TermVar

Workshop on Terminological Variation

Needs of language experts and a proposal for a frame-based representation of terminological variants

21.07.2017

Dr. Laura Giacomini
Universität Hildesheim / Universität Heidelberg

Topics:

- Project overview
- Needs of language experts and coverage of variation in reference resources
- A frame-based representation of term variants

Project overview

"Morphologically related term variants. A multi-layered database design for the technical domain".

Goal: to design a lexicographic resource for specialised text production

Topics:

- A morphological approach to synonymous term variation:
  - Mainstream approaches to terminology variation
  - A definition of variation

- The technical domain and its language: the selected subdomains

- A typology of morphologically related variants:
  - Types of terms
  - Types of variants

- Data modelling:
  - Conceptual level: ontology
  - Semantic level: frames
  - Lexical level: terminology

Needs of language experts and coverage of variation in reference resources

Text production and variation:

- Unavailability of adequate lexicographic/ terminographic resources for **text production**.

- Consequence: the workflow and the quality of text production are negatively influenced by typical issues related to synonymous variants,

  e.g. their **existence** (does a variant 'exist'?)
  acceptability (is a variant acceptable in a context?), and
  use (how can a variant be used in that context?).

- **Consistency** in terminology = consistency in variation use

Technical language(s):

- Limited degree of standardisation in many subfields (e.g. in the construction industry, cf. Dainty et al. 2007: 23) and specific (con)textual features are the main factors which promote proliferation of synonymous variants.
- It is difficult to distinguish **systematic** and **non-systematic** usage (independently of motivation).

Which **notion of variation** is relevant for translation studies and technical writing?

Premise:

translated/ produced texts are usually **diasystemically homogeneous** (= no geographical, chronological, communicative, … variation in the same text)

Focus of the study on variants which match the following criteria:

a) they display a **morphological similarity** (similarity = shared lexical morphemes)

b) are totally or partially **synonymous**, and

c) build clusters which mostly belong to the **same diasystemic level**.

Reference resources (specialised dictionaries, glossaries, termbases, etc.)

display major gaps in variant representation and presentation, and provide their users with **insufficient data and metadata.**

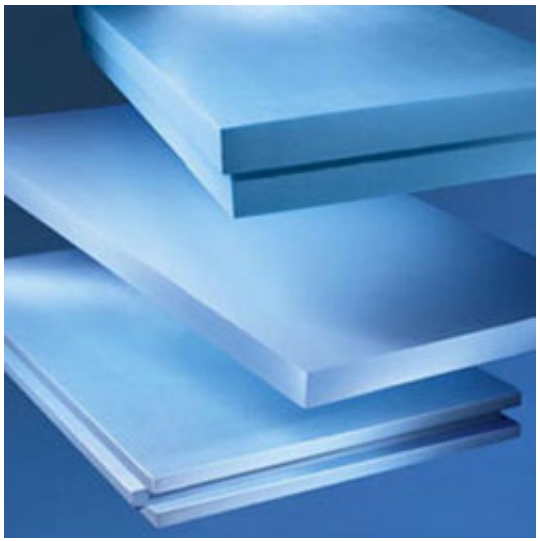
The main problems concern the following aspects:

- generally restricted number of covered variants

- multiword terms (n-grams with n>2) are rarely recorded

- coherency issues with the presentation of variants (macrostructure, microstructure, mediostructure)

Thermal insulation products:



PUR insulation board

PUR insulation foam

perlite loose fill insulation

mineral wool
insulation batt

mineral wool
insulation roll

Wood fibre natural insulation from Pavatex provides a high level of thermal resistance in roofs, walls (both internally and externally), floors and attics to reduce heat loss and to save on energy costs. (…)

For our added health and well-being, **wood fibre softboards** are a totally natural, sustainable and eco-friendly product made from new timber off-cuts from local sawmills. (…)

Pavatex receive this material as chippings, splinters and slabs and then further process it to create **wood fibre thermal insulation boards** that form part of the ecological lifecycle. (…)

There is no requirement to wear any protective clothing when installing **Pavatex wood fibre boards**. (…)

The **Pavatex boards** are manufactured using both the dry and wet processes, without the addition of any harmful toxins. (…)

This excess heat can be stored in the **wood fibre natural insulation boards** for as much as 10 to 12 hours due to their excellent thermal mass properties. (…)

http://www.acaraconcepts.com/wood-fibre-insulation

wood fibre (thermal) insulation/insulating board

is the usual denomination of a product category

Results obtained when searching bilingual terminographic/lexicographic resources for ***wood fibre thermal insulation board:***

