

# Brug og integration af termekstraktion fra IATE i oversættelsesprogrammer

Tommi Bernitz, Europa-Parlamentet, 13.6.2017

# Tommi Bernitz

Europa-Parlamentet, Den Danske Oversættelsesenhed (1995-)

Oversætter og terminolog

Enhedens terminologikoordinator med ansvar for samarbejdet mellem EU-institutionernes terminologer

Medlem af Terminologigruppen i Danmark

Koordinator for CAT-værktøjer



I. EU-oversættelse, TermCoord

II. Dokumentspecifik termekstraktion

III. Tematiske termbaser

# EU-oversætteres behov for terminologi

- **høj kvalitet og relevant**
- **nem og hurtig adgang**
- **søgning i flere ressourcer på én gang**
- **integreret i værktøjer (CAT-værktøjer)**

# EU-oversætteres værktøjer

- **CAT-værktøjer (SDL Studio, CAT4TRAD)**

**med integrerede søgemuligheder bl.a.:**

- **Quest-søgefunktion med søgning i flere ressourcer på én gang bl.a.**
  - **Euramis-databasen (sætninger)**
  - **IATE og andre termbaser (termer)**
- **Maskinoversættelse MT@EC**

# EP's TermCoord-Enhed - rolle

- at få nye og opdaterede termer ud:
  - til EU-terminologer og -oversættere via:
    - proaktiv terminologi med tematiske termprojekter for de enkelte sprogenheders terminologer
  - til eksterne brugere via <http://termcoord.eu/>
    - [IATE Term of the week](#)
    - [EU-glossarer](#)
    - [Glossary links](#)

# EP's TermCoord-Enhed - opgaver

- at vedligeholde og koordinere indholdet af IATE
- at teste den nye IATE 2 (sommer 2018)
- at teste nye termekstraktionsprogrammer
  
- at koordinere termarbejdet i EP's sprogenheder
- at organisere workshops om terminologi
- at have kontakt til uddannelsesinstitutioner

## II. Dokumentspecifik termekstraktion

SDL Studios termgenkendelse bruges til at integrere termekstraktionen fra IATE:

- dokument til sendes IATE, og man får en termbase (i tre formater: TBX, SDLTB og CSV) tilbage med alle de termer, som IATE har fundet i dokumentet
- SDLTB-format vælges, gemmes og tilføjes i oversættelsesprojektet i SDL Studio



