Status	Register	Size in words
HBC	ST EN	Business	106,035
HBC	Manuscript T DE	Business	112,810
HBC	Published T DE	Business	106,958
CroCo	ST EN	Share, Popsci	62,952
CroCo	Published T DE	Share, Popsci	61,791
CroCo	ST DE	Share, Popsci, Speech, Essay	124,926

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)
- English: CLAWS tagger (Garside & Smith 1997) using the CLAWS 7 tagset for English data

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)
- English: CLAWS tagger (Garside & Smith 1997) using the CLAWS 7 tagset for English data

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)
- English: CLAWS tagger (Garside & Smith 1997) using the CLAWS 7 tagset for English data

Feature extraction

- cqp script (Fest et al. 2019; Neumann & Evert 2021)

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)
- English: CLAWS tagger (Garside & Smith 1997) using the CLAWS 7 tagset for English data

Feature extraction

- cqp script (Fest et al. 2019; Neumann & Evert 2021)
- based on the lexico-grammatical features developed by Neumann (2013)

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)
- English: CLAWS tagger (Garside & Smith 1997) using the CLAWS 7 tagset for English data

Feature extraction

- cqp script (Fest et al. 2019; Neumann & Evert 2021)
- based on the lexico-grammatical features developed by Neumann (2013)
- normalized to an appropriate unit of measurement (e.g. nominalizations/word or passive/finite verb)

Methods

POS tagging

- German: TreeTagger (Schmid 1994) using STTS tagset (Schiller et al. 1999)
- English: CLAWS tagger (Garside & Smith 1997) using the CLAWS 7 tagset for English data

Feature extraction

- cqp script (Fest et al. 2019; Neumann & Evert 2021)
- based on the lexico-grammatical features developed by Neumann (2013)
- normalized to an appropriate unit of measurement (e.g. nominalizations/word or passive/finite verb)
- final set contains 37 features

Methods

Geometric Multivariate Analysis

- GMA procedure (Diwersy et al. 2014; Evert & Neumann 2017) consisting of visual inspection and linguistic interpretation of PCA and LDA

Methods

Geometric Multivariate Analysis

- GMA procedure (Diwersy et al. 2014; Evert & Neumann 2017) consisting of visual inspection and linguistic interpretation of PCA and LDA
- data pre-processing and analysis – R scripts (based on scripts by Stephanie Evert, Friedrich-Alexander Universität Erlangen-Nürnberg)

Methods

Geometric Multivariate Analysis

- GMA procedure (Diwersy et al. 2014; Evert & Neumann 2017) consisting of visual inspection and linguistic interpretation of PCA and LDA
- data pre-processing and analysis – R scripts (based on scripts by Stephanie Evert, Friedrich-Alexander Universität Erlangen-Nürnberg)
- every text is projected into a multi-dimensional feature space as a feature vector comprising the log-transformed z-scores of 37 indicators

Methods

Geometric Multivariate Analysis

- GMA procedure (Diwersy et al. 2014; Evert & Neumann 2017) consisting of visual inspection and linguistic interpretation of PCA and LDA
- data pre-processing and analysis – R scripts (based on scripts by Stephanie Evert, Friedrich-Alexander Universität Erlangen-Nürnberg)
- every text is projected into a multi-dimensional feature space as a feature vector comprising the log-transformed z-scores of 37 indicators
- Euclidean distances between the feature vectors are assumed to represent meaningful differences between texts