1) *bau.net*: no results

2) *Baulexikon*, Beuth 2016: no results

3) *Fachwörterbuch Architektur und Bauwesen*, Beuth 2014: no results

4) *Technik Fachwörterbuch*, Langenscheidt 2015:

wood fiber insulating board <*build.mat*>, **wood-fibre insulating board**	Holzfaserdämmplatte
wood-fibre insulating board <*build.mat*>, **wood fiber insulating board**	Holzfaserdämmplatte

These resources usually record only

- single words (e.g. *board*) and
- binary word combinations (e.g. *insulating board, wood fibre)*

Linguee

Deutsch ↔ Englisch ä ö ü ß

wood fibre insulation board

Wörterbuch Englisch-Deutsch

wood fibre insulation *n* – Holzfaserdämmung *f*
wood fibre *n* – Holzfaser *f*
wood fibre[BE] *n* – Holzfaser *f*
wood fibre board *n* – Holzfaserplatte *f*

Externe Quellen (nicht geprüft)

⚠ PAVATEX developed the innovative PAVADENTRO wood fibre insulation board specially to meet this demanding challenge of building science. *pavatex.co.uk*	⚠ Speziell für diese bauphysikalisch sehr anspruchsvolle Aufgabe entwickelte PAVATEX die innovative Holzfaserdämmplatte PAVADENTRO. *pavatex.de*
⚠ The PAVADENTRO wood fibre insulation board in conjunction with a clay render and finish seemed to us to be a very good combination. *pavatex.co.uk*	⚠ Die PAVADENTRO-Holzfaserdämmplatte in Verbindung mit einem Lehmuntergrund und -feinputz erschien uns als sehr gute Kombination. *pavatex.de*
⚠ Dieffenbacher has developed a unique concept fort he production of stiff or flexible wood fibre insulation board. *dieffenbacher.de*	⚠ Dieffenbacher liefert hierfür ein einzigartiges Anlagenkonzept zur Herstellung von festen und flexiblen Holzfaser-Dämmplatten. *dieffenbacher.de*
⚠ The introduction of the flexible wood-fibre insulation board UdiFLEX and the system UdiSTEAM for complete air *unger-diffutherm.de*	⚠ Einführung der flexiblen Holzfaserdämmplatte UdiFLEX und das System UdiSTEAM für komplette Luft- und Winddichtigkeit *unger-diffutherm.de*
⚠ [...] market for high quality ecologically produced wood fibre insulation board with the new dry process specifically for application [...] *dieffenbacher.de*	⚠ [...] Trockenverfahrens zur Produktion von Qualitäts-Holzfaser-Dämmplatten ein enormes Wachstumspotential im Bereich der hochwertigen, [...] *dieffenbacher.de*
⚠ PAVATEX wood fibre insulation boards are natural wood products that people in the carpentry trade can work in a simple and [...] *pavatex.co.uk*	⚠ PAVATEX-Holzfaserdämmplatten sind natürliche Holzprodukte, die für uns Zimmerleute einfach und bequem zu verarbeiten sind", [...] *pavatex.at*
⚠ The wet process used makes it possible to integrate a mineral functional layer into the wood fibre insulation board. *pavatex.co.uk*	⚠ Das Nassverfahren erlaubt es, eine mineralische Funktionsschicht in die Holzfaserdämmplatte einzubringen. *pavatex.de*
⚠ In early 2007 Homatherm has taken the decision to invest in another production line for wood fibre insulation board. *dieffenbacher.de*	⚠ Anfang des Jahres 2007 fiel bei Homatherm die Entscheidung, in eine weitere Anlage zur Herstellung von Holzfaserdämmplatten zu investieren. *dieffenbacher.de*
⚠ This means that the wood fibre insulation board offers benefits in terms of costs and labour. *pavatex.co.uk*	⚠ Die Holzfaserdämmplatte hat damit Vorteile beim Kosten- und Arbeitsaufwand. *pavatex.de*

Are text producers expected to build longer multiword terms from single terms or binary word combinations in a compositional way?

How can lexicographic or terminographic resources satisfy the needs of translators and technical writers for what concerns term variation?

