Terminology:
Don't only collect it, use it!

Vesna Lušicky & Tanja Wissik
University of Vienna

Translating and the Computer Conference

London, 29 & 30 November 2012

Overview

▸ LISE project

▸ Workflow research

▸ Highlights – needs analysis

▸ LISE Service Version Two

▸ Demo

▸ Outlook

The LISE project

- EU funded project within the ICT-PSP programme

- Consortium: academic, industrial and institutional partners

- Improve the quality of large (inter-institutional or inter-departmental) terminological resources by efficiently cleaning them in a collaborative online environment

- Efficiently spot inconsistencies and gaps

- Collaborate with peers from other institutions

- Benefit from proven best practices

- Achieve interoperability

universität wien

CrossLang
WWW.CROSSLANG.COM

EURAC
research

ESTeam

REPUBLIK ÖSTERREICH
Parlament

Two directions of the project

- Workflow research and best practices in (inter-institutional) terminology work

- Development of a terminology service platform to support inter-institutional terminology work and to (semi)automate processes

User groups and evaluation

▶ **User groups:**

– IATE

– EU Institutions involved in IATE

European Commission, Parliament, Council, Court of Justice. Court of Auditors, Economic & Social Committee, Committee of the Regions, European Central Bank, European Investment Bank, Translation Centre for the Bodies of the EU

– Austrian Parliament

REPUBLIK ÖSTERREICH
Parlament

▶ **Evaluation phase**

CrossLang
WWW.CROSSLANG.COM

Workflow research

- Interviews
    - Interview guideline
    - 17 conducted
    - Recorded–transcribed–analysed

- Survey
    - On-line
    - Available until end of March 2012
    - Disseminated via relevant networks
    - 22 completed

Workflow research

Interviewed institutions

- Auswärtiges Amt (Ministry of Foreign Affairs, Germany)
- Translation Centre for the Bodies of the EU
- DGA3 Terminology and Documentation, General Secretariat, EU Council
- DG TRAD – Terminology Coordination at the European Parliament
- Court of Justice of the European Union, Terminology section
- DG Translation – D.03 - Terminology Coordination at the European Commission)
- Sektion Terminologie der Schweizerischen Bundeskanzlei
- Staatskanzlei des Kantons Bern – Zentraler Terminologiedienst
- Translation Workflow and Terminology Officer – Food and Agriculture Organization of the United Nations
- Ausschuss der Deutschsprachigen Gemeinschaft für die deutsche Rechtsterminologie – Ministerium der Deutschsprachigen Gemeinschaft Belgiens
- An Coiste Téarmaíochta – Foras na Gaeilge
- Terminologicentrum TNC
- Terminology Standardization Directorate/Translation Bureau, Canada
- TERMCAT
- Österreichische Nationalbank (Austrian National Bank)
- Ministerie van Buitenlandse Zaken (Ministry of Foreign Affairs, Netherlands)

Highlights–needs analysis

Maintaining TB	Harmonising and merging TB	Enriching TB	Review with peers – collaborative work
batch export - review – batch import	harmonising	monolingual entries in multilingual TB	communication via e-mails
discrepancy: working method of consolidation vs. workload	merging	crosscheck between TM and TB	communication via phone
redundant entries, redundant/missing information		language specific, subject - oriented projects	informal communication
basic collaborative work			formal communication
manual correction of misspellings			collaboration tool
			extranet

TB = Termbase
TM = Translation memory

Maintaining a termbase

"If we go about cleaning up in about 20 to 30 terms at the time, it will take about 100 years before we get anywhere. We would like to clean up, but frankly, I think, we need some assistance. Currently, with the cleaning up operation, I don't think we can go very far with that." (INT6)

"…this is one side and cleaning is the other side. I mean, content-wise the database is well structured, but needs cleaning because we have lots of duplications, lots of things that are not necessary and others we need and are not there yet." (INT7)

