

# Outline

- Aim of the procedural method: to support translation through LSP sample text corpora
- Description of the method
- Case study: testing the suitability of the method in 5 steps
- Conclusions on small-scale LSP sample text corpora

# Aim of the procedural method

- The aim of the method is to support the translation of texts with a moderate degree of specialisation
  - general newspaper articles on specialised topics
  - marketing texts
  - specialised magazine articles for a broad public
  - institutional texts for the citizens
- The method is intended to complement usual reference resources for translators
  - LGP dictionaries
  - LSP dictionaries (if available)
  - online documentation: parallel (i.e. functionally comparable) texts, translations (e.g. multilingual websites)
  - termbases in TMs

Most frequent problems in translating texts with a moderate degree of specialisation:

LSP dictionaries do not offer sufficient information:

- outdated lexicographic resources
- no domain-specific resources (e.g. technical domain)
- lack of context-specific equivalents, collocations, pragmatic markers etc.

specialised corpora:

- not publicly available
- domain content is not specific enough
- text type / genre do not match those of the source text

# Description of the method

- Given a source text with a moderate degree of specialisation, to build small-sized corpora in the source/ target language with texts matching the topic, type and genre of the source text and to use these as a complement to traditional reference works/tools
- theoretical frame: DIY/ disposable/ ad hoc/ ... corpora in translation, cf. F. Zanettin et al. (2003-2014: *Corpora in Translator Education*); LSP: legal domain (J. Scott 2012: “Towards professional uptake of DIY electronic corpora in legal genres”)

NEW:

texts with a **low level of specialisation** and **broad topics**, belonging to LSP **less standardised** domains, and in which greater **terminological variation** is to be expected and **corpus compilation** requires an initial effort

- **sample text corpus:**

“a corpus that is designed in order to be representative of a particular language variety (...). It may consist of complete texts or samples taken from parts of texts. In general, the more specialised the language variety, the smaller the sample needs to be” (Baker et al. 2006, *Glossary of Corpus Linguistics*)

# Case study

- source texts: newspaper texts on the topic “Climate change and renewable energies”
- type of translation assignment: passive translation
- language combination: German (L2), Italian (L1)

## Steps:

1. analysis of the source text
2. classification of search items
3. selection of keywords
4. sample text corpus creation
5. sample text corpus analysis and translation

# 1- Textual analysis

- Analysis of the L2 (German) newspaper source texts according to established translation theories (e.g. Christiane Nord 2009 ) and/or text linguistic theories (e.g. Brinker et al. 2000)
- The purpose is to assess:
  - text type (mainly: narrative, argumentative, expository)
  - text genre (news, commentary, review, etc.)
  - text function (mainly: informative)
  - level of specialisation (vertical dimension): relation author-recipient (mainly: texts for a broad public, laymen with prior knowledge of the topic/domain and of most of its terminology)
  - text-specific ratio LSP-LGP

## 2- Search items classification

- classification and clustering of LSP linguistic phenomena that can pose transfer problems when translating into the

L1:

phenomena regarding the reception phase (L2)

e.g. understanding polysemous words, identifying collocations

phenomena regarding the production phase (L1)

e.g. finding context-dependent equivalents, variants of collocations, choosing specific translation procedures such as recategorization, reproducing realia



<p>Noch immer sind 1,3 Milliarden Menschen ohne Zugang zu <b>kommerziell erzeugtem Strom</b>. Dazu kommt die Frage, wie die Unternehmen Strom erzeugen und zu bezahlbaren Preisen anbieten können, ohne dass die <b>Klimaziele</b> geopfert werden. (...)</p>	<p>context-dependent equivalents (PROD)</p>
<p>Der Energiesektor ist nach Angaben der <b>Internationalen Energie-Agentur (IEA)</b> für rund zwei Drittel der <b>weltweiten Treibhausgasemissionen</b> verantwortlich. Die Agentur warnte im Juni, das <b>Ziel der Erderwärmung von nicht mehr als zwei Grad Celsius</b> aus den Augen zu verlieren. Gebe es kein Umdenken, drohe eine Erwärmung um etwa vier Grad bis zum Jahr 2100. (...)</p>	<p>realia (PROD)  reategorization (PROD)  collocations (REC/PROD)</p>
<p>Nicht nur Öl und Gas, auch die erneuerbaren Energien, die Kernkraft, Kohle und neue Energiequellen werden trotz der Herausforderungen künftig eine Rolle beim <b>Energiemix</b> spielen, sind sich die Experten einig.</p>	<p>polysemy (REC)</p>

### 3- Keywords selection

- manual selection of a set of terminological units in the source text (DE) that explicitly characterises its semantic content
- keywords (key terms) can be both single lexical items and word combinations, for instance

DE

Strom N	(power)
Treibhausgasemissionen N	(greenhouse emissions)
Klimaziele N	(climate targets)
Energie N	(energy, power)
Windenergie N	(wind energy)
nachhaltig A	(sustainable)
erneuerbar A	(renewable)
Energiebedarf decken V+N	(to cover power needs)