Principle Component Analysis

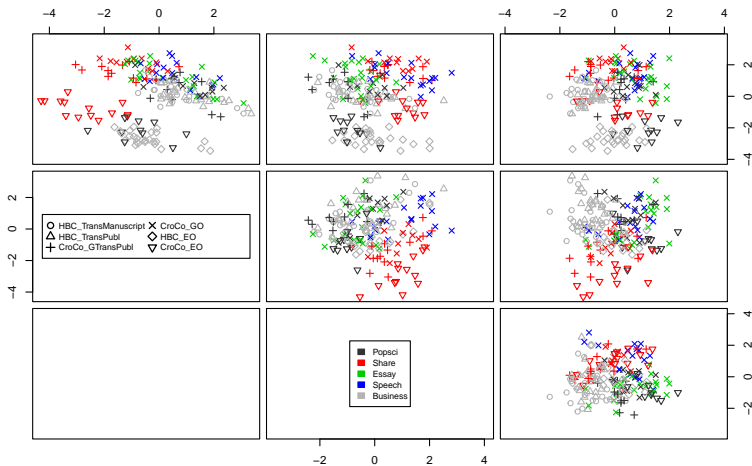


Figure: Scatterplot matrix of the first four PCA dimensions

Principle Component Analysis

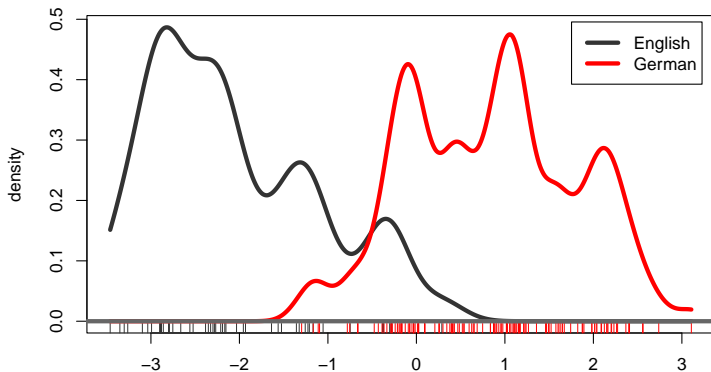


Figure: Discriminant plot of the 1st PCA dim according to language

Principle Component Analysis

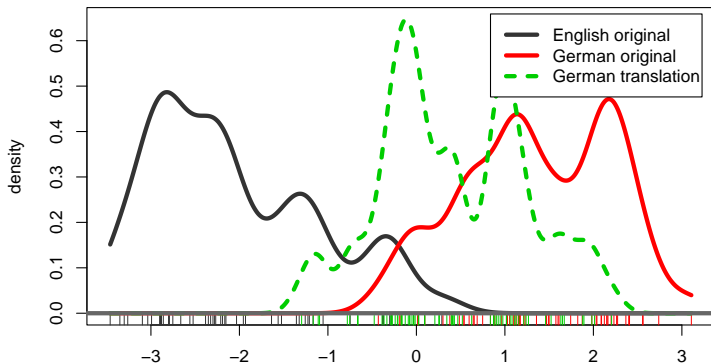


Figure: Discriminant plot of the 1st PCA dim according to language and translation status

