



From lexicography to translation via proofing tools and terminology: an enjoyable journey

Euralex 40th Anniversary Online Lecture Series
13 October 2023

Thierry Fontenelle

t.fontenelle@eib.org

Thierry.fontenelle@hotmail.com

Who am I?

Currently Head of the Linguistic Services Division
at the [European Investment Bank](#) (EIB)
(Luxembourg)

(EIB = the EU Bank, a.k.a. the climate bank)



Academia, private sector and international organisations

Education

- Studied English & Dutch languages & literature (University of Liège, Belgium, 1986)
- English translation (University of Liège, 1987)
- PhD in English linguistics (University of Liège, 1995)

Languages and multilingualism

Career in a nutshell

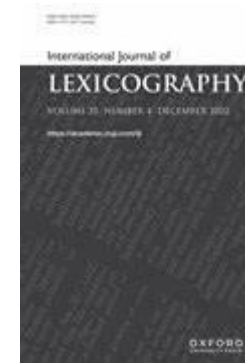
- 1986-1996: University of Liège
- 1991-1992: Siemens (MT development) (Leuven/Liège)
- 1996-1999: European Commission Translation Service, Luxembourg
- 1999-2001: NAMSA (NATO Agency), Luxembourg
- 2001-2009: Microsoft Natural Language Group (Redmond, USA)
- 2009-2019: Translation Centre for the Bodies of the European Union (Luxembourg)
- 2019 – now: European Investment Bank (EIB), Linguistic Services Division (Luxembourg)

My relationship with Euralex & IJL

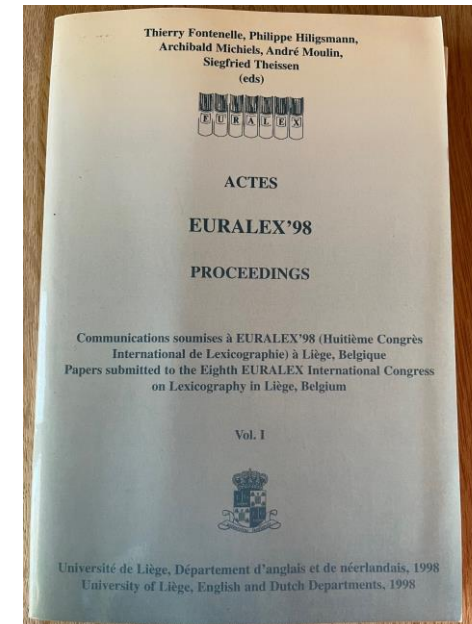


- Joined Euralex in 1990 (Malaga)
 - Member of the Euralex Executive Board: 1994-2006
 - Secretary-treasurer : 1998-2000
 - Vice-President : 2000-2002
 - President: 2002 – 2004
-
- Co-organised Euralex 1998 at the University of Liège (Belgium) in 1998

Fontenelle, Th., Hiligsmann, Ph., Michiels, A., Moulin, A. & Theissen, S. (eds) *Actes EURALEX'98 Proceedings (Proceedings of the 8th International Congress of the European Association for Lexicography, EURALEX'98)*, [Volume 1](#) – [Volume 2](#), Université de Liège, Liège, 1998, 674 pp.