Requirements:

- Need for systematic coverage of **non-diasystemic variation**
- Need for **syntactic and semantic** information concerning variants
- Need for **pragmatic** information concerning text sources, genres, type of communication, etc.
- Need for a clear and coherent **link between domain terminology and domain ontology**

A frame-based representation of term variants

Frame: system of concepts related in such a way that to understand any one of them it is necessary to understand the whole structure in which it fits (Fillmore 2006)

frame (TRANSPORTATION)

frame_elements (MOVER(S), MEANS, PATH)

scene (MOVER(S) move along PATH by MEANS)

subframes: frame (DRIVING), frame (RIDING_1), …

Frame Semantics and specialised language:

- Frame-Based Terminology (cf. Faber 2015)

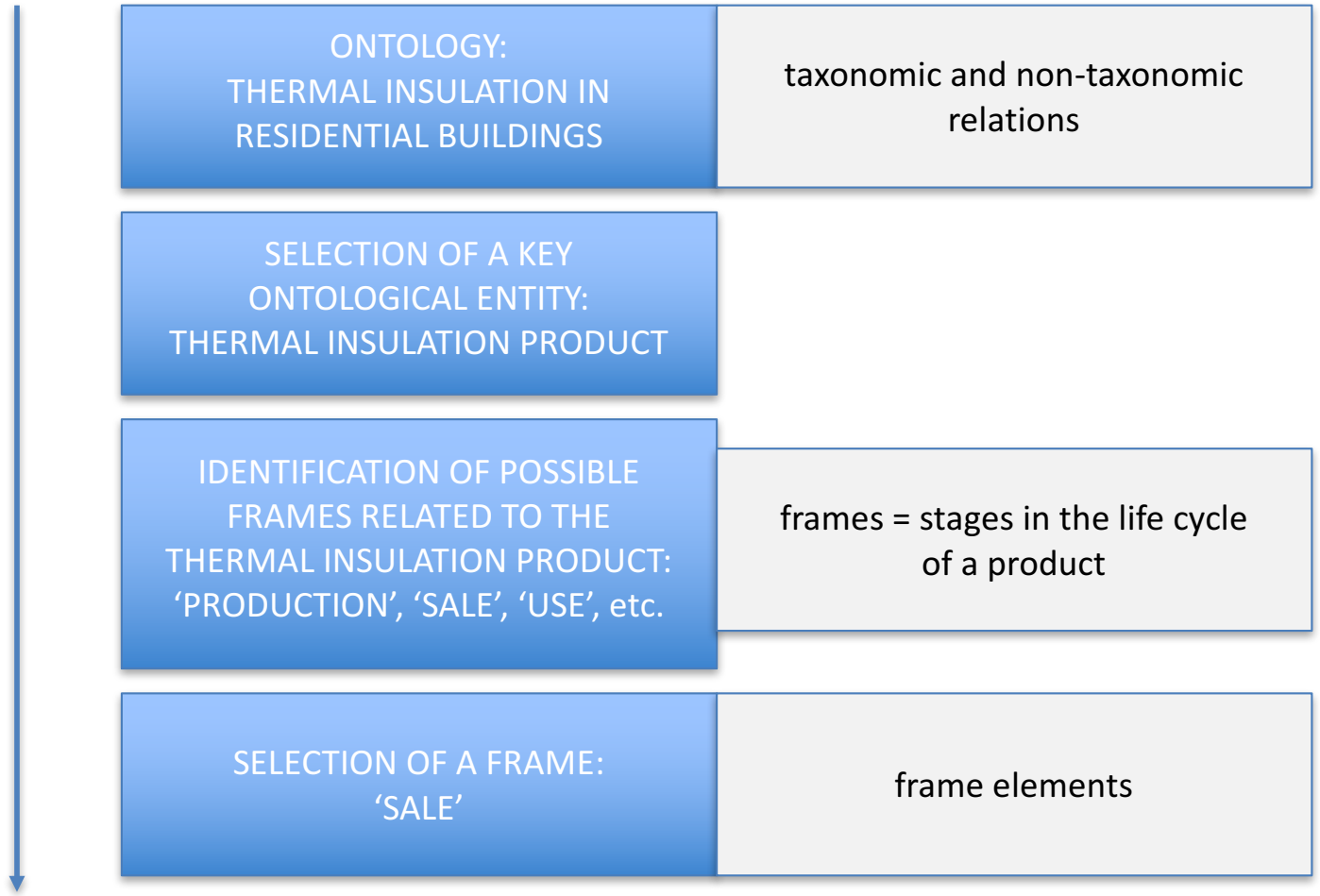
- medical language (Dolbey/ Ellsworth/ Scheffczyk 2006; Kokkinakis/ Gronostaj 2008; Wandji-Tchami 2017)

Frame-based lexicographic applications:

- Kicktionary (Schmidt 2008)

- EcoLexicon (Buendía Castro et al. 2014)

Frames: the interface between ontological (conceptual) and terminological layer



- In the selected frame, SALE, the insulation product is an object with distinctive features that is sold by producers or traders to specific users in order for them to thermally insulate one or more components/ areas of a building.

