



Termwise

*CREATING RESOURCES FOR
SPECIALISED USE*

Prof dr Frieda Steurs

TermWise

Project type:

Industrial Research Fund (IOF) KU Leuven

- Knowledge platform: developing practical applications based on multidisciplinary fundamental research
- 800.000 Euro (2009-2013)

Project aims:

- Developing software to support translators in dealing with highly specialised language use (case study: legal texts) based on fundamental research in terminology, translations studies, lexicology, corpus linguistics, computer science and software engineering


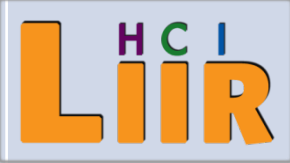



Termwise

Targets :

- Target group: the translation section of the Federal Justice Department (Belgium)
- Target languages: Dutch-French, Belgian legal system
- Authentication

Possible extensions:

- Spinoffs : German community in Belgium
- Other countries and specialised domains (e.g. medical, patents,...)

KU LEUVEN	LINGUISTS	COMPUTER SCIENTISTS
IOF-grant KP/09/001 BASIC RESEARCH	PROMOTORS: <i>Dirk Geeraerts</i> Dirk Spielman  STAFF: Kris Heylen	 PROMOTOR: Sien Moens STAFF: Ivan Vulić
APPLIED RESEARCH	PROMOTORS: Frieda Steurs Hendrik Kockaert  STAFF: Dirk De Hertog	 PROMOTORS: Herman Crauwels Joost Vennekens STAFF: Stephen Bond
USER PARTNER	 Federale Overheidsdienst Justitie Translation Services (Martine Perpet)	