Principle Component Analysis

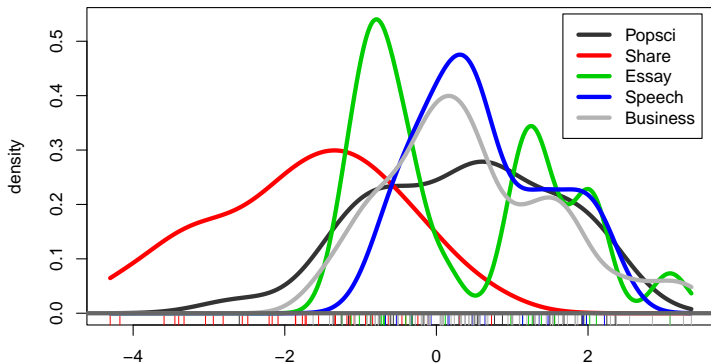


Figure: Discriminant plot of the 2nd PCA dim according to register

Principle Component Analysis

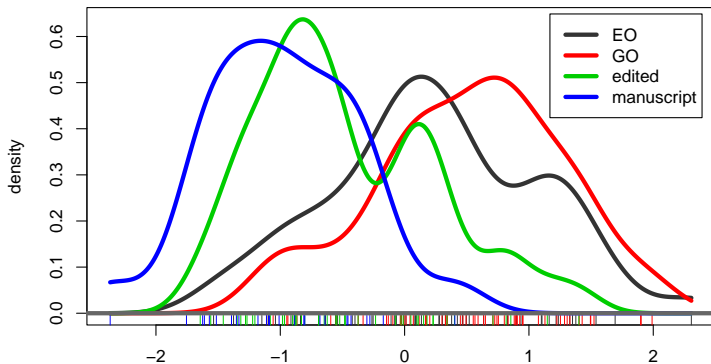


Figure: Discriminant plot of the 4th PCA dim according to editorial intervention

Linear Discriminant Analysis

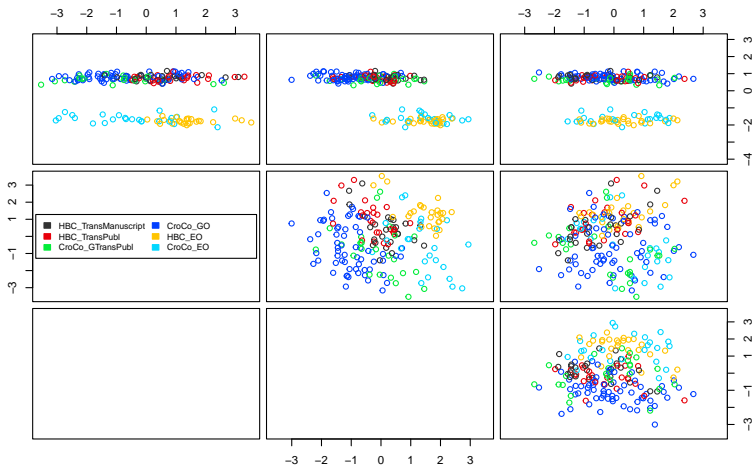


Figure: LDA with language as discriminant

Linear Discriminant Analysis

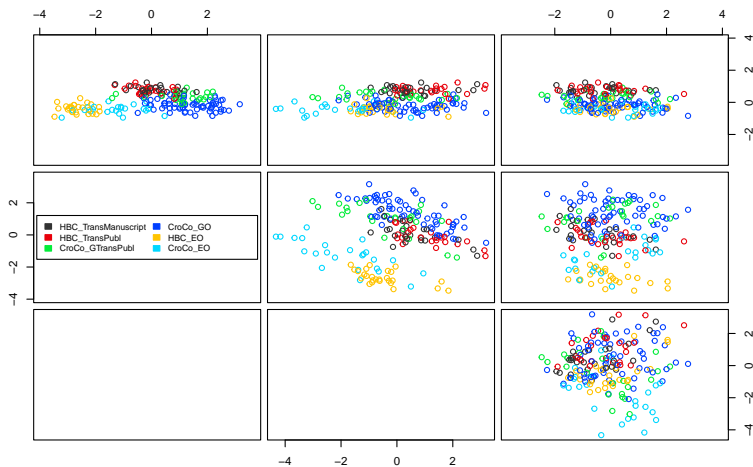


Figure: LDA with translation status as discriminant

Interaction of the four explanatory variables

- language and translation status seem to account for most of the variation (see also previous GMA studies on the role and interaction of these two factors (Diwersy et al. 2014; Evert & Neumann 2017))

Interaction of the four explanatory variables

- language and translation status seem to account for most of the variation (see also previous GMA studies on the role and interaction of these two factors (Diwersy et al. 2014; Evert & Neumann 2017))
- in contrast to Serbina et al. (2021), register effect is less profound

Interaction of the four explanatory variables

- language and translation status seem to account for most of the variation (see also previous GMA studies on the role and interaction of these two factors (Diwersy et al. 2014; Evert & Neumann 2017))
- in contrast to Serbina et al. (2021), register effect is less profound
- considering the whole range of linguistic features and other explanatory variables, editorial intervention does not appear to contribute much to linguistic variation

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):
 - focus only on a limited number of linguistic features, such as passive constructions or nominalisations

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):
 - focus only on a limited number of linguistic features, such as passive constructions or nominalisations
 - translators are aware of the specific editorial guidelines → could to some extent adjust or self-revise their texts

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):
 - focus only on a limited number of linguistic features, such as passive constructions or nominalisations
 - translators are aware of the specific editorial guidelines → could to some extent adjust or self-revise their texts
- norms-as-conventionalization and norms-as-legitimization (Kruger & van Rooy 2017):

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):
 - focus only on a limited number of linguistic features, such as passive constructions or nominalisations
 - translators are aware of the specific editorial guidelines → could to some extent adjust or self-revise their texts
- norms-as-conventionalization and norms-as-legitimization (Kruger & van Rooy 2017):
 - legitimization or enforcement of certain norms by the publishing house

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):
 - focus only on a limited number of linguistic features, such as passive constructions or nominalisations
 - translators are aware of the specific editorial guidelines → could to some extent adjust or self-revise their texts
- norms-as-conventionalization and norms-as-legitimization (Kruger & van Rooy 2017):
 - legitimization or enforcement of certain norms by the publishing house
 - the more frequently translators encounter the constructions used in edited published texts, the more entrenched they get → a process of conventionalization (Kruger & van Rooy 2017)