## 4- Corpus compilation

- keywords (DE) and immediate keyword equivalents (IT) are used to identify comparable texts in
  - A) large online databases with comparable text genres, e.g. LexisNexis
    - + selected text sources
    - restricted access
  - and/or B) web text collections through seed words in CQS, e.g. SketchEngine
    - + large text collections
    - no specific text sources (types/genres)
  - and/or C) institutional parallel corpora, e.g. Europarl
    - + controlled language in source and target texts
    - restricted number of specific topics
- result: 2 comparable corpora with the keywords in different text-internal distributions, non-annotated, each 3m words

## 5- Sample text corpus analysis and translation

- the research items identified during step 2 are analysed by means of the CQS AntConc 3.4.3w (KWIC, collocates lists)
- **reception phase L2:** understanding the text by using the L2 corpus

class	item	findings
polysemy	<i><u>Energiemix</u></i>	2 readings: “power generation mix” / “proportion of different forms of energy used in a country”
collocation (?)	<i><u>kommerziell</u> erzeugter Strom</i>	> not in the cp as a collocation “power generated for commercial purposes”

- **production phase L1**: testing possible equivalents in the L1 corpus (competence control), assessing context-dependent usage

class	item	findings
recategorization (?)	<i>weltweite Treibhausgasemi ssionen</i>	Adj: “mondiale”, “globale” PP: “nel mondo”, “a livello mondiale/ globale”
variants of MWEs (?)	<i>Klimaziele</i>  <i>Ziel der Erderwärmung von... 2 Grad C.</i>	“obiettivi climatici” (+F), “obiettivi di politica climatica”, “obiettivi in materia di clima” ≠ “obiettivi connessi ai cambiamenti climatici”  “obiettivo di un surriscaldamento globale” non idiomatic; a verb is needed “obiettivo di limitare il surriscaldamento...”
realia (?)	<i>Internationale Energie-Agentur (IEA)</i>	“International Energy Agency IEA” / “Agenzia internazionale dell’/per l’energia (AIE)” (> it. version is not official)

Noch immer sind 1,3 Milliarden Menschen ohne Zugang zu **kommerziell erzeugtem Strom**. Dazu kommt die Frage, wie die Unternehmen Strom erzeugen und zu bezahlbaren Preisen anbieten können, ohne dass die **Klimaziele** geopfert werden. (...)

Der Energiesektor ist nach Angaben der **Internationalen Energie-Agentur (IEA)** für rund zwei Drittel der **weltweiten** Treibhausgasemissionen verantwortlich. Die Agentur warnte im Juni, das **Ziel der Erderwärmung** von nicht mehr als zwei Grad Celsius aus den Augen zu verlieren. Gebe es kein Umdenken, drohe eine Erwärmung um etwa vier Grad bis zum Jahr 2100. (...)

Nicht nur Öl und Gas, auch die erneuerbaren Energien, die Kernkraft, Kohle und neue Energiequellen werden trotz der Herausforderungen künftig eine Rolle beim **Energiemix** spielen, sind sich die Experten einig.

Ci sono ancora 1,3 miliardi di persone senza accesso alla **corrente elettrica**. A questo proposito ci si chiede come sia possibile per le imprese produrre elettricità, e offrirla a prezzi contenuti, senza però sacrificare gli **obiettivi climatici**. (...)

Secondo l'**International Energy Agency (IEA)**, il settore energetico è responsabile di due terzi di tutte le emissioni di gas serra **nel mondo**. A giugno la IEA ha messo in guardia dal perdere di vista l'**obiettivo di limitare il surriscaldamento globale** entro i 2°C. Se non si cambia rotta, si rischia un surriscaldamento di 4 gradi entro il 2100. (...)

Gli esperti sono certi che, nonostante le difficoltà, non solo il petrolio e il gas, ma anche le energie rinnovabili, l'energia nucleare, il carbone e le nuove fonti di energia giocheranno un ruolo nell'**assetto energetico** del futuro.

# Conclusions on small-scale LSP sample text corpora

- advantages

- helpful for the intended users: translation students, translators working on long-term domain-specific translations; (improvement of software literacy among translators)
- qualitative point of view: relevant contextual information (e.g. semantic disambiguation, variation)
- useful complement of lexicographic/terminographic resources
- periodic terminology maintenance: updating texts (e.g. newspaper texts)
- data expandability (the corpus is not intended to be static)

- disadvantages

- initially time consuming corpus building
- quantitative evidence (frequency, statistical measures) needs to be relativised to the size and terminological specificity of the corpus
- low probability of finding relevant data on general language items
- (infrequent occurrence of realia)

- outlook:

- need for a greater awareness and understanding of corpus practices among professional translators
- need for further studies on
  - keyword-based LSP corpus building methods
  - statistical data interpretation
  - interoperability of small ad hoc corpora and TMs



THANK YOU FOR YOUR ATTENTION

[laura.giacomini@iued.uni-heidelberg.de](mailto:laura.giacomini@iued.uni-heidelberg.de)