

# Translation of taxonomies: challenges, methods and synergies



**suva**

Mehr als eine Versicherung



Silvia Giacomotti, Dr. Annette Weilandt, Beate Früh  
London October 17, 2017

# Who we are

Silvia Giacomotti  
Suva  
Head of Language Management.  
Lucerne, Switzerland



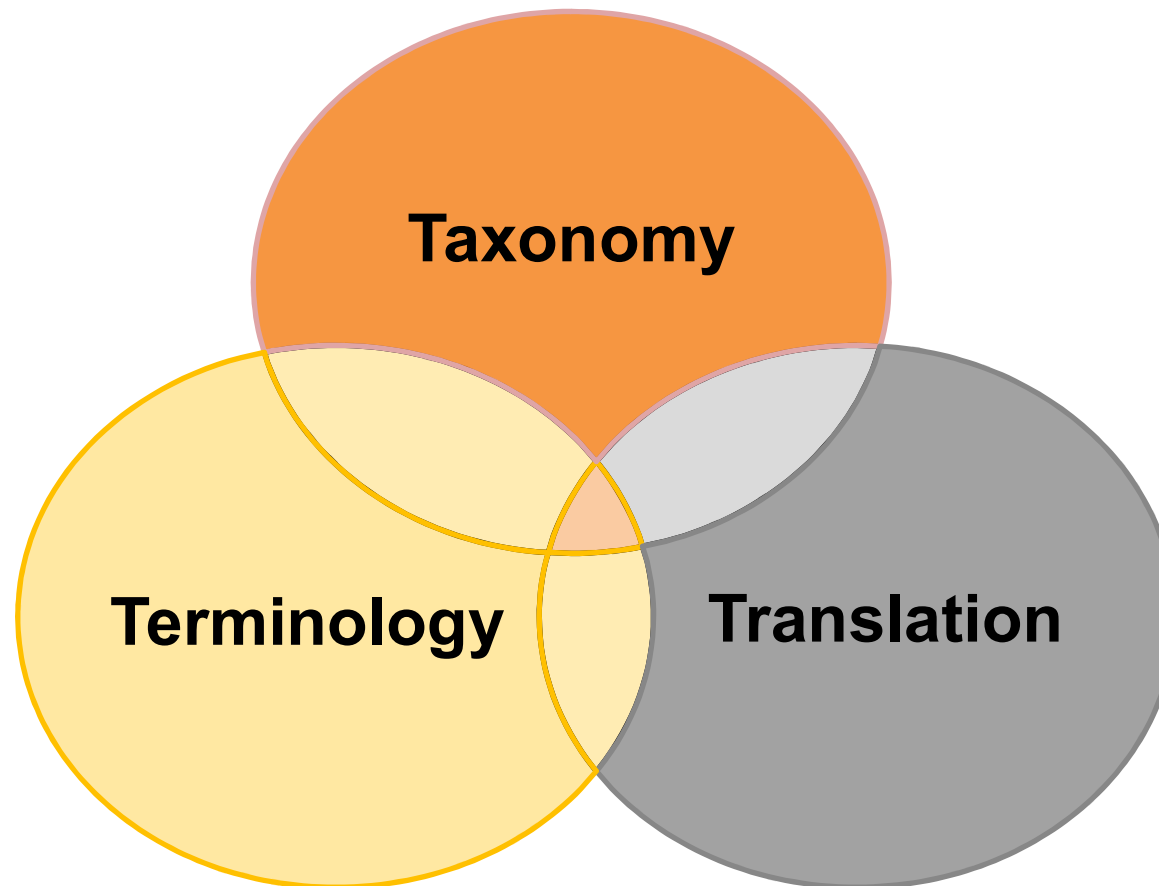
Dr. Annette Weilandt  
eccenca GmbH  
Senior Consultant  
Leipzig, Germany



Beate Fröh  
Büro b3 Terminologiemangement  
Senior Consultant and Terminologist,  
Nürnberg, Germany



# Synergies



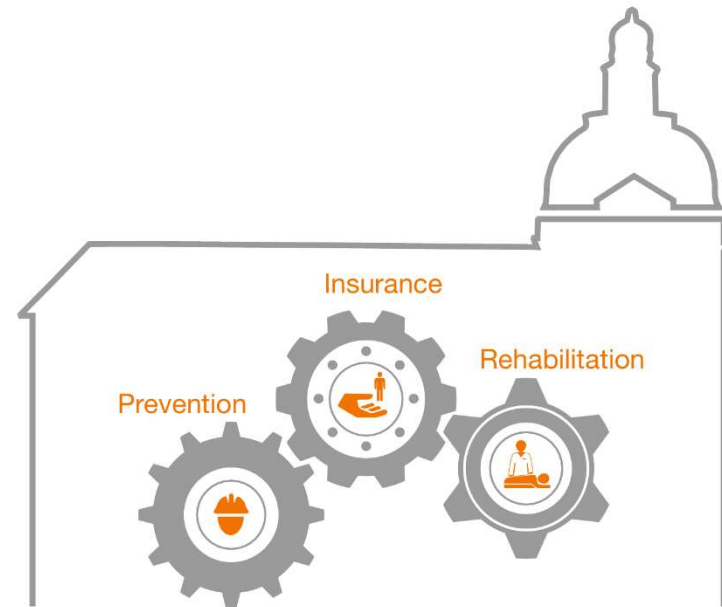
# Agenda

- Suva - More than just insurance
- Content and documentation at Suva
- Knowledge Management at Suva
- Suva Taxonomy
- Terminology Management
- Computer-assisted translation
- Translation of Sharepoint taxonomies
- Summary and conclusion

# More than just insurance

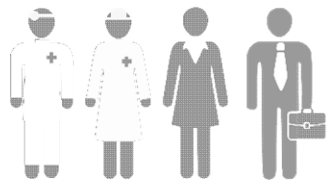
## Suva and its unique business model

- Suva, the Swiss National Accident Insurance Fund, is an independent company under public law
- It mainly insures industries with high accident risk as well as unemployed persons in Switzerland
- In addition, it manages the Military Insurance on behalf of the Federal Government
- Suva offers a comprehensive product mix of prevention, insurance and claims management