Possible reasons

- a tendency for linguistic features of English originals to 'shine through' in German translations → priming/SL magnetism (Halverson 2017)
- the role of editorial guidelines (Bisiada 2016; 2018b):
 - focus only on a limited number of linguistic features, such as passive constructions or nominalisations
 - translators are aware of the specific editorial guidelines → could to some extent adjust or self-revise their texts
- norms-as-conventionalization and norms-as-legitimization (Kruger & van Rooy 2017):
 - legitimization or enforcement of certain norms by the publishing house
 - the more frequently translators encounter the constructions used in edited published texts, the more entrenched they get → a process of conventionalization (Kruger & van Rooy 2017)
 - entrenchment of certain norms may be limited to a particular context

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium
 - editorial intervention – low

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium
 - editorial intervention – low
- however, as shown in the previous studies, each factor in its own right has a significant effect on individual linguistic features

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium
 - editorial intervention – low
- however, as shown in the previous studies, each factor in its own right has a significant effect on individual linguistic features
- future research should

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium
 - editorial intervention – low
- however, as shown in the previous studies, each factor in its own right has a significant effect on individual linguistic features
- future research should
 - include further explanatory factors

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium
 - editorial intervention – low
- however, as shown in the previous studies, each factor in its own right has a significant effect on individual linguistic features
- future research should
 - include further explanatory factors
 - study a greater variety of languages

Conclusion and outlook

- in the current combination of factors, the respective contribution of each factor to the overall linguistic profile appears to be the following:
 - language in combination with translation profile – high
 - register – medium
 - editorial intervention – low
- however, as shown in the previous studies, each factor in its own right has a significant effect on individual linguistic features
- future research should
 - include further explanatory factors
 - study a greater variety of languages
 - consider larger datasets with rich meta-data (similar to the MUST corpus and DPC 2.0)

Thank you for your attention!

References I



Bisiada, Mario. 2016. "Lösen Sie Schachtelsätze möglichst auf": The impact of editorial guidelines on sentence splitting in German business article translations. *Applied Linguistics* 37(3). 354–376. <https://doi.org/10.1093/applin/amu035>.



Bisiada, Mario. 2017. Universals of editing and translation. In Silvia Hansen-Schirra, Oliver Czulo, Sascha Hofmann & Bernd Meyer (eds.), *Empirical modelling of translation and interpreting*, 241–275. Berlin: Language Science Press.



Bisiada, Mario. 2018a. Editing nominalisations in English–German translation: When do editors intervene? *The Translator* 24(1). 35–49. <https://doi.org/10.1080/13556509.2017.1301847>.



Bisiada, Mario. 2018b. The editor's invisibility: Analysing editorial intervention in translation. *Target* 30(2). 288–309. <https://doi.org/10.1075/target.16116.bis>.



Bisiada, Mario. 2018c. Translation and editing: A study of editorial treatment of nominalisations in draft translations. *Perspectives: Studies in Translation Theory and Practice* 26(1). 24–38. <https://doi.org/10.1080/0907676X.2017.1290121>.



Bisiada, Mario. 2019. Translated language or edited language? A study of passive constructions in translation manuscripts and their published versions. *Across Languages and Cultures* 20(1). 35–56. <https://doi.org/10.1556/084.2019.20.1.2>.



Boas, Hans C. (ed.). 2010. *Contrastive studies in construction grammar*. Amsterdam: Benjamins.

References II



Carl, Michael & Moritz Schaeffer. 2017. Sketch of a noisy channel model for the translation process. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting*, 71–116. Berlin: Language Science Press.



Delaere, Isabelle. 2015. *Do translations walk the line? visually exploring translated and non-translated texts in search of norm conformity*. Ghent. (Doctoral dissertation).



Diwersy, Sascha, Stefan Evert & Stella Neumann. 2014. A weakly supervised multivariate approach to the study of language variation. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: linguistic variation in text and speech*, 174–204. Berlin: de Gruyter.



Evert, Stefan & Stella Neumann. 2017. The impact of translation direction on characteristics of translated texts: a multivariate analysis for English and German. In Gert de Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies: new theoretical and methodological traditions*, 47–80. Berlin: Mouton de Gruyter.



Fest, Jennifer, Arndt Heilmann, Oliver Hohlfeld, Stella Neumann, Helge Reelfs, Marco Schmitt & Alina Vogelgesang. 2019. Determining response-generating contexts on microblogging platforms. In *Proceedings of the 15th conference on natural language processing (konvens)*, 171–182.