Harmonising and merging

"...of course there is a huge need for coordination and harmonisation of what is done. Especially with regard to feeding the database... and updating the database." (INT9)

(as translated by V.L., T.W.; original: „natürlich großen Bedarf an Koordinierung und Harmonisierung dessen, was gemacht wird. Insbesondere im Hinblick auf die Einspeisung in die Datenbank und … die Aktualisierung der Datenbank" (INT9))

"And when it is larger and all that and if it involves mappings, scripts and things like that, this has to be done on a... Sometimes we would deal with that in batches, depending on the size of the project.." (INT17)

Enriching with new languages

"If you look at the figures for the monolingual, bilingual and trilingual entries, more than half of the database is made of monolingual and bilingual entries, which probably shouldn't be there, which are probably quite useless." (INT6)

"It would be interesting, would be a comparison of the term candidates from the translation memories and a comparison of what is in there [the termbank]." (INT1)

(as translated by VL;TW; original „Was interessant wäre, einen Abgleich zwischen wirklich, was Termkandidaten aus der Translation Memory und Abgleich mit [der Termbank], was schon drinnen ist." (INT1))

Review with peers – collaborative work

"What we maybe lack, is this communication…, this platform, to communicate. " (INT5)

(as translated by VL;TW; original „Was uns vielleicht fehlt momentan, ist so dieses Kommunikations…, diese Plattform, so …zu kommunizieren."(INT5))

"They send us an e-mail or come and talk to us. We are in the same building." (INT6)

"And then they send us the list, "Could you please delete or merge your data on to ours, so that we could eliminate a couple of duplicates in this area." (INT7)

LISE Service Version Two

- LISE Collaboration Portal as a central hub:
  - Access software
  - Discuss issues
  - Exchange data

- Fully featured collaboration functionality

- Semi-automatic collaborative processing – the terminologist has the last word!