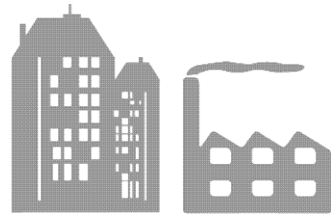
# Suva at a glance

## Key figures for 2016



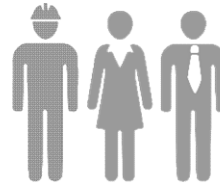
**4,200**

employees incl. clinics



**127,900**

companies insured



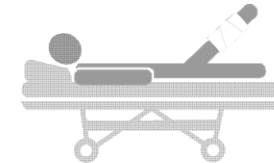
**2 m**

people insured



**CHF 147,8 bn**

insured payroll



**461,000**

accidents and  
occupational diseases



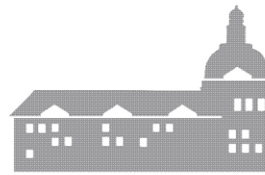
**CHF 4,1 bn**

premium income



**CHF 4,2 bn**

insurance benefits



**CHF 700 m**

operating costs



**CHF 48 bn**

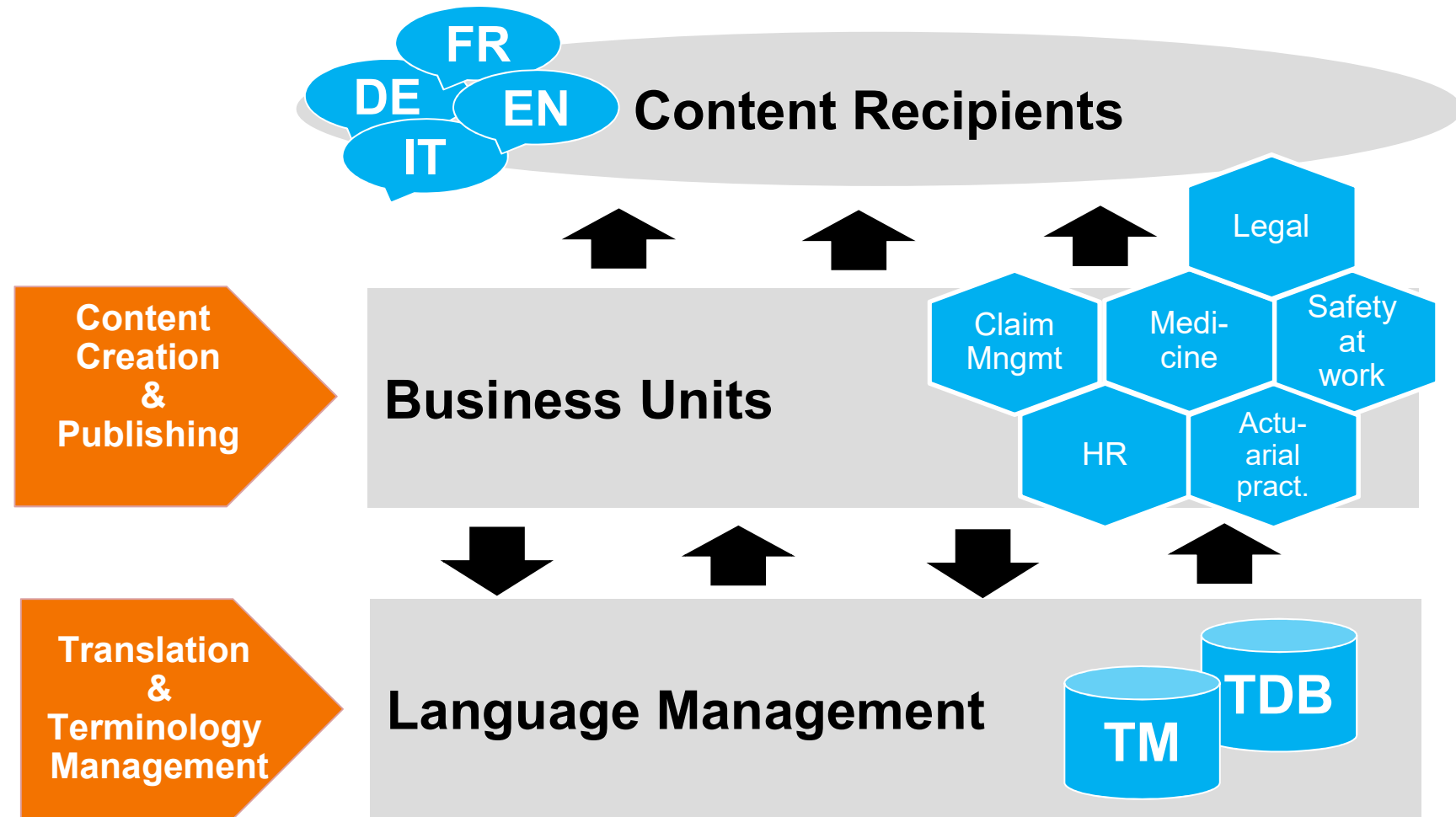
fixed assets (market value)



**CHF 32,2 m**

operating result

# Content and documentation at Suva: n fields of knowledge, 4 languages



# Knowledge Management at Suva: the challenge of complexity

## Challenges

**Highly multidisciplinary organisation**

**Multilingualism**

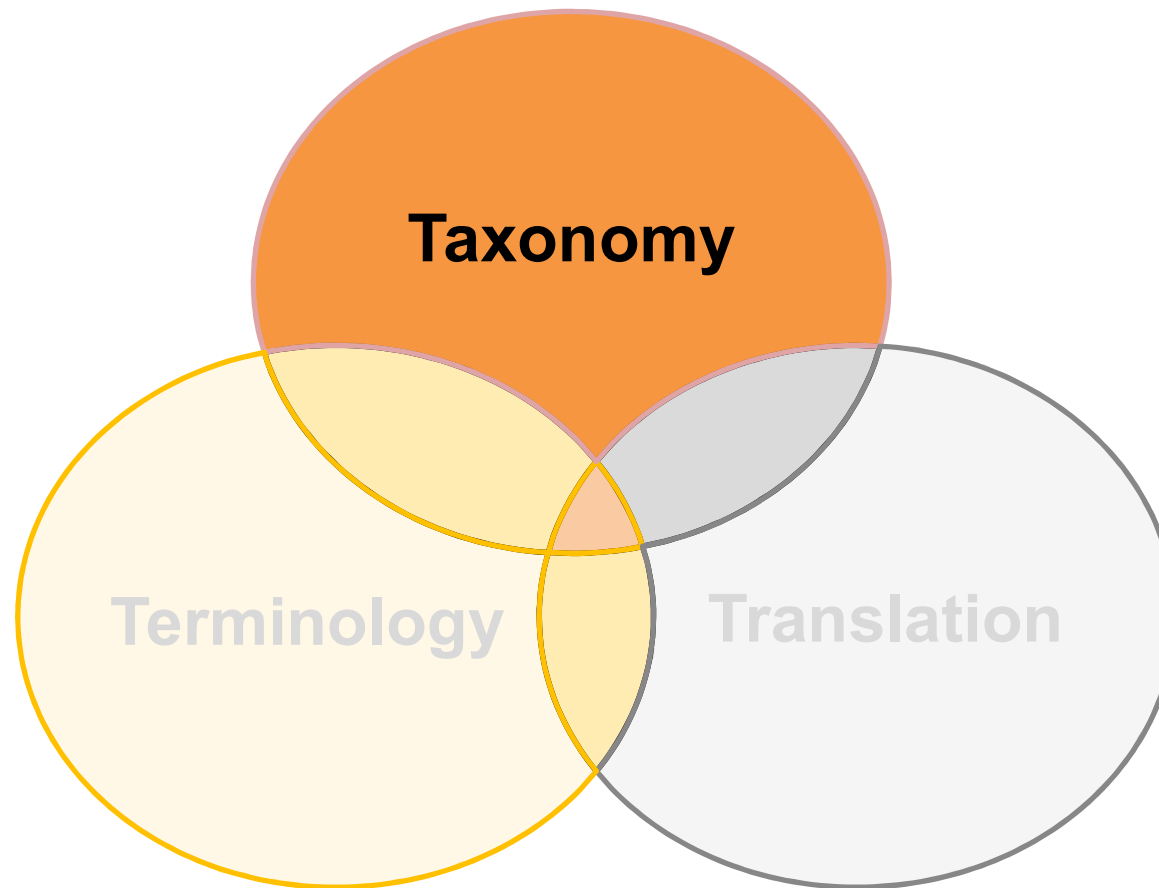
**Decentralized responsibility for knowledge management**

**Different file stores and knowledge management tools**

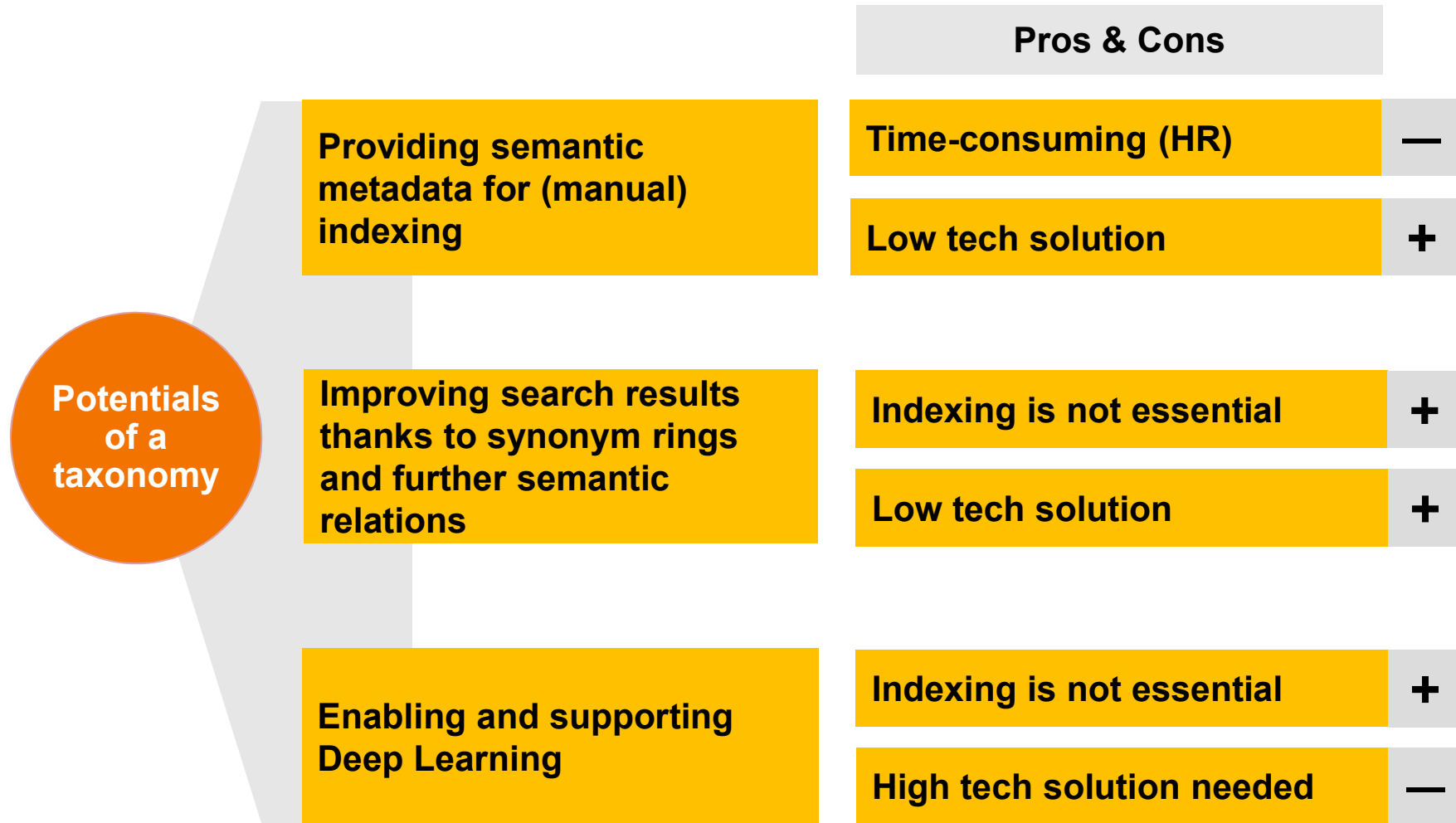
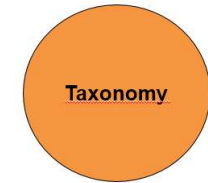
**No systematic indexing of content**