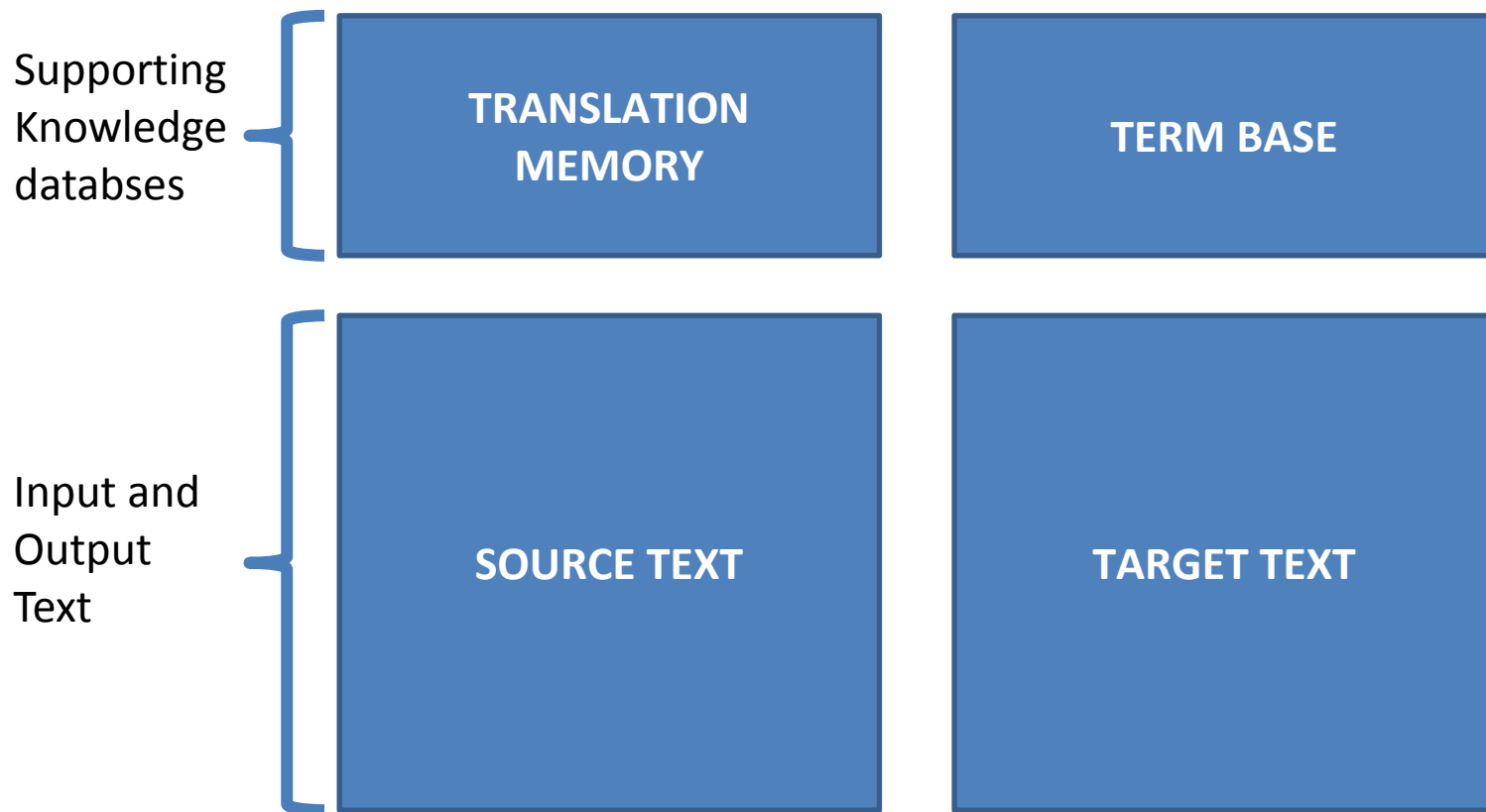
User Needs of Legal Translators

THE TERMWISE PROJECT IS USER-DRIVEN:

- Need for terminological support expressed by the translators at the **Belgian Federal Justice Services**
- Their current state-of-the-art **commercial CAT-tool** gives insufficient support:

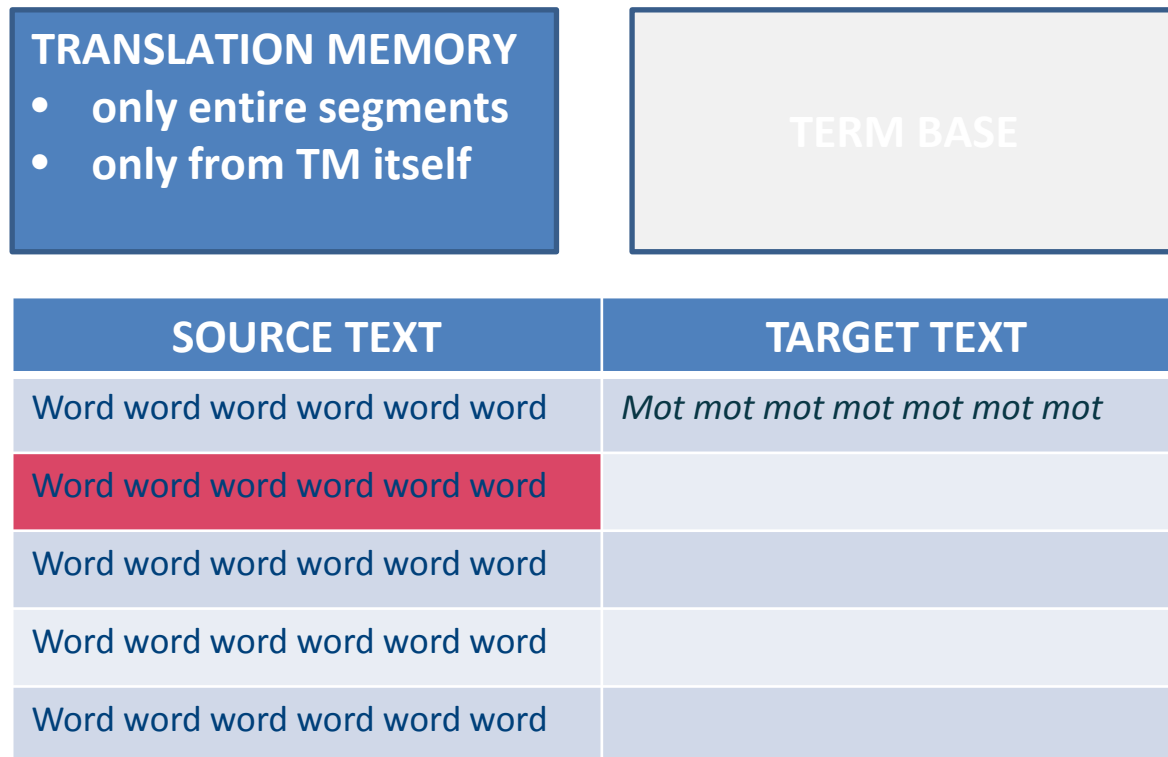
User Needs of Legal Translators

CAT-TOOL (schematic)