Frame (SALE)

Subframe: (FEATURES AND FUNCTIONALITY)

Frame Elements (MATERIAL, FORMAT, ...)

Scene:

An insulation product is made from a specific MATERIAL and has PRODUCT FEATURES.

It has a FORMAT and a NET COVERAGE, and is sold in a DELIVERY FORM in a PACKAGING (or a CONTAINER) in a specific QUANTITY.

The product is used by a USER with the GOAL of thermically insulating BUILDING ELEMENTS of a residential building made from specific BUILDING MATERIALS.

The product is used by means of an APPLICATION TECHNIQUE in the context of a specific CONSTRUCTION PROJECT.

- Each term or term component directly or indirectly referring to a thermal insulation product can be reduced to **frame elements** of the selected frame.

- Frame elements, e.g. MATERIAL, DELIVERY FORM, PART OF THE BUILDING, APPLICATION TECHNIQUE, can be understood as **potential semes** which often coincide with the previously identified ontological entities.

- The relation between the ontological, the semantic and the terminological level of the proposed model can be visualised as follows:

ontological level:	ontological macrocategory ontological category	FORM DELIVERY FORM	FUNCTION APPL. TECHNIQUE
semantic level:	frame element	DELIVERY FORM	APPL. TECHNIQUE
term level:	terms/ variants	*insulation boards*, *insulation batts*	*spray foam insulation*, *blow-in insulation*

Relation between Frame elements, terms and variation types:

● = total variation
◐ = partial variation
○ = no variation
GV = (ortho)graphical variation
MV = morphological variation
SV = syntactic variation

Frame elements	Terms	Variation	
		MV	SV
	[stone wool] [insulation] [batts] (pref. term)		
MATERIAL = [stone wool] DELIVERY FORM = [batt] GOAL = [insulation]	[mineral wool] [insulation] [batts]	◐	○
	[stone wool] [thermal insulation] [batts]	◐	●
	[stone wool] [batts] for [thermal insulation]	◐	●
	[stone wool] [batts] for [insulating] …	○	●
	[insulation] [batts] made of [stone wool]	○	●

Frame elements can be employed to semantically annotate terms/ variants:

e.g. Frame elements: MATERIAL (Ma), DELIVERY FORM (DF), GOAL (Go)

Wood fibre natural insulation from Pavatex	$N_{Ma} \bullet N_{Go}$
wood fibre softboards	$N_{Ma} \bullet N_{DF}$
wood fibre thermal insulation boards	$N_{Ma} \bullet N_{Go} \bullet N_{DF}$
Pavatex wood fibre boards	$N_{Ma} \bullet N_{DF}$
Pavatex boards	N_{DF}
wood fibre natural insulation boards	$N_{Ma} \bullet N_{Go} \bullet N_{DF}$

e.g. Frame elements: MATERIAL (Ma), DELIVERY FORM (DF), GOAL (Go)

Holzfaserdämmplatten	$N_{Ma} \bullet N_{Go} \bullet N_{DF}$
Dämmplatten aus Holzfasern	$N_{Go} \bullet N_{DF}\ p_{aus} N_{Ma}$
aus Holzfasern hergestellte Dämmplatten	$N_{Go} \bullet N_{DF}$ *hergestellt* $p_{aus} N_{Ma}$
Holzfaserplatten zur Dämmung (von X)	$N_{Ma} \bullet N_{DF}\ p_{zu} N_{Go}$
Dämmung mit Holzfaserplatten	$N_{Go}\ p_{mit} N_{Ma} \bullet N_{DF}$
Wärmedämmung mit Holzfaserplatten	$N_{Go}\ p_{mit} N_{Ma} \bullet N_{DF}$
dämmen mit Holzfaserplatten	$V_{Go}\ p_{mit} N_{Ma} \bullet N_{DF}$