# Taxonomy



# Improving information retrieval: what a taxonomy can do

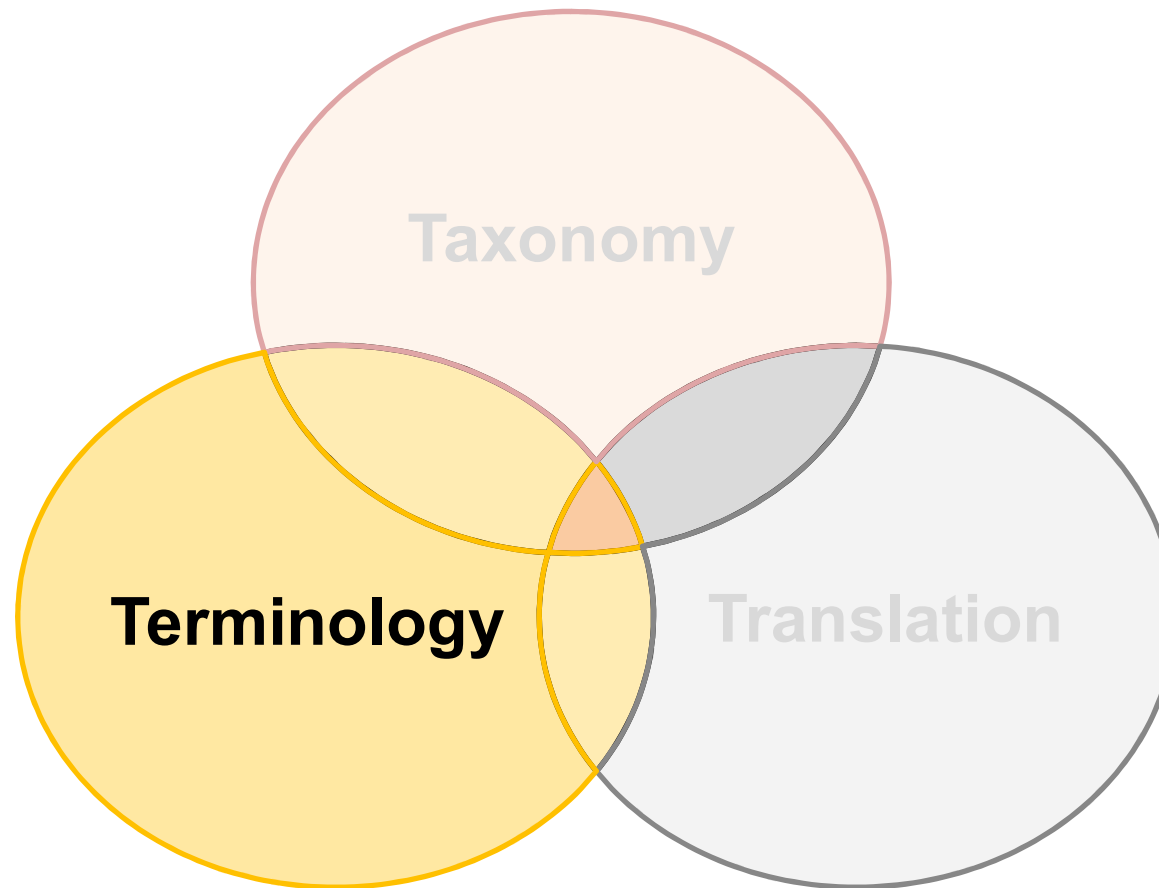


# xls Taxonomy

Taxonomy

<b>Casualty insurance</b>			
Casualty insurance	<b>Insurance products</b>		
Casualty insurance	Insurance products	<b>Occupational accident insurances (OAI)</b>	
Casualty insurance	Insurance products	<b>Non-occupational accident insurances (NOAI)</b>	
Casualty insurance	Insurance products	<b>Military insurances</b>	
Casualty insurance	Insurance products	Military insurances	<b>Voluntary basic insurance for pensioners</b>
Casualty insurance	Insurance products	<b>Accident insurances for the unemployed</b>	
Casualty insurance	Insurance products	<b>Insurances for entrepreneurs</b>	
Casualty insurance	Insurance products	<b>Interim accident insurances</b>	
Casualty insurance	<b>Insured events</b>		
Casualty insurance	Insured events	<b>Accidents</b>	
Casualty insurance	Insured events	Accidents	<b>Occupational accidents</b>
Casualty insurance	Insured events	Accidents	<b>Non-occupational accidents</b>
Casualty insurance	Insured events	<b>Occupational diseases</b>	
Casualty insurance	Insured events	Occupational disease	<b>Unfitness</b>
Casualty insurance	<b>Accident consequences</b>		
Casualty insurance	Accident consequences	<b>Injuries</b>	
Casualty insurance	Accident consequences	<b>Adverse health effects</b>	
Casualty insurance	Accident consequences	<b>Disability</b>	
Casualty insurance	Accident consequences	<b>Helplessness</b>	
Casualty insurance	Accident consequences	<b>Loss of integrity</b>	
Casualty insurance	<b>Insurance benefits</b>		
Casualty insurance	Insurance benefits	<b>Care benefits</b>	
Casualty insurance	Insurance benefits	<b>Reimbursements of expenses</b>	
Casualty insurance	Insurance benefits	<b>Rehabilitation</b>	
Casualty insurance	Insurance benefits	<b>Disability pension</b>	
Casualty insurance	Insurance benefits	<b>Survivors' pension</b>	
Casualty insurance	Insurance benefits	Survivors' pension	<b>Widows' pension</b>
Casualty insurance	Insurance benefits	Survivors' pension	<b>Children pension</b>
Casualty insurance	Insurance benefits	<b>Compensation for the loss of bodily functions</b>	
Casualty insurance	Insurance benefits	<b>Helplessness allowance</b>	
Casualty insurance	<b>Damage segments</b>		
Casualty insurance	Damage segments	<b>Normal segments</b>	
Casualty insurance	Damage segments	Normal segments	<b>Normal case</b>
Casualty insurance	Damage segments	<b>Complex segments</b>	

# Terminology



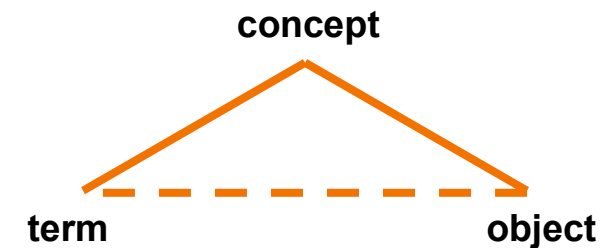
# Terminology



Specific concepts and terms of a domain or organization.