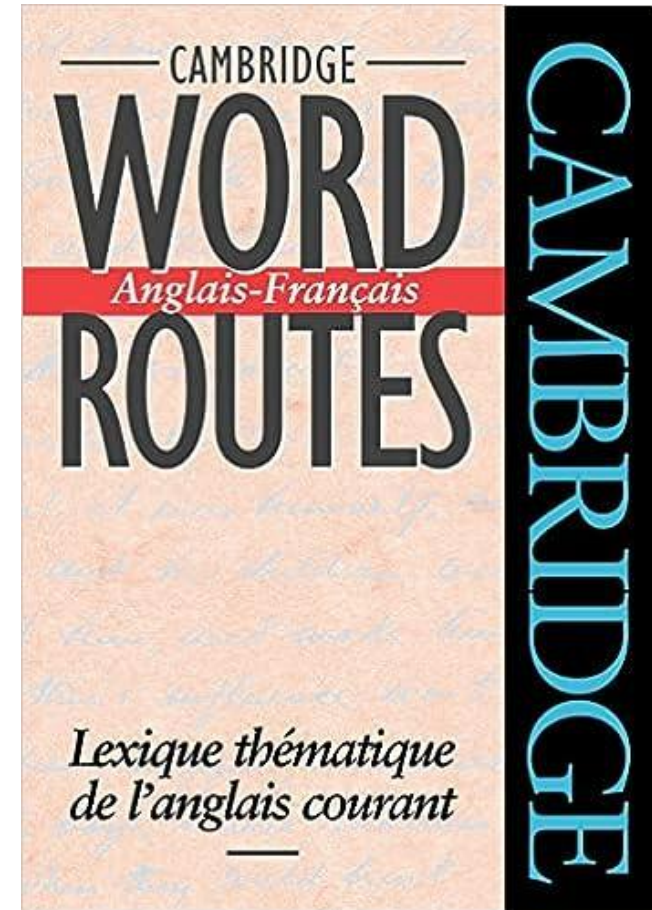
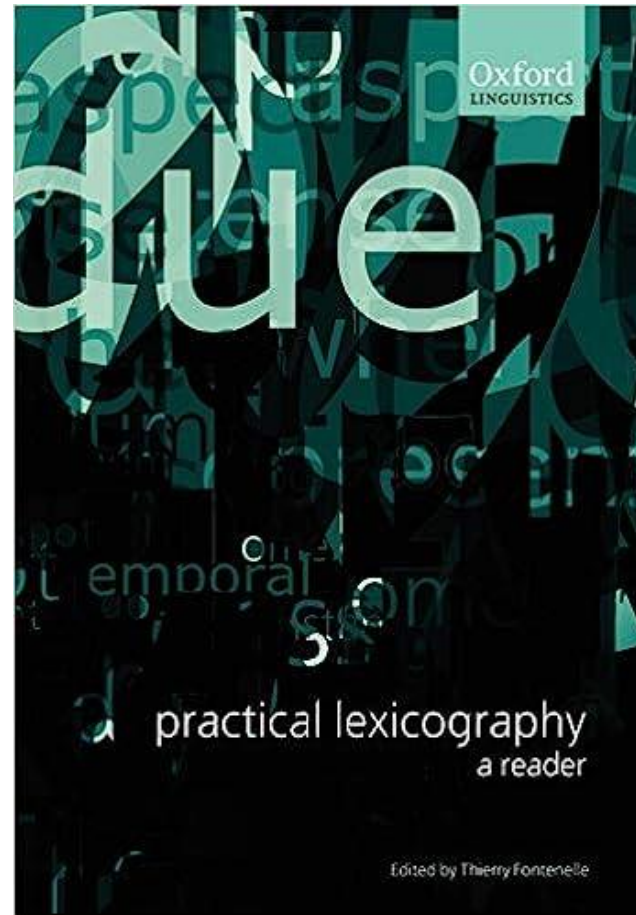
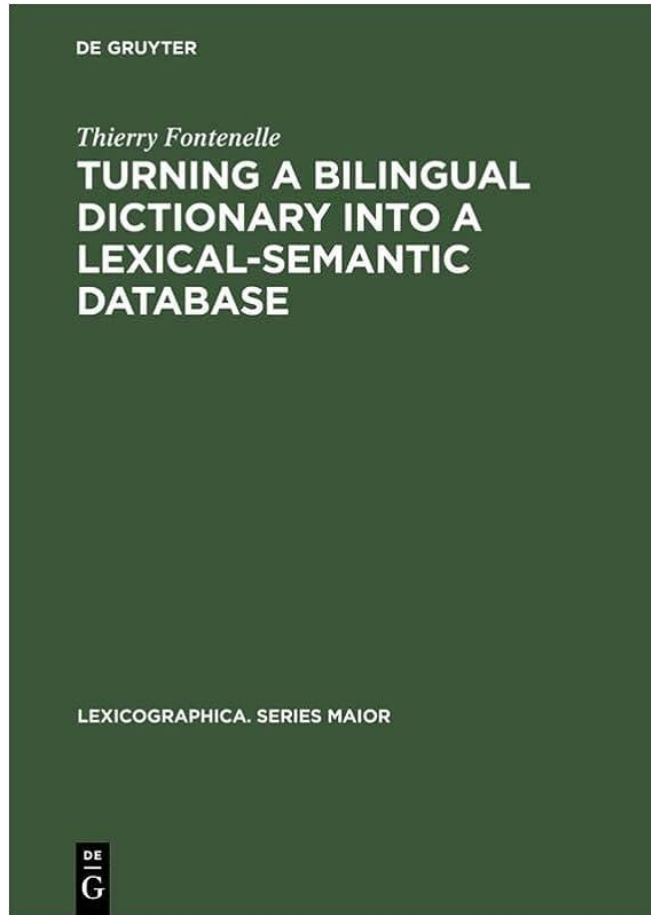
- International Journal of Lexicography (IJL, Oxford University Press)
- Member of IJL Editorial Board between 1994-1997 and Associate Editor since 1998



13 papers in Euralex proceedings