The LISE Workflow

Before		
Redundant	Inconsistent	Incomplete

Spot the Flaws		
Doublettes	Inconsistencies	Gaps

Review with Peers		
Clean	Add	Approve

After		
Cleaned	Consistent	Enhanced
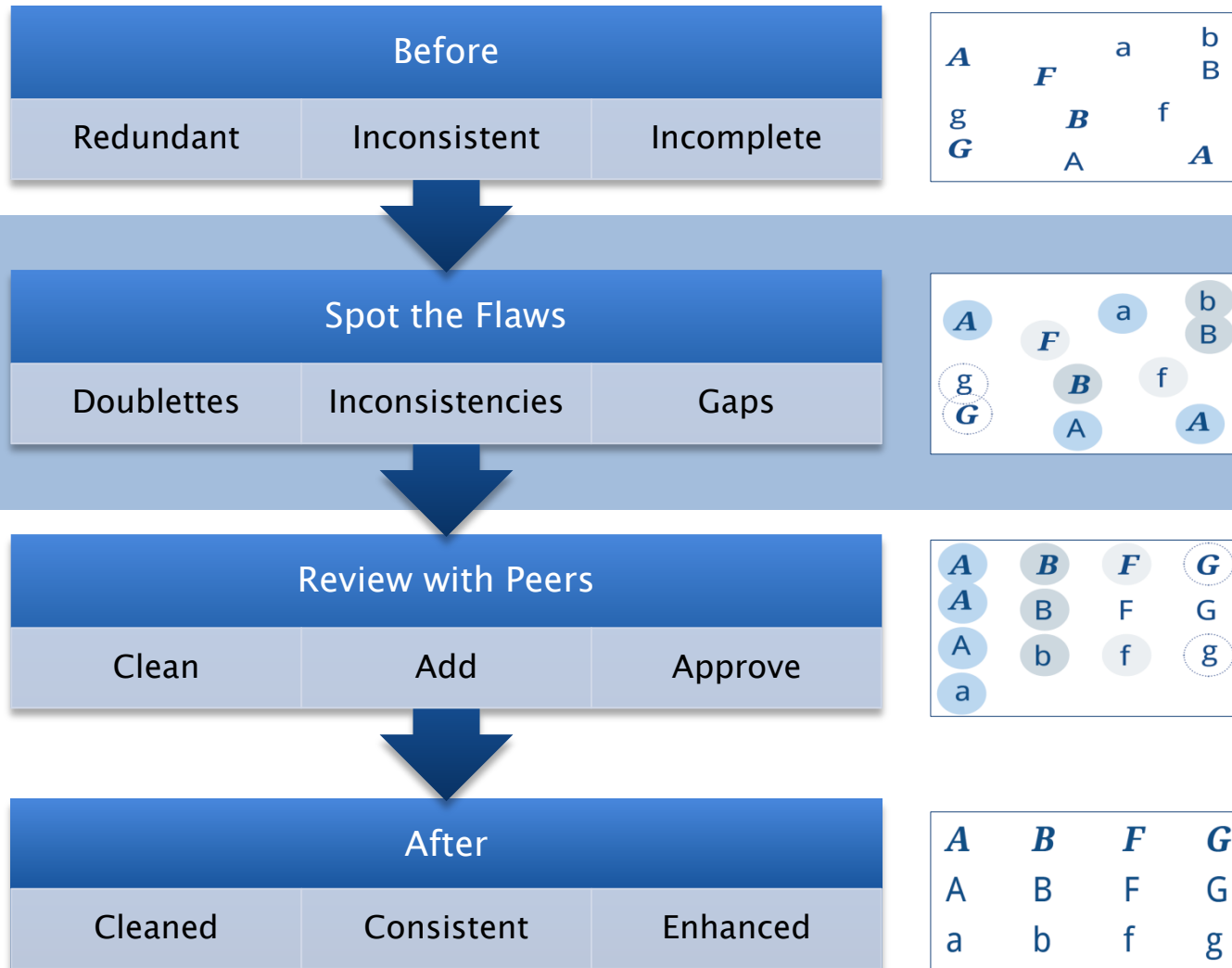
Maintaining a termbase: Cleanup

▸ Misspellings

▸ Missing domain assignments

▸ Equivalents („doublettes")

▸ Subsets (redundant concepts)

▸ Translation errors

▸ Wrong language assignments

Cleanup – spelling error

German spelling error „Ertäge" – should be „Erträge"

Internal harmonising and merging: OMEO

- Identical concepts should no longer be stored separately

- Identical terms in different entries are detected

- Term variants for one concept are discovered in order to be merged

- Results: a more complete collection of terminology, which can be harmonised semi-automatically

- Users can accept or reject the suggestions

Sample: IATE

Apparently doublettes, i.e. redundant entries

Omeo – Identifying Doublettes

Three IATE records suggested to become one.

Fillup

▸ Fillup: Enhance the database into other languages by using translation memory resources