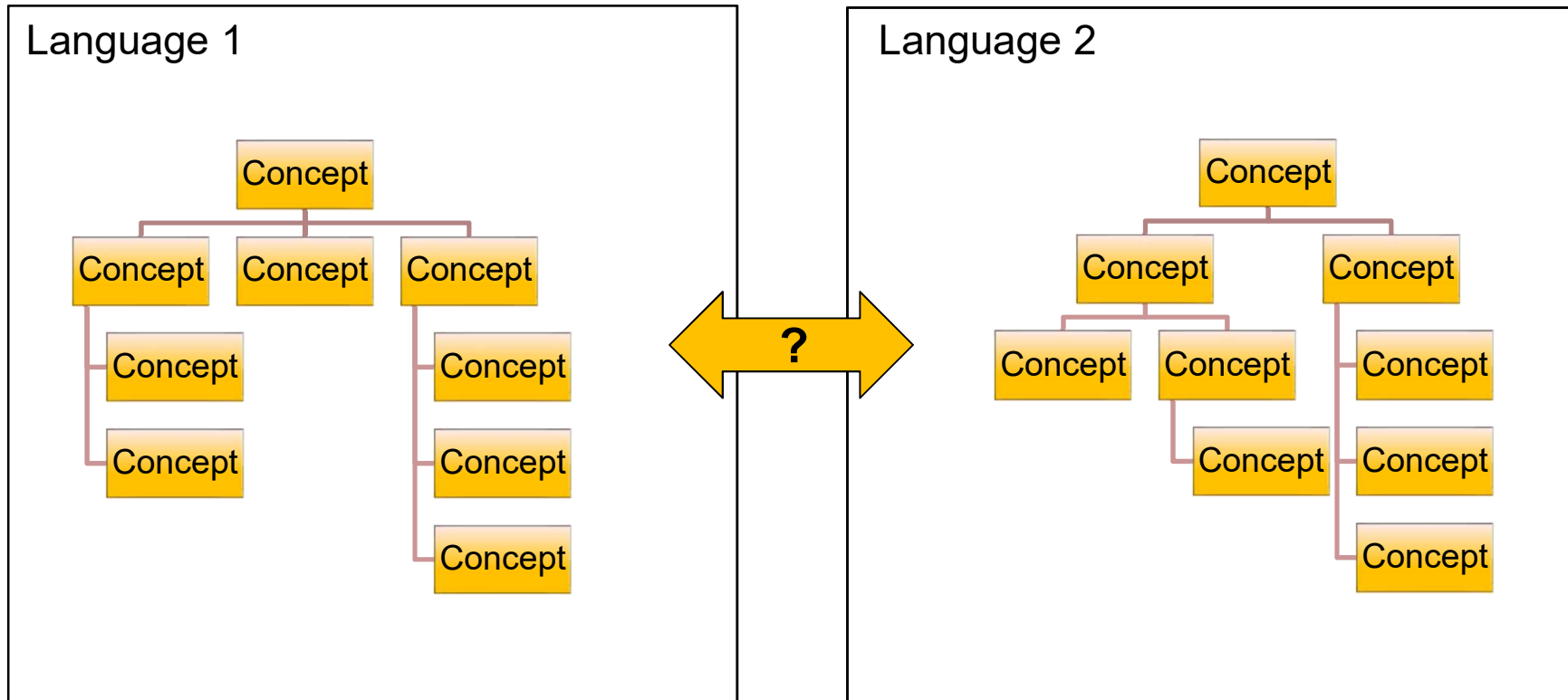
## Terminology management

- Unambiguous definition of concepts
- Harmonization of terms
- Consistent usage
- Systematic approach



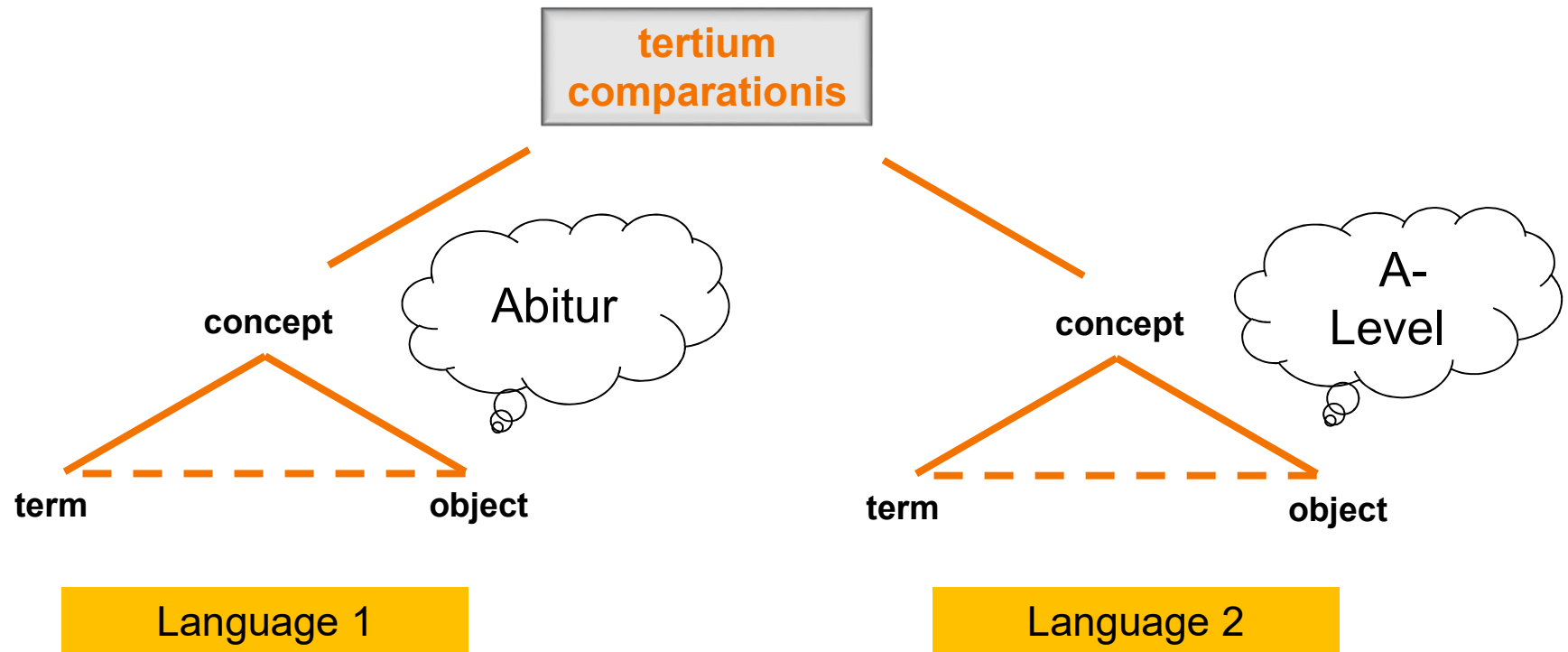
# Systematic approach

Terminology





# Terminology in different languages

Terminology



# Terminology database

Terminology

Definition (DE):	Versicherungsprodukt, das mit allen geeigneten Mitteln Arbeitsunfälle und Berufskrankheiten verhüten soll, und nach Eintritt von Arbeitsunfällen oder Berufskrankheiten die Gesundheit der Versicherten mit allen geeigneten Mitteln wieder herstellen soll, und sie oder ihre Hinterbliebenen durch Geldleistungen entschädigt.
Subject matter:	Insurance plans
Definition (EN):	A social security plan that primarily covers the losses and injuries to persons caused by professional accidents and occupational diseases.
 <b>German</b>	
<b>Berufsunfallversicherung</b>	
Usage:	allowed
Geographical usage:	CH
PoS:	noun
Gender:	f
Numerus:	Sg
Term type:	full form
<b>BU</b>	
Usage:	allowed
Geographical usage:	CH
Term type:	acronym
<b>gesetzliche Unfallversicherung</b>	
Usage:	allowed
Geographical usage:	DE
PoS:	noun
Gender:	f
Numerus:	Sg
 <b>English</b>	
<b>occupational accident insurance</b>	
Usage:	allowed
Geographical usage:	CH
PoS:	noun
Numerus:	Sg
Term type:	full form
<b>OAI</b>	
Usage:	allowed
Term type:	acronym
Geographical usage:	CH
<b>work accident insurance</b>	
Usage:	allowed
Geographical usage:	GB
PoS:	noun
<b>insurance for accidents at work</b>	
Usage:	rejected



# Terminology database

Terminology

Definition (DE): Versicherungsprodukt, das mit allen geeigneten Mitteln Arbeitsunfälle und Berufskrankheiten verhüten soll, und nach Eintritt von Arbeitsunfällen oder Berufskrankheiten die Gesundheit der Versicherten mit allen geeigneten Mitteln wieder herstellen soll, und sie oder ihre Hinterbliebenen durch Geldleistungen entschädigt.

Subject matter: Insurance plans

Definition (EN): A social security plan that primarily covers the losses and injuries to persons caused by professional accidents and occupational diseases.