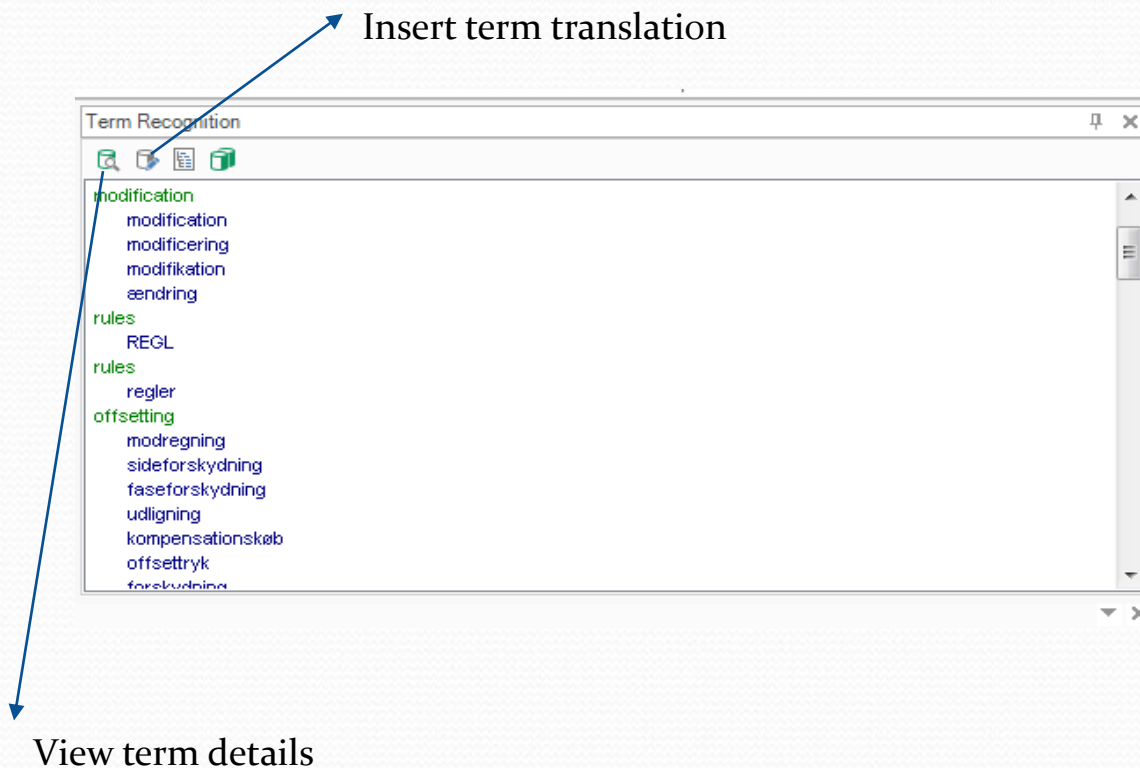
# Studio-interface med termgenkendelse

The screenshot displays the Studio interface for a translation project. The main window shows a translation table with the following content:

Source Text	Target Text	Quality
Furthermore, a modification of the rules on offsetting <a href="#">&lt;fn id="19"&gt;</a> led to the reopening of 21 files in 2016 and resulted in payments of transitional allowance of EUR 752 696.	Endvidere førte en ændring af reglerne om godtgørelse <a href="#">&lt;fn id="19"&gt;</a> til, at 21 sager blev behandlet på ny i 2016 og førte til udbetaling af overgangsgodtgørelser på 752 696 EUR.	100%
The decrease of commitments against Item 1004 'Ordinary travel expenses' over 2015 and the increase of commitments against Item 1005 'Other travel expenses' can both be explained by the change in Parliament's activity calendar.	Både faldet i forpligtelser på konto 1004 'Ordinære rejseudgifter' i forhold til 2015 og stigningen i forpligtelser på konto 1005 'Andre rejseudgifter' kan forklares ved ændringen i Parlamentets aktivitetskalender.	97%
The increase of external parliamentary activities weeks led to a decrease in the level of expenditure concerning ordinary travels whereas travels in the Member State of election tended to increase.	Stigningen i uger med eksterne parlamentariske aktiviteter førte til et fald i udgifter til ordinære rejser, mens rejser i den medlemsstat, hvor medlemmer var valgt, havde tendens til at stige.	98%
In 2016, Members made a total of 29 641 journeys <a href="#">&lt;fn id="17"&gt;</a> .	Medlemmeme foretog i 2016 i alt 29 641 rejser <a href="#">&lt;fn id="17"&gt;</a> .	98%
Most journeys were made to Brussels (18 615) and to Strasbourg (8 687).	Målet for de fleste rejser var Bruxelles (18 615) og Strasbourg (8 687).	98%
There were 1 348 and 991 journeys, respectively, to attend meetings elsewhere in the European Union and outside the European Union.	Møder andre steder i Den Europæiske Union udgjorde 1 348 rejser, og der blev foretaget 991 rejser i forbindelse med møder uden for Unionen.	98%
Based on Members' reimbursement claims as at 13 February 2017.	På baggrund af medlemmernes anmodninger om godtgørelse indtil den 13. februar 2017.	100%
A journey is connected to a Member's participation at a single meeting place (ordinary travel).	En rejse er knyttet til et medlems deltagelse på et enkelt mødested (ordinære rejser).	100%
Item 1006 is intended to cover the cost of Members' parliamentary activities.	Konto 1006 skal dække udgifter til medlemmernes parlamentariske aktiviteter.	90%
This allowance amounted to EUR 4 320 a month for each Member in 2016 (unchanged over 2015) and was paid monthly at Members' request.	Denne godtgørelse beløb sig til 4 320 EUR pr. måned og pr. medlem i 2016 (uændret i forhold til 2015) og blev udbetalt månedligt efter anmodning fra medlemmet.	
The reduction against Item 1020 'Transitional allowances' between 2015 and 2016 reflects the phasing-out effect of Members who left office after the elections in 2014 and applied for a transitional allowance <a href="#">&lt;fn id="18"&gt;</a> .	Reduktionen på konto 1020 'Overgangsgodtgørelser' mellem 2015 og 2016 afspejler den gradvise reducerende virkning af medlemmer, som forlod Parlamentet efter valget i 2014 og ansøgte om overgangsgodtgørelse <a href="#">&lt;fn id="18"&gt;</a> .	
In December 2016, only 2 former Members were paid a transitional allowance, compared to 12 in July 2016, 25 at the end of December 2015 and still 51 in July 2015.	I december modtog kun 2 tidligere medlemmer en overgangsgodtgørelse sammenlignet med 12 i juli 2016, 25 ved udgangen af december 2015 og endnu 51 i juli 2015.	
Furthermore, a modification of the rules on offsetting <a href="#">&lt;fn id="19"&gt;</a> led to the reopening of 21 files in 2016 and resulted in payments of transitional allowance of EUR 752 696.	Endvidere førte en ændring af reglerne om godtgørelse <a href="#">&lt;fn id="19"&gt;</a> til, at 21 sager blev behandlet på ny i 2016 og førte til udbetaling af overgangsgodtgørelser på 752 696 EUR.	
This allowance is paid in accordance with Article 9 of the Statute for Members of the European Parliament, whose terms can be summarised as follows:	Denne godtgørelse ydes i medfør af artikel 9 i statuten for Europa-Parlamentets medlemmer, hvis betingelser kan sammenfattes som følger:	