Translation procedures > variation ‘procedures’:

- paraphrase	insulation materials > **materials for insulation**
- transposition [e.g. PoS change]	insulation materials > materials for **insulating**
- explication/ implication [e.g. hyperonymy-hyponymy]	insulation > **thermal** insulation
- expansion/ reduction [e.g. adding a new frame element]	polyurethane insulation > polyurethane **roof** insulation

Model for data representation in a termbase (excerpt):

Syntactic-semantic structure	Variation	Procedure	Term	Frame elements	Ontology
N_{Ma}•N_{Go}•N_{DF}	main term		stone wool Insulation batts	INSULATION PRODUCT **GOAL: INSULATION** **MATERIAL: STONE WOOL** **DELIVERY FORM: BATT** FORMAT PACKAGING BUILDING ELEMENT BUILDING MATERIAL APPL. TECHNIQUE CONSTRUCTION PROJ. PRODUCT FEATURE …	INSULATION PRODUCT - **MATERIAL** DELIVERY FORM FORMAT PACKAGING BUILDING ELEMENT BUILDING MATERIAL APPL. TECHNIQUE CONSTRUCTION PROJ. PRODUCT FEATURE …
N_{Ma}•N_{Go}•N_{DF}	◐ MV ● SV	explic.	stone wool thermal insulation batts		
N_{Ma}•N_{DF} $p_{for}$$N_{Go}$	◐ MV ● SV	paraph.	stone wool batts for thermal insulation		
N_{Go}• N_{DF} $p_{made\_of}$$N_{Ma}$	○ MV ● SV	paraph.	insulation batts made of stone wool		
N_{Ma}•N_{DF} $p_{for}$$V_{Go}$	○ MV ● SV	paraph. transp.	stone wool batts for insulating…		

+ examples, sources (and source types), communication type

Conclusions and topics for discussion:

The selected frame serves as an **interface** between the ontology and the lexicon of the subdomain, and provides a relevant tool for **semantic categorisation** of terms as well as for lexicographic disambiguation of variants.

Benefits of a frame-based data modelling:

- syntactic structures of terms (and rule-based variation description) can be provided with frame-based semantic annotation
- semantic annotation serves the purpose of detecting clusters of synonymous variants, which often display morphological similarity
- semantic annotation can be combined with variant annotation (GV, MV, SV) and with procedural annotation (paraphrase, transposition, etc.)

Variation can be understood as a relation between a preferred term and its (morphologically-related) variants. Which criteria can be adopted to identify preferred terms (e.g. standardisation, frequency, pragmatic criteria)?

How far is this frame-based model applicable to other specialised domains (cf. conceptual features, level of standardisation, etc.)?

Which are the requirements data presentation in a lexicographic resource needs to comply with in order to support text production?

Is frame-based annotation suitable for (semi)-automatic applications?

References

- Buendia Castro, M., Montero Martínez, S., & Faber, P. (2014). Verb collocations and phraseology in EcoLexicon. In: *Yearbook of Phraseology*, *5*(1), 57-94.
- Dainty, A., Moore, D., & Murray, M. (2007). *Communication in construction: Theory and practice*. Routledge.
- Dolbey, A., Ellsworth, M., & Scheffczyk, J. (2006). In: *KR-MED* (Vol. 222).
- Faber, P. (2015). Frames as a framework for terminology. In: Kockaert H.J. & Steurs F. (Eds.), *Handbook of terminology* (Vol. 1). John Benjamins, 14-33.
- Fillmore, C. J. (2008). Frame semantics. In: Geeraerts D. (Ed.), *Cognitive linguistics: Basic readings*. Walter de Gruyter, 373-400.
- Hoffmann, L. (1985). *Kommunikationsmittel Fachsprache. Eine Einführung*. Gunter Narr.
- Kokkinakis, D., & Toporowska Gronostaj, M. (2008). MedLex+: An Integrated Corpus-Lexicon Medical Workbench for Swedish. In: *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008).*
- Schmidt, T. (2008). The Kicktionary: Combining corpus linguistics and lexical semantics for a multilingual football dictionary. In: Lavric E. et al. (Eds.), *The linguistics of football*. Gunter Narr, 11-22.
- Wandji-Tchami (2017). A frame-based study of semantic variation of verbs in medical corpora. *Presentation at TermVar2017 – Workshop on Terminological Variation*, University of Hildesheim, 20.-21-07.2017.