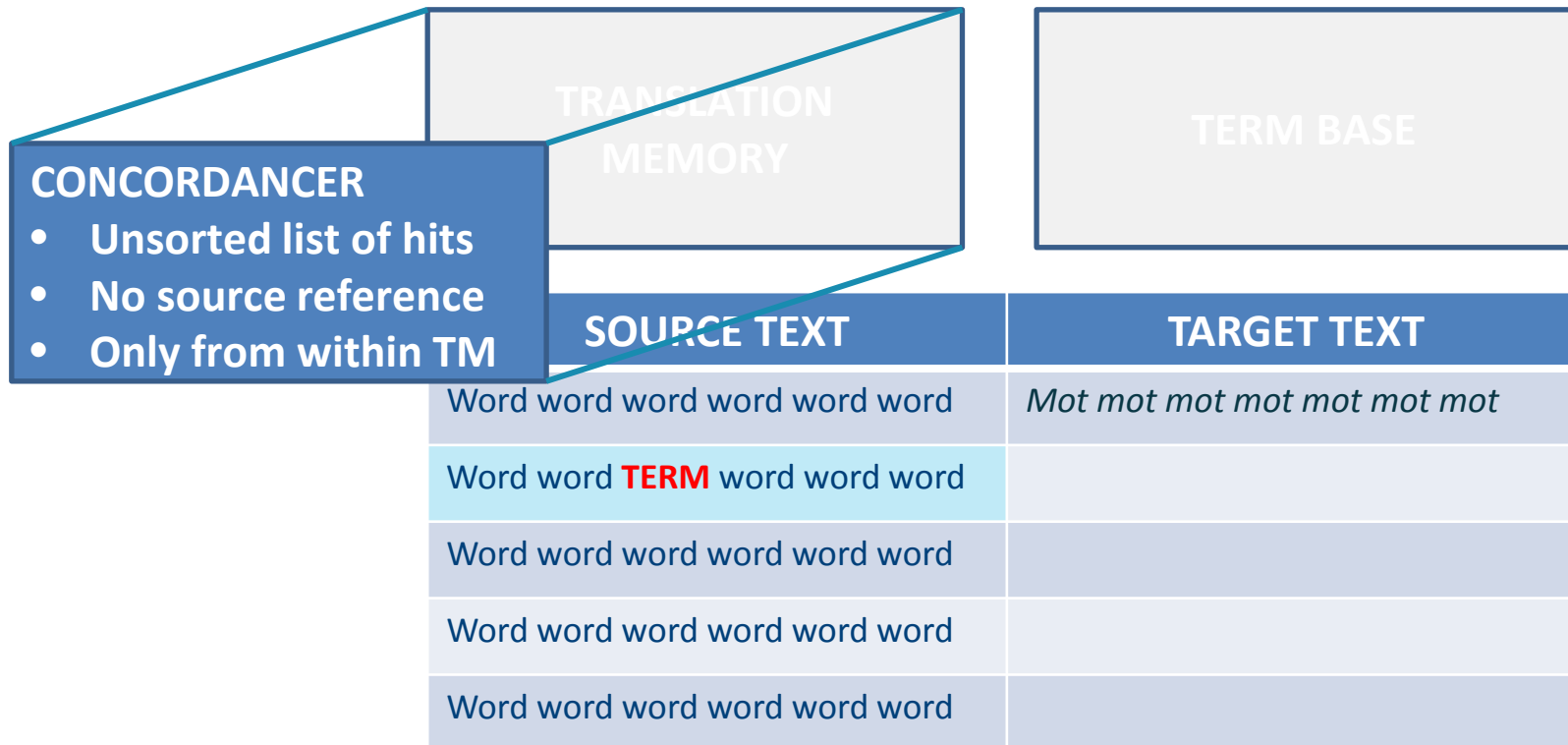
User Needs of Legal Translators

CAT-TOOL (schematic)



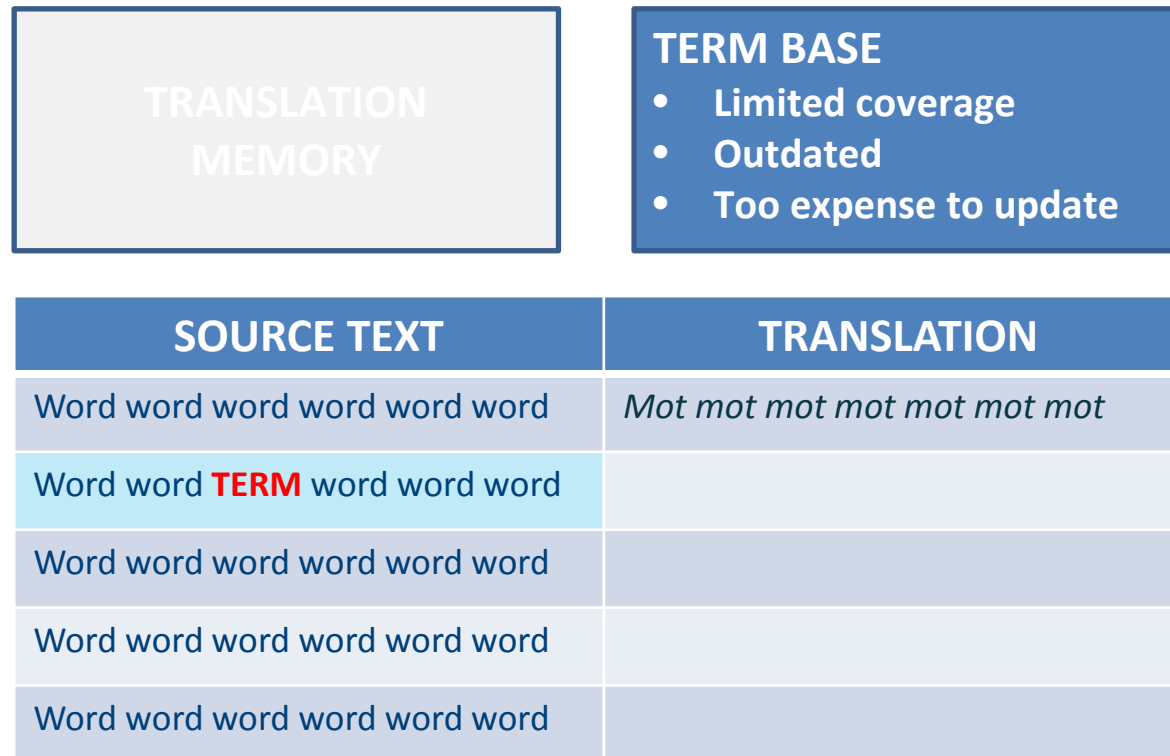
User Needs of Legal Translators

CAT-TOOL (schematic)



User Needs of Legal Translators

CAT-TOOL (schematic)



User Needs of Legal Translators

TERMWISE IS USER-DRIVEN:

- Need for terminological support expressed by the translators at the **Belgian Federal Justice Services**
- Their current state-of-the-art **commercial CAT-tool** gives insufficient support:
 - **Translation Memory**: only gives translation suggestions for entire segments, not for specific terminology
 - **Concordancer**: gives a list of hits from the TM, but not sorted by relevance, without source information, and examples from outside the TM are not included
 - **Termbase**: has limited coverage and is outdated. The department has no resources for terminological work

User Needs of Legal Translators

LARGE ONLINE BILINGUAL DOCUMENTS COLLECTIONS



The image shows a screenshot of the 'Juridat' website. On the left, there are two side-by-side covers of the Belgian legal gazettes: 'BELGISCH STAATSBLAD' and 'MONITEUR BELGE'. Each cover includes a small crest and publication details in both Dutch and French. On the right, the 'Juridat' logo is displayed in a dark blue box, with the text 'J u r i d a t' in white. Below the logo is a stylized icon of a classical building with columns, and the text 'DE KAMER.BE' in a bold, sans-serif font.



TIME CONSUMING MANUAL SEARCHES

User Needs of Legal Translators

IDENTIFIED USER NEEDS:

- Access to previous translations of **domain-specific single and multi-word expressions**
- Examples of **usage in context** to infer correct phraseology
- Information about the **source** documents of the translation examples
- Examples from all relevant **document collections** that are available **online**
- Sorting the examples by **relevance** to the current translation assignment
- **Easy access** to the examples from within the CAT-tool

TermWise project concept

ADDITIONAL KNOWLEDGE BASE

BELGISCH
STAATSBLAD

MONITEUR
BELGE

U r r a t

**TERM & PHRASE
MEMORY**

TRANSLATION
MEMORY

TERM BASE

SOURCE TEXT

TRANSLATION

TermWise project concept

ADDITIONAL KNOWLEDGE BASE



**TERM & PHRASE
MEMORY**

**TRANSLATION
MEMORY**

TERM BASE



- Different types of domain specific expressions
- Looked up automatically
- Previous translations in context and with source reference
- Sorted for relevance to current source text

SOURCE TEXT	TRANSLATION
Word word word word word word	<i>Mot mot mot mot mot mot mot</i>
Word word TERM word word word	
Word MUITIWORD TERM word	
Word word FORMULA word word	
Word word word PHRASEME word	

Knowledge Extraction & Retrieval

TASKS INVOLVED

- Bilingual online documents from the Belgian legal domain
 - *Web crawling of Belgisch Staatsblad/Moniteur Belge*
- Different types of domain-specific expressions
 - *Term Extraction*
- French-Dutch Translation of the expressions
 - *Bilingual Term Alignment*
- Suggestions of translation examples relevant to assignment
 - *Context-sensitive database querying*

TermWise Tool

- Ideally, the Term&Phrase Memory functionality would be integrated in the current CAT-tool of the translators
- IP-restrictions on commercial tools make that unfeasible for now
- A stand-alone interface was implemented as light-weight java-executable that functions as a client to communicate with the TermWise server.
- Uses the segmented xliiff file of the CAT tool as input
- All document comparisons and database querying is done on the server (cloud-based)


Computer assisted translation tool

File View Info


Line	Source
1	Rapport 2012
2	Rapport 2012 en application de l'article 90decies du Code d'instruction criminelle (2011)
3	Table des matières
4	Introduction
5	Conformément à l'article 90decies du Code d'instruction criminelle, le ministre de la Justice est tenu de faire rapport annuellement au Parlement sur l'application des mesures d'écoute, des témoignages anonymes (anonymat complet), de la protection des témoins menacés, des méthodes particulières de recherche et autres méthodes d'enquête.
6	Le présent rapport répond à cette obligation légale.
7	La mission concrète ainsi que la récolte et le traitement des données du rapport sont exposés ci-dessous.
8	Mission
9	L'article 90decies du Code d'instruction criminelle dispose que :
10	« Le Ministre de la Justice fait rapport annuellement au Parlement sur l'application des articles 90ter à 90novies.
11	Il informe le Parlement du nombre d'instructions ayant donné lieu à des mesures visées par ces articles, de la durée de ces mesures, du nombre de personnes concernées et des résultats obtenus.
12	Il fait en même temps rapport sur l'application des articles 40bis, 46ter, 46quater, 47ter à 47decies, 56bis, 86bis, 86ter, 88sexies et 89ter.
13	Il informe le Parlement du nombre d'instructions ayant donné lieu à des mesures visées par ces articles, du nombre de personnes concernées, des infractions concernées et des résultats obtenus.
	Il fait en même temps rapport sur l'application des articles 102 à 111 et 317 et informe les Chambres législatives fédérales du nombre de dossiers, de personnes et d'infractions