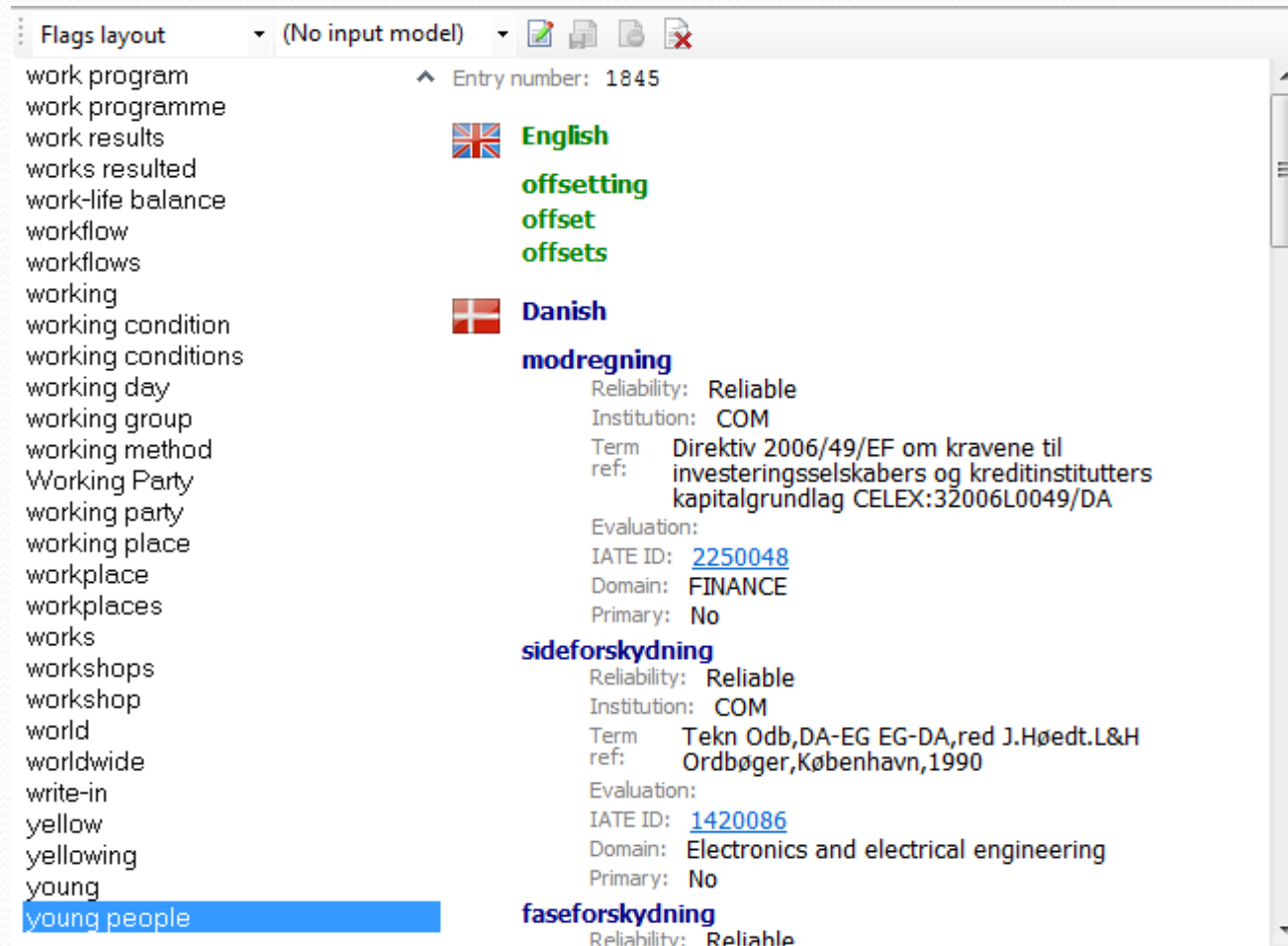
The right-hand panel shows the Term Recognition interface, listing terms and their corresponding document numbers and types.