## German

### Berufsunfallversicherung

Usage: allowed  
Geographical usage: CH  
PoS: noun  
Gender: f  
Numerus: Sg  
Term type: full form

## BU

Usage: allowed  
Geographical usage: CH  
Term type: acronym

### gesetzliche Unfallversicherung

Usage: allowed  
Geographical usage: DE  
PoS: noun  
Gender: f  
Numerus: Sg

## English

### occupational accident insurance

Usage: allowed  
Geographical usage: CH  
PoS: noun  
Numerus: Sg  
Term type: full form

## OAI

Usage: allowed  
Term type: acronym  
Geographical usage: CH

### work accident insurance

Usage: allowed  
Geographical usage: GB  
PoS: noun

### insurance for accidents at work

Usage: rejected

## Italian

### assicurazione contro gli infortuni professionali

Usage: allowed  
Geographical usage: CH  
PoS: noun  
Gender: f  
Numerus: Sg

### assicurazione contro gli infortuni sul lavoro

Usage: allowed  
Geographical usage: IT  
PoS: noun  
Gender: f  
Numerus: Sg

# Terminology database

Terminology

- Concept-oriented structure
- Definitions
- Relations between concepts
- Synonyms
- Usage instructions
- Equivalents in different languages

**Gefahr**

Entry number: 7  
Definition (EN): The possibility that someone may be harmed or killed.  
Definition (DE): Die Möglichkeit dass jemand verletzt oder getötet werden kann.

**German**  
**Gefahr**  
Usage: allowed  
Geographical usage: CH, DE  
PoS: noun  
Gender: f  
Numerus: Sg

**English**  
**danger**  
Usage: allowed  
Geographical usage: US, GB  
PoS: noun  
Numerus: Sg

**Italian**  
**pericolo**  
Usage: allowed

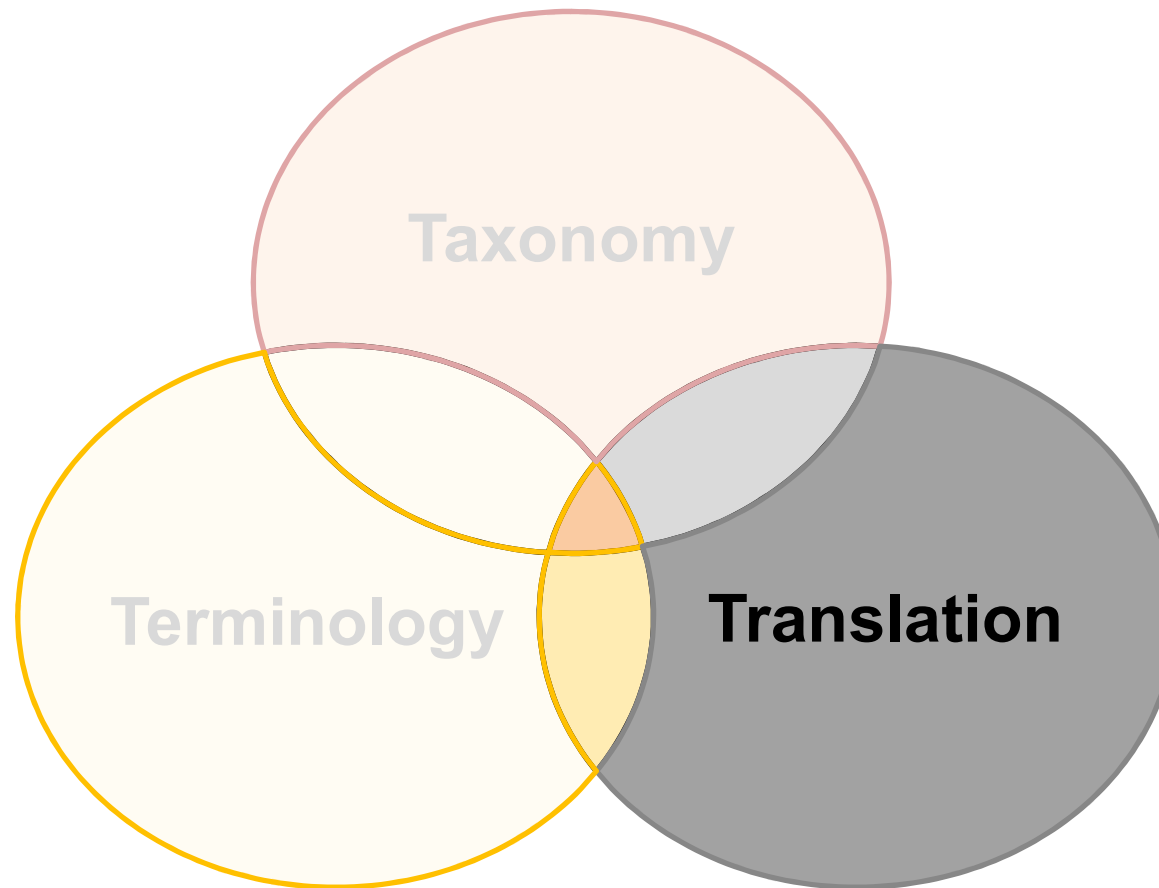
Definition (EN): Something which could be dangerous to a person, her health or safety.  
Definition (DE): Etwas, das einer Person, ihrer Gesundheit oder Sicherheit Schaden bringen kann.

**German**  
**Gefahr**  
Usage: allowed  
Geographical usage: CH, DE  
PoS: noun  
Gender: f  
Numerus: Sg

**English**  
**hazard**  
Usage: allowed  
Geographical usage: US, GB  
PoS: noun  
Numerus: Sg

**Italian**  
**pericolo**  
Usage: allowed  
Geographical usage: CH  
PoS: noun  
Gender: m  
Numerus: Sg

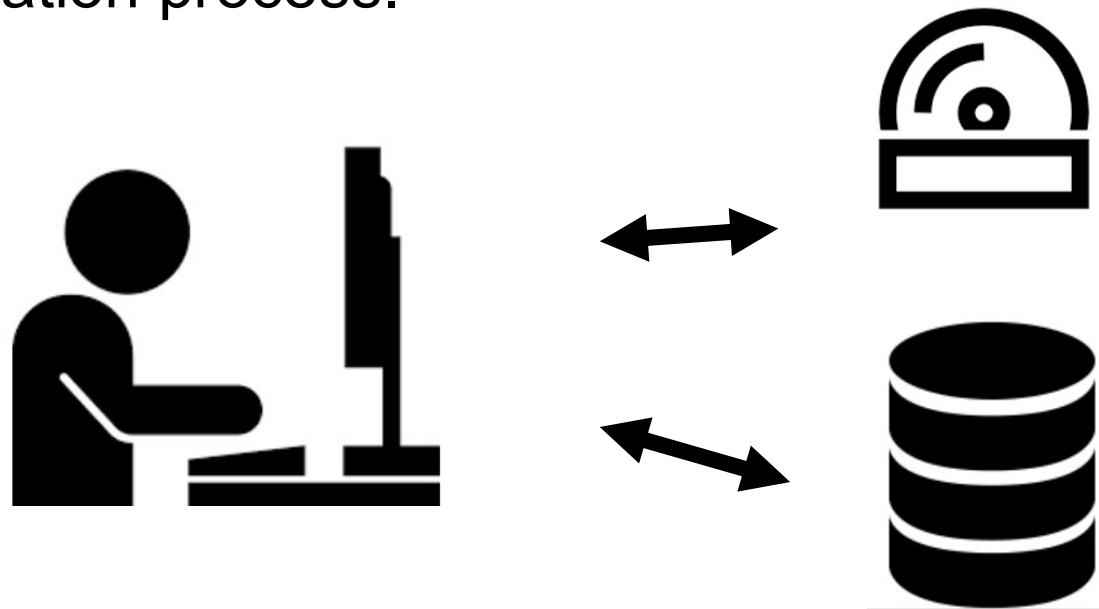
# Translation



# CAT – computer-assisted translation

## What is CAT?

CAT is a form of translation where a human translator uses computer hardware and software to support and facilitate the translation process.