▸ Extraction of translation candidates
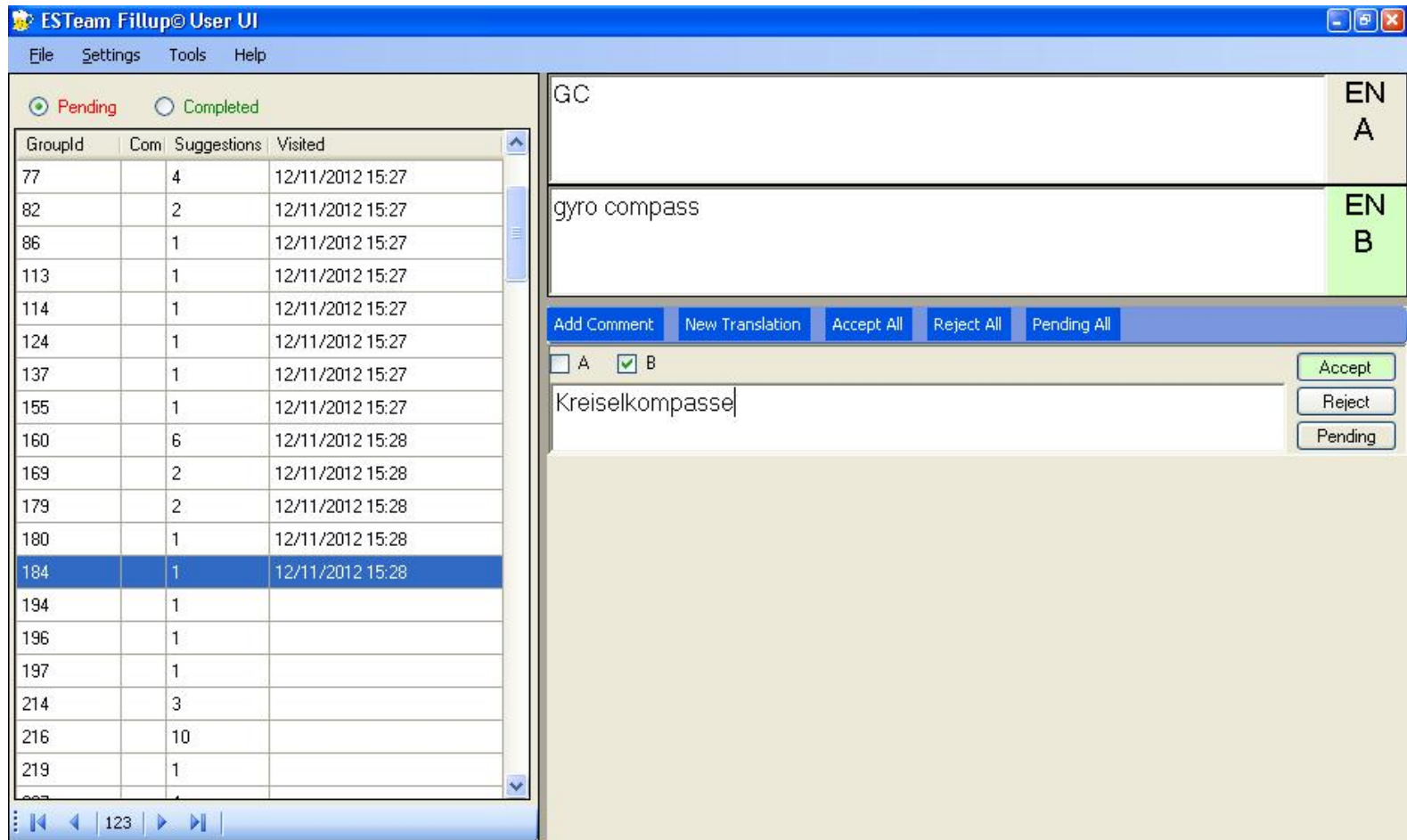
Fillup – Suggesting translations

Out of a translation memory, Fillup suggests for the existing English terms „GC" / „gyro compass" the German term „Kreiselkompasse"

The LISE collaboration platform

- Achieve interoperability & collaboration

- Improve quality of multilingual legal term banks

- Facilitate (inter-institutional) meta-communication

- Workflow-oriented

The LISE collaboration platform

LISE in action: demonstration

Outlook: LISE Service Version Three

- Available in July 2013

- Functionalities to be implemented:
  - Workgroups
  - File attachments to posts
  - Full-text search & filtering

- Feedback on adaptation to functioning with IATE metadata

Further information

Project website: http://www.lise-termservices.eu

Thank you for your attention!

Tanja Wissik   tanja.wissik@univie.ac.at

Deputy Project Coordinator & Researcher, University of Vienna

Vesna Lušicky   vesna.lusicky@univie.ac.at

Researcher, University of Vienna