# Studio interface – Term Recognition-vindue



**Term Recognition-vindue** kan placeres hvor som helst. Markering af term og valg af Insert term translation indsætter termen i Studio. View term details giver flere IATE-oplysninger om term.

# Studio interface – View term details



IATE-oplysninger:

Reliability

Institution

Reference

Domæne

IATE-link

(valgfri men tager meget plads på skærmen)

# Pilotforsøg med dokumentsspecifik termekstraktion fra IATE (jan-feb. 2017)

- 19 sprogenheder – 70 personer

Vigtigste resultater:

- gjorde livet lettere – øget kvalitet og konsistens i tekster
- færre manuelle termsøgninger
- identificering af ord, som man ikke ville have betragtet som termer
- post-2004 målsprog langt mindre støj end pre-2004
- generel brug kræver uddannelse af brugere (indstillinger)

# Ulemper/problemer

- støj især pre-2004 (især EN, DE, FR,) med mange termer i IATE pga.:
  - homonymer
  - duplikater
  - almindelige ord / småord genkendes som termer
- længere termer med flere ord genkendes ikke
- uhensigtsmæssig rangordning (efter rækkefølge i teksten)
- ikke forskel mellem store og små bogstaver who vs. WHO

# Løsninger/fremtidig udvikling

- løsning af støjproblemet for især pre-2004 (EN, DE, FR,)
  - stopliste (med de mest almindelige ord)
    - findes allerede, kan evt. udvides
  - domæne
    - problematisk pga. IATE's domænestruktur (Eurovoc)
  - IATE-pålidelighedskode
  - institution

Er det værd at bruge energi på at rense ud?

- Interinstitutionelt samarbejde om integrering af Studio og IATE

## II. Tematiske termbaser - behov

- specifik, afgrænset pålidelig terminologi
  - i specifikke emneområder
  - i specifikke lovgivningsprocedurer
    - Kommissionens forslag sendes til Parlamentet (EP), derefter til Rådet
    - EP er ofte bundet af Kommissionens terminologi; det er vigtigt, at der anvendes den samme terminologi
- sikres ved en manuel udvælgelse og validering af termer fra referencedokumentet:
  - f.eks. basisretsakter, Kommissionens forslag

# Tematiske termbaser - fremgangsmåde

- manuelt oprettede tematiske termbaser vha. et program for termekstraktion Syncroterm der foreslår en række termer i en liste,
- der eksporteres til programmet Glossary Converter,
- der omdanner listen til en termbase til Studio SDLTB-format, der bruges på samme måde som i den dokumentbaserede termekstraktion fra IATE.



# Eksempler på tematiske termbaser

- Finansforordningen
- Den flerårige finansielle ramme
- Den Europæiske Globaliseringsfond
- Emissioner fra biler (EMIS)

# Konkret eksempel

- Ved oversættelse af tekster om Den Europæiske Globaliseringsfond (EGF) er det vigtigt at vide, om hvordan ”financial contribution” oversættes i forbindelse med EGF:
  - termbasen med udelukkende termer fra basisretsakten<sup>1</sup> om oprettelse af EGF indeholder termen ”økonomisk støtte”
  - ikke behov for at søge andre steder

# Eksempel - illustration

1125436\_Upload\_EN-DA,1125436\_BRTM\_EN-DA,FullCat translation,fullcatTM,1125436\_restEN\_DA\_en-GB\_da-DK,1125436\_WTM\_EN-DA - Translation Results

Project Settings

whereas the adoption of the EGF Regulation reflects the agreement reached between the Parliament and the Council to reintroduce the crisis mobilisation criterion, to set the Union financial contribution to 60% of the total estimated cost of proposed measures, to increase efficiency for the treatment of EGF applications in the Commission and by the Parliament

whereas the adoption of the EGF Regulation reflects the agreement reached between the Parliament and the Council to reintroduce the crisis mobilisation criterion, to set the Union financial contribution to 60% of the total estimated cost of proposed measures, to increase efficiency for the treatment of EGF applications in the Commission and by the Parliament and the Council by shortening the time for assessment and approval, to widen eligible actions and beneficiaries by introducing self-employed persons and young people and to finance incentives for setting up own businesses;

100% der henviser til, at vedtagelsen af EGF-forordningen afspejler den aftale, der er indgået mellem Parlamentet og Rådet om at genindføre kriseanvendelseskriteriet, fastsætte Unionens økonomiske støtte til 60% af de samlede anslåede omkostninger ved de foreslåede foranstaltninger, øge effektiviteten af behandlingen af EGF-ansøgninger i Kommissionen og af Parlamentet og Rådet ved at forkorte perioden for vurdering og godkendelse, udvide omfanget af støtteberettigede aktioner og modtagere ved at inkludere selvstændige og unge og yde støtte til incitament til virksomhedsopretning;