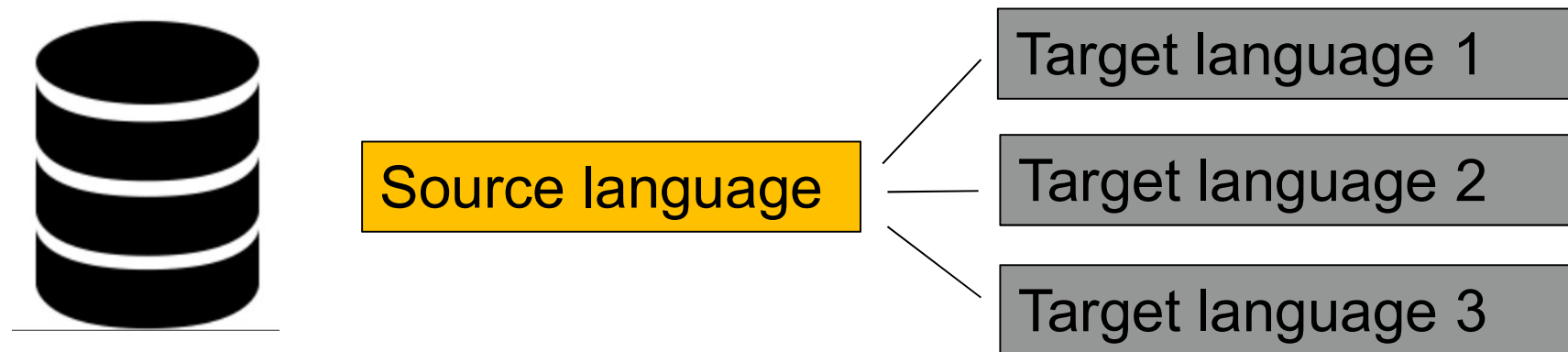


Not to be confused with machine translation.

# CAT – computer-assisted translation

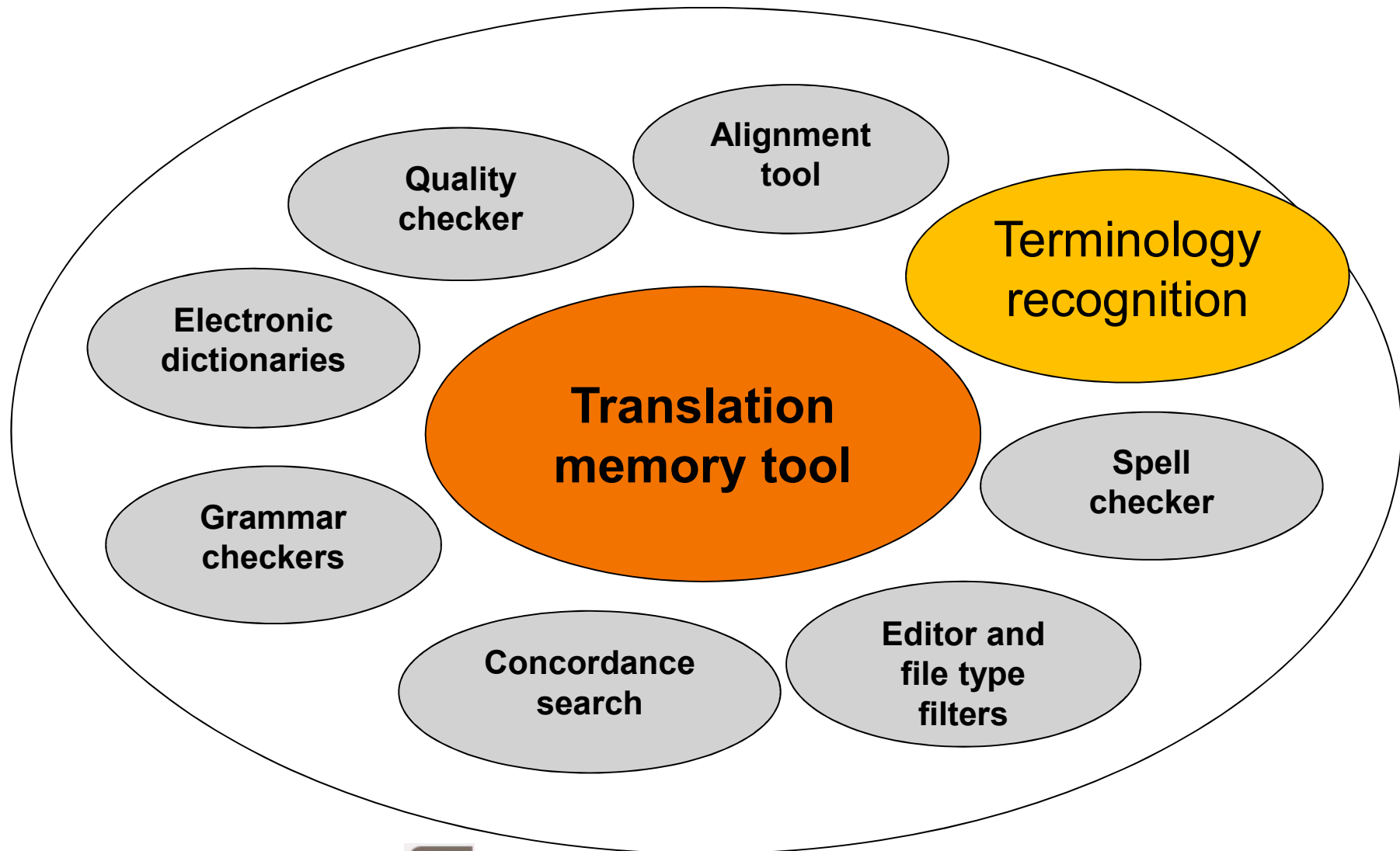
## What is CAT?

**Translation memory tools** consist of a database of previously translated text segments in a source language and their translations in one or more target languages.



# CAT – computer-assisted translation

Translation



# CAT – computer-assisted translation

## Advantages & benefits

- Reuse of translation segments
  - Reduction of translation costs due to reduced prices for fuzzy matches and 100 % matches
- Less time needed for translation jobs
- Supported by company-specific terminology database or other resources like dictionaries



# CAT – computer-assisted translation Transit

Translation

The screenshot displays the StarTransit CAT software interface. It features three main panels:

- Source Language View (Top):** A table with columns for line numbers and text. The text is in German (Deutsch (Schweiz)).

Line	Text
9	MV
11	Berufskrankheit
13	Berufsunfall
- Target Language Editor (Middle):** A table with columns for line numbers and text. The text is in English (UK).

Line	Text
9	MI
11	occupational disease
13	Berufsunfall
15	Arbeitsunfall
17	Retriebsunfall
- Terminologie (Right):** A panel showing a terminology database entry for "Berufsunfall" (occupational accident). It includes details like "Sachgebiet: Medizin; Baugewerbe; Industrie und Gewerbe", "Status: 3", "Projekt: Gerüste; Suva-Wegleitung", "Benutzer3: HSR, LUH", "Online: Ja", and "Wörterbuch: SuvaTerm (StarTransitNXT)".
- Fuzzy Ausgang (Bottom):** A panel showing a fuzzy match output. It includes a reference sentence: "we no longer focus primarily on occupational accidents, but also on leisure-time accidents." and a fuzzy match score of 70%.

source language view

target language editor

synchronisation  
with  
terminology  
database

Reference material for  
pretranslates and context  
sentences



# CAT – computer-assisted translation

## SDL Trados Studio

Translation

Fuzzy match from a previous translation

The screenshot displays the SDL Trados Studio interface. The top-left pane shows project settings for 'Betriebsunfallversicherung'. The main workspace is divided into three panes: the left pane shows the source language view with a list of terms and their corresponding target language translations; the middle pane shows the target language editor with a list of terms and their corresponding target language translations; the right pane shows the terminology database with a list of terms and their corresponding target language translations.

source language view

target language editor

synchronisation with terminology database

# Sharepoint Term Store Management

Translation

- Manual assignment of source language equivalents
- or
- XML translation file
- One-to-one mapping of source taxonomy in target language



# Translation of Sharepoint taxonomies

Translation

Export of a  
translation package  
of term set

The screenshot shows the 'Term Store Management Tool' interface. The 'TRANSLATION' tab is selected. The term set 'Unfallversicherung' is displayed. Three options are available:

- ☐ **Machine Translate**  
Send the contents of this term set to be machine translated.
- ☐ **Create Translation Package**  
Export the contents of this term set into XLIFF file for professional translation.
- ☐ **Upload Translation**  
Upload translated XLIFF files from a professional translator.

A 'Continue' button is located at the bottom right. The 'Create Translation Package' option is highlighted with an orange rectangle.