Garside, Roger & Nicholas Smith. 1997. A hybrid grammatical tagger: claws4. In Roger Garside, Geoffrey Leech & Anthony McEnery (eds.), *Corpus annotation: linguistic information from computer text corpora*, 102–121. London: Longman.

References III



Goldberg, Adele. 2006. *Constructions at work: the nature of generalizations in language*. Oxford: Oxford Univ. Press.



Halverson, Sandra L. 2017. Gravitational pull in translation: testing a revised model. In Gert de Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies: new theoretical and methodological traditions*, 9–45. Berlin: Mouton de Gruyter.



Halverson, Sandra L. & Haidee Kotze. 2022. Sociocognitive constructs in translation and interpreting studies (tis): do we really need concepts like norms and risk when we have a comprehensive usage-based theory of language? In Sandra L. Halverson & Álvaro Marín García (eds.), *Contesting epistemologies in cognitive translation and interpreting studies*, 51–79. New York: Routledge.



Hansen-Schirra, Silvia, Stella Neumann & Erich Steiner. 2012. *Cross-linguistic corpora for the study of translations: insights from the language pair english-german*. Berlin: de Gruyter.



Hansen-Schirra, Silvia, Jean Nitzke & Katharina Oster. 2017. Predicting cognate translation. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting*, 3–22. Berlin: Language Science Press.



König, Ekkehard & Volker Gast. 2018. *Understanding English-German contrasts*. 4th edn. Berlin: Erich Schmidt Verlag.



Kotze, Haidee. 2022. Translation as constrained communication: principles, concepts and methods. In Sylviane Granger & Marie-Aude Lefer (eds.), *Extending the scope of corpus-based translation studies*, 67–97. London: Bloomsbury.

References IV



Kruger, Haidee. 2012. A corpus-based study of the mediation effect in translated and edited language. *Target* 24(2). 355–388. <https://doi.org/10.1075/target.24.2.07kru>.



Kruger, Haidee & Bertus van Rooy. 2017. Editorial practice and the progressive in black south african english. *World Englishes* 36. 20–41.



Kruger, Haidee & Bertus van Rooy. 2018. Register variation in written contact varieties: a multidimensional analysis. *English World-Wide: A Journal of Varieties of English* 39(2). 214–242.



Langacker, Ronald W. 2008. *Cognitive grammar: a basic introduction*. Oxford: Oxford Univ. Press.



Leino, Jaakko. 2010. Results, cases, and constructions: argument structure constructions in english and finnish. In Hans C. Boas (ed.), *Contrastive studies in construction grammar*, 103–135. Amsterdam: Benjamins.



Neumann, Stella. 2013. *Contrastive register variation: a quantitative approach to the comparison of english and german*. Berlin: de Gruyter.



Neumann, Stella. 2021. Register and translation. In Mira Kim, Jeremy Munday, Zhenhua Wang & Pin Wang (eds.), *Systemic functional linguistics and translation studies*, 65–82. London: Bloomsbury.



Neumann, Stella & Stefan Evert. 2021. A register variation perspective on varieties of english. In Elena Seoane & Douglas Biber (eds.), *Corpus based approaches to register variation*, 143–178. Berlin: de Gruyter.

References V



Schiller, Anne, Simone Teufel, Christine Stöckert & Christine Thielen. 1999. *Guidelines für das tagging deutscher textcorpora mit stts*. Universität Stuttgart: Universität Stuttgart.
<http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.



Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing*.



Serbina, Tatiana, Mario Bisiada & Stella Neumann. 2021. Linguistic profiles of translation manuscripts and edited translations. In *Proceedings for the first workshop on modelling translation: translology in the digital age*, 34–45. Association for Computational Linguistics.
<https://aclanthology.org/2021.motra-1.4>.



Serbina, Tatiana & Stella Neumann. 2022. Translation product and process data: a happy marriage or worlds apart? In Sandra L. Halverson & Álvaro Marín García (eds.), *Contesting epistemologies in cognitive translation and interpreting studies*, 131–152. New York: Routledge.



Steiner, Erich. 2001. Translations English–German: Investigating the relative importance of systemic contrasts and of the text-type ‘translation’. *SPRIKreports* 7. 1–48.



de Sutter, Gert & Marie-Aude Lefer. 2020. On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28(1). 1–23.



Volansky, Vered, Noam Ordan & Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities* 30(1). 98–118.