Doc. No. 1117166  
Obs. 2017/2014(BUD)  
index F. shared Budget-Discharge  
date F. 17:20 10/02/2017  
user F. tbernitz  
Doc.Type F.

fullcatTM 2/10/2017 5:20:00 PM tbernitz(shared)

1125436\_Upload\_EN-DA,1125436\_BRTM\_EN-DA,FullCat translation,fullcatTM,1125436\_restEN\_DA\_en-GB\_da-DK,1125436\_WTM\_EN-DA - Translation Results

1117166EN docx.sdxliff [Translation]

33	having regard to the letter of the Committee on Employment and Social Affairs,		
34	—		P
35	having regard to the letter of the Committee on Regional Development,		
36	—		P
37	having regard to the report of the Committee on Budgets (A8-0036/2017),		
38	A.		P
39	whereas the Union has set up legislative and budgetary instruments to provide additional support to workers who are suffering from the consequences of major structural changes in world trade patterns or of the global financial and economic crisis and to assist their reintegration into the labour market;		
40	B.		P
41	whereas the Union's financial assistance to workers made redundant should be dynamic and made available as quickly and efficiently as possible, in accordance with the Joint Declaration of the European Parliament, the Council and the Commission adopted during the conciliation meeting on 17 July 2008, and having due regard to the IIA of 2 December 2013 in respect of the adoption of decisions to mobilise the European Globalisation Adjustment Fund (EGF);		
42	C.		P
43	whereas the adoption of the EGF Regulation reflects the agreement reached between the Parliament and the Council to reintroduce the crisis mobilisation criterion, to set the Union financial contribution to 60% of the total estimated cost of proposed measures, to increase efficiency for the treatment of EGF applications in the Commission and by the Parliament and the Council by shortening the time for assessment and approval, to widen eligible actions and beneficiaries by introducing self-employed persons and young people and to finance incentives for setting up own businesses;	100%	der henviser til, at vedtagelsen af EGF-forordningen afspejler den aftale, der er indgået mellem Parlamentet og Rådet om at genindføre kriseanvendelseskriteriet, fastsætte Unionens økonomiske støtte til 60% af de samlede anslåede omkostninger ved de foreslåede foranstaltninger, øge effektiviteten af behandlingen af EGF-ansøgninger i Kommissionen og af Parlamentet og Rådet ved at forkorte perioden for vurdering og godkendelse, udvide omfanget af støtteberettigede aktioner og modtagere ved at inkludere selvstændige og unge og yde støtte til incitament til virksomhedsopretning;

Term Recognition

- financial contribution
- økonomisk støtte
- measures
- foranstaltninger
- Applications
- Ansøgninger
- Eligible actions
- Støtteberettigede aktioner
- beneficiary
- støttemodtager
- self-employed persons
- selvstændige

# Nuværende anvendelse

- i testfasen i nogle enheder for oversættelser om ændring af finansforordningen
- nemt at anvende Syncroterm:
  - efter træning - ca. 1 time at lave en termbase (f.eks. 40 termer)
  - et sprogpar EN-ET kan nemt duplikeres til andre EN-DA, etc.
- deling af termbaser i SDL Studio
  - flere brugere samtidigt tilkoblet samme base
  - flere brugere kan samtidigt tilføje termer til samme base

# Fremtidig anvendelse

- Systematisk anvendelse:
  - på enhedsplan: filkoordinator til central terminologi, f.eks. EN-DA
  - i EP: udbredes til andre sprogenheder, f.eks. EN-FR, EN-DE etc.
- Udveksling af tematiske termbaser:
  - mellem EU-institutionerne
  - lovgivningsprocedurer: Kommissionen opretter basen og sender den videre til EP og Rådet
  - sikring af kohærent terminologi og øget kvalitet

# Hindringer/muligheder for anvendelse

- Hindringer for anvendelse:
  - Manglende licenser til programmer pga.
    - finansielle ressourcer
    - IT-sikkerhed
  - Manglende menneskelige ressourcer:
    - terminologer
- Muligheder
  - Opbakning fra oversættere og terminologer
  - Nemmere at integrere egne termbaser i CAT-værktøjer (Multiterm)

# Spørgsmål?

[tommi.bernitz@europarl.europa.eu](mailto:tommi.bernitz@europarl.europa.eu)

Tak for jeres opmærksomhed 😊