# Translation of Sharepoint taxonomies

Translation

Whole term set  
or only modified and  
new terms

For each language  
one file (source and  
target language)

The screenshot displays the 'Term Store Management Tool' interface. The 'TRANSLATION' tab is selected, showing options for 'Unfallversicherung'. Two radio buttons are present: 'Machine Translate' (disabled) and 'Create Translation Package' (selected and highlighted with an orange box). Below this, the 'Create Translation Package' dialog is open. It includes a section 'Select Terms to Translate' with instructions. Two radio buttons are available: 'Translate all terms' (selected) and 'Translate updated terms'. Below these are dropdown menus for 'Translate from' (set to 'German (Germany)') and 'Translate to' (set to 'Italian (Italy)'). At the bottom right of the dialog are 'Continue' and 'Cancel' buttons.

# Sharepoint translation file

Translation

## XLIFF

(an OASIS standard)

```
<?xml version="1.0" encoding="utf-8"?>
<xliff version="1.2" xmlns="urn:oasis:names:tc:xliff:document:1.2">
```

## XML Localization Interchange File Format

## Export XLIFF

```
<file original="9473554c-9e5f-413d-a8f8-5a5155e3bed" source-language="de-DE" target-language="en-US" datatype="text">
  <header>
    <note>type=TermSetItem</note>
    <note>translatorType=Vendor</note>
    <tool tool-id="SharePoint Server 2016" tool-name="Microsoft® SharePoint® Server 2016" />
    <note>termStoreId=15e113b9-6047-4f36-8736-296e5a7cc3d0</note>
    <note>termSetId=71fa6a73-852c-4e31-861d-261e6250eb7e</note>
    <note>waterline=636432549112182109</note>
    <note>path=Versicherungsprodukte;Berufsunfallversicherung</note>
  </header>
  <body>
    <trans-unit id="TermLabel0" datatype="plaintext">
      <source>Berufsunfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermLabel1" datatype="plaintext">
      <source>Betriebsunfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermLabel2" datatype="plaintext">
      <source>BU</source>
    </trans-unit>
    <trans-unit id="TermLabel3" datatype="plaintext">
      <source>gesetzliche Unfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermDescription" datatype="plaintext">
      <source>Versicherungsprodukt, das mit allen geeigneten Mitteln Arbeitsunfälle und Berufskrankheiten ver<
    </trans-unit>
  </body>
</file>
<file original="251fdedc-4c11-4465-af28-5a7865b9e2b5" source-language="de-DE" target-language="en-US" datatype="text">
  <header>
```



# Sharepoint translation file

Translation

```
<file original="9473554c-9e5f-413d-a8f8-5a5155e3bed3" source-language="de-DE" target-language="en-US" datatype="text">
  <header>
    <note>type=TermSetItem</note>
    <note>translatorType=Vendor</note>
    <tool tool-id="SharePoint Server 2016" tool-name="Microsoft® SharePoint® Server 2016" />
    <note>termStoreId=15e113b9-6047-4f36-8736-296e5a7cc3d0</note>
    <note>termSetId=71fa6a73-852c-4e31-861d-261e6250eb7e</note>
    <note>waterline=636432549112182109</note>
    <note>path=Versicherungsprodukte;Berufsunfallversicherung</note>
  </header>
  <body>
    <trans-unit id="TermLabel0" datatype="plaintext">
      <source>Berufsunfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermLabel1" datatype="plaintext">
      <source>Betriebsunfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermLabel2" datatype="plaintext">
      <source>BU</source>
    </trans-unit>
    <trans-unit id="TermLabel3" datatype="plaintext">
      <source>gesetzliche Unfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermDescription" datatype="plaintext">
      <source>Versicherungsprodukt, das mit allen geeigneten Mitteln Arbeitsunfälle und Berufskrankheiten ver</source>
    </trans-unit>
  </body>
</file>
```

**Description**  
Descriptions will help users know when to use this term, and to disambiguate amongst similar terms.

Versicherungsprodukt, das mit allen geeigneten Mitteln Arbeitsunfälle und Berufskrankheiten verhüten soll, und nach Eintritt von Arbeitsunfällen oder

**Default Label**  
Enter one label as the default for this language.

Berufsunfallversicherung

**Other Labels**  
Enter synonyms and abbreviations for this term. (You can enter a word or phrase per line.)

Betriebsunfallversicherung  
BU  
gesetzliche Unfallversicherung

# CAT – computer aided translation

## SDL Trados Studio

Translation

The screenshot displays the SDL Trados Studio interface. At the top, the 'Projekteinstellungen' (Project Settings) window is open, showing the project name 'Betriebsunfallversicherung' and a 90% match with 'Professional accident'. Below this, the 'Übersetzungsergebnisse' (Translation Results) window shows a list of translation units. The table below highlights several units with orange boxes:

Unit ID	Source Text	Match %	Target Text	Quality
16	Berufsunfallversicherung	100%	Occupational accident insurance	TU+
17	Betriebsunfallversicherung	90%		TU+
18	BU	100%	OAI	TU+
19	gesetzliche Unfallversicherung			TU+
20	Versicherungsprodukt, das mit allen geeigneten Mitteln Arbeitsunfälle und Berufskrankheiten verhüten soll, und nach Eintritt von Arbeitsunfällen oder Berufskrankheiten die Gesundheit der Versicherten mit allen geeigneten Mitteln wieder herstellen soll, und sie oder ihre Hinterbliebenen durch Geldleistungen entschädigt.			TU+

On the right side, a 'Terminologieerkennung' (Terminology Recognition) window is open, showing a list of terms and their status:

- Betriebsunfallversicherung rejected, CH|DE
- occupational accident insurance allowed, CH
- OAI allowed, CH
- work accident insurance allowed, GB
- insurance for accidents at work rejected

**Problem:** Alternate labels are included in the translation, but often there is not really an equivalent in the target language. QA checker in CAT tool demands for a translation. Untranslated alternate labels are filled by Sharepoint automatically with source language term.

# Sharepoint translation file

Translation

## Translated import XLIFF

```
<file original="9473554c-9e5f-413d-a8f8-5a5155e3bed3" source-language="de-DE" target-language="en-US" datatype="plaintext">
  <header>
    <note>type=TermSetItem</note>
    <note>translatorType=Vendor</note>
    <tool tool-id="SharePoint Server 2016" tool-name="Microsoft® SharePoint® Server 2016" />
    <note>termStoreId=15e113b9-6047-4f36-8736-296e5a7cc3d0</note>
    <note>termSetId=71fa6a73-852c-4e31-861d-261e6250eb7e</note>
    <note>waterline=636432549112182109</note>
    <note>path=Versicherungsprodukte;Berufsunfallversicherung</note>
  </header>
  <body>
    <trans-unit id="TermLabel0" datatype="plaintext">
      <source>Berufsunfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermLabel1" datatype="plaintext">
      <source>Betriebsunfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermLabel2" datatype="plaintext">
      <source>BU</source>
    </trans-unit>
    <trans-unit id="TermLabel3" datatype="plaintext">
      <source>gesetzliche Unfallversicherung</source>
    </trans-unit>
    <trans-unit id="TermDescription" datatype="plaintext">
```

```
8-5a5155e3bed3" source-language="de-DE" target-language="en-US" datatype="plaintext" tool-id="SharePoint Server 2016">
  <tool tool-id="SharePoint Server 2016" tool-name="Microsoft® SharePoint® Server 2016" />
  <note>termStoreId=15e113b9-6047-4f36-8736-296e5a7cc3d0</note>
  <note>termSetId=71fa6a73-852c-4e31-861d-261e6250eb7e</note>
  <note>path=Versicherungsprodukte;Berufsunfallversicherung</note>
  <trans-unit id="TermLabel0" datatype="plaintext">
    <source>Berufsunfallversicherung</source>
    <target>Occupational accident insurance</target>
  </trans-unit>
  <trans-unit id="TermLabel1" datatype="plaintext">
    <source>Betriebsunfallversicherung</source>
    <target>Professional accident insurance</target>
  </trans-unit>
  <trans-unit id="TermLabel2" datatype="plaintext">
    <source>BU</source>
    <target>OAI</target>
  </trans-unit>
  <trans-unit id="TermLabel3" datatype="plaintext">
    <source>gesetzliche Unfallversicherung</source>
    <target>Work accident insurance</target>
  </trans-unit>
```



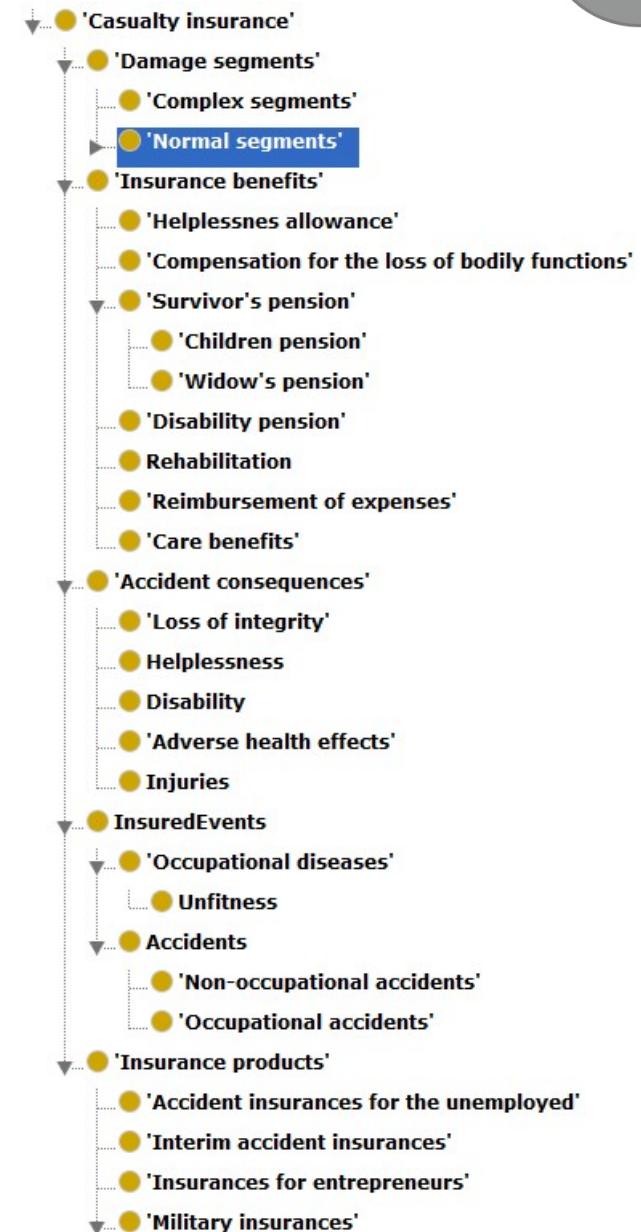
# Taxonomy with Protegé

Translation

- Manual assignment of source language equivalents by taking them over from translated xls



- One-to-one mapping of source taxonomy in target language



# Sharepoint Term Store Management

Translation



**Different sorting due to automatic alphabetical sorting in Sharepoint Term Store Management**

# Sharepoint Term Store Management

Translation

TAXONOMY TERM STORE

German

- Unfallversicherung
  - Schadenssegmente
  - Unfallfolgen
    - Gesundheitsschaden
    - Hilfslosigkeit
    - Integritätsschaden
    - In
    - V
  - Vers
  - Vers
  - Vers
- Hilfslosenentschädigung
- Hinterlassenenrente
- Integritätsentschädigung
- Invalidenrente
- Kostenvergütungen
- Pflegeleistungen

There is already a term with the same default label and parent term.

TAXONOMY TERM STORE

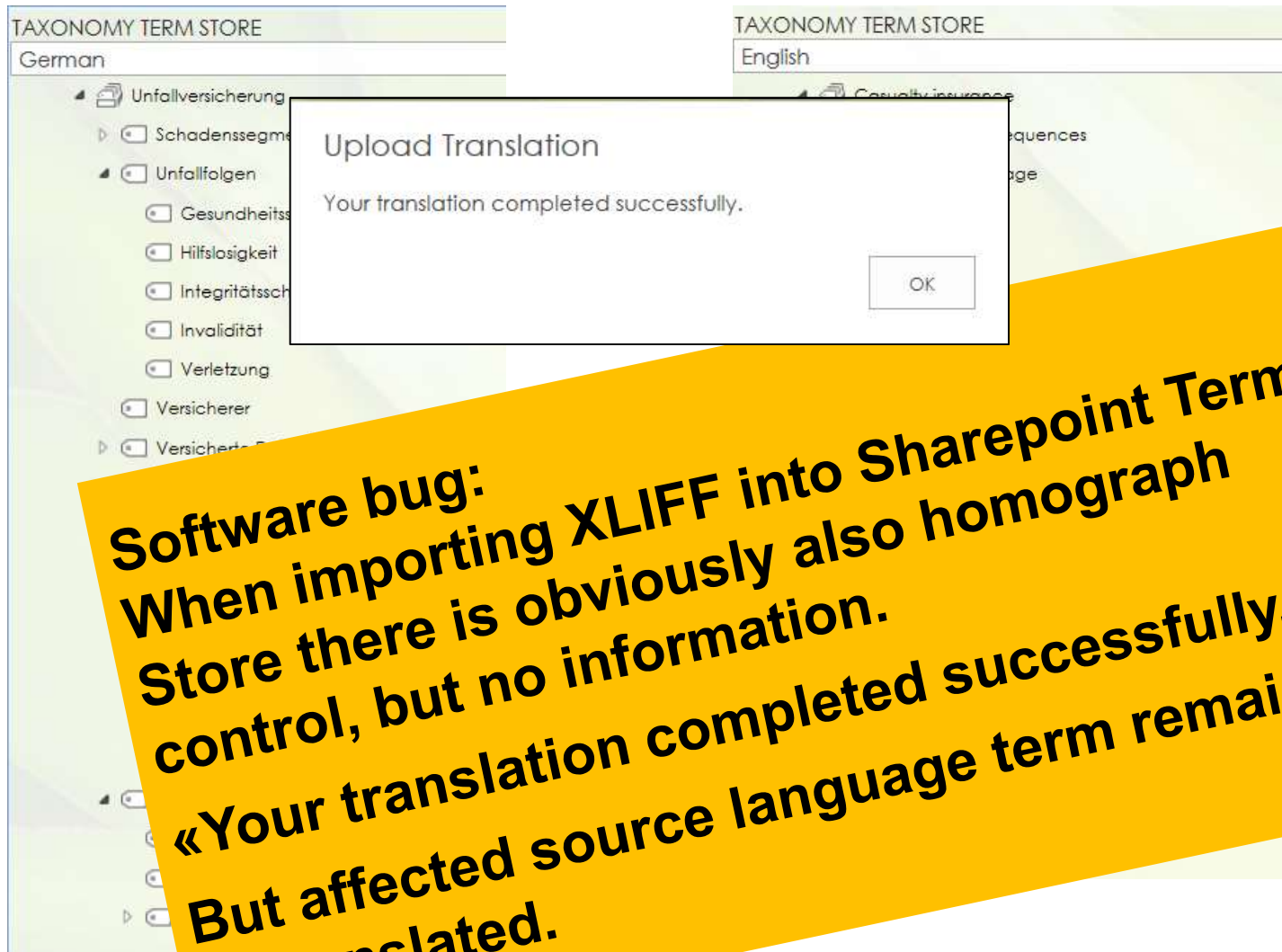
English

- Casualty insurance
  - Accident consequences
    - Health damage
    - Helplessness
    - Injury
    - Invalidity
  - Casualty insurance for unemployed people
  - Military insurance
  - Helplessness allowance
  - Nursing payments
  - Rehabilitation

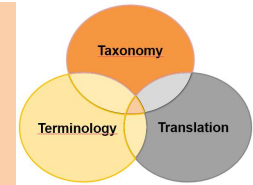
**When manually inserting a new term or translation into Sharepoint Term Store there is homograph control.**

# Sharepoint Term Store Management

Translation

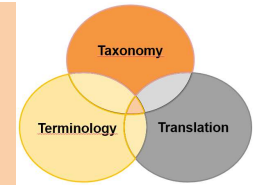


**Software bug:  
When importing XLIFF into Sharepoint Term  
Store there is obviously also homograph  
control, but no information.  
«Your translation completed successfully.»  
But affected source language term remains  
untranslated.**



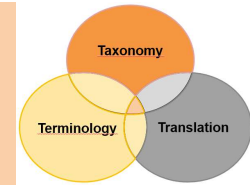
# Conclusion 1

**Terminology supports taxonomy creation: a terminology database provides concepts, preferred terms and part of the semantic relations between them**



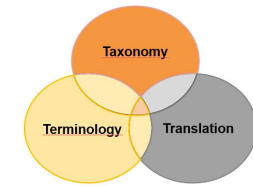
## Conclusion 2

**Language vs. culture: a mere translation of the original language taxonomy is not enough to cover the semantic field of all target cultures**



## Conclusion 3

**The combination of a CAT-Tool and postediting creates economies of scale: the creation of a multilingual taxonomy can be automated to a large extent by using a CAT-Tool; postediting provides for cultural adaptation**



[silvia.giacomotti@suva.ch](mailto:silvia.giacomotti@suva.ch)

[annette.weilandt@eccenca.com](mailto:annette.weilandt@eccenca.com)

[frueh@buerob3.de](mailto:frueh@buerob3.de)