Translation Memory Terms and Phrases


70% "« - 833 voor het jaar 2012. » ;"

 -- FEDERALE OVERHEIDSDIENST VOLKSGEZONDHEID, VEILIGHEID VAN DE VOEDSELKETEN EN LEEFMILIEU, "11 JULI 2005 - Koninklijk besluit tot wijziging van het koninklijk besluit van 30 mei 2002 betreffende de planning van het medisch aanbod, bl. 34100." (pdf)

57% "« - 500 voor het jaar 2012. » ;"

 -- FEDERALE OVERHEIDSDIENST VOLKSGEZONDHEID, VEILIGHEID VAN DE VOEDSELKETEN EN LEEFMILIEU, "11 JULI 2005 - Koninklijk besluit tot wijziging van het koninklijk besluit van 30 mei 2002 betreffende de planning van het medisch aanbod, bl. 34100." (pdf)

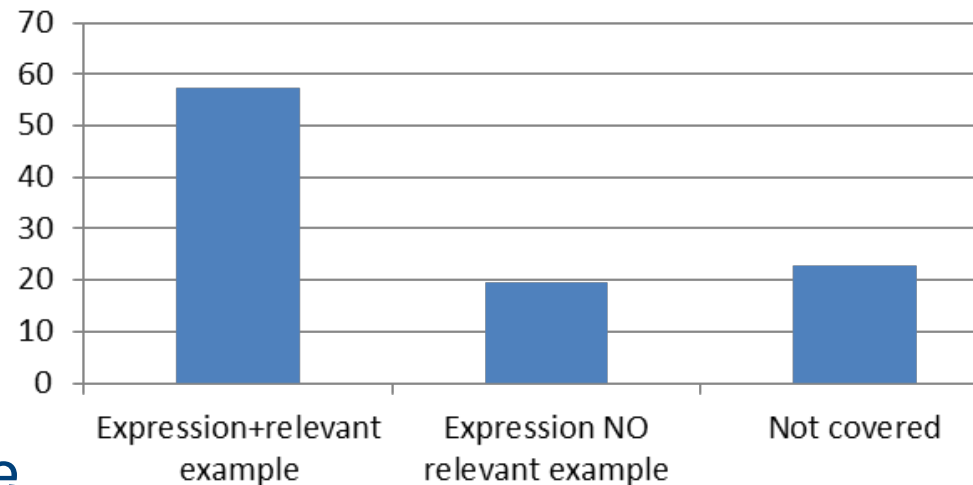
57% "« - 333 voor het jaar 2012. »."

 -- FEDERALE OVERHEIDSDIENST VOLKSGEZONDHEID, VEILIGHEID VAN DE VOEDSELKETEN EN LEEFMILIEU, "11 JULI 2005 - Koninklijk besluit tot wijziging van het koninklijk besluit van 30 mei 2002 betreffende de planning van het medisch aanbod, bl. 34100." (pdf)

TermWise Tool

USER-BASED EVALUATION

- **19 students of legal translation in Antwerp:**
- *For expressions whose previous translations you would like to look up, how many were covered by the Term&Phrase Memory?*
- **5 professional translators at the Federal Justice Services (ongoing right now)**



Conclusion

www.termnet.org

"There is no knowledge without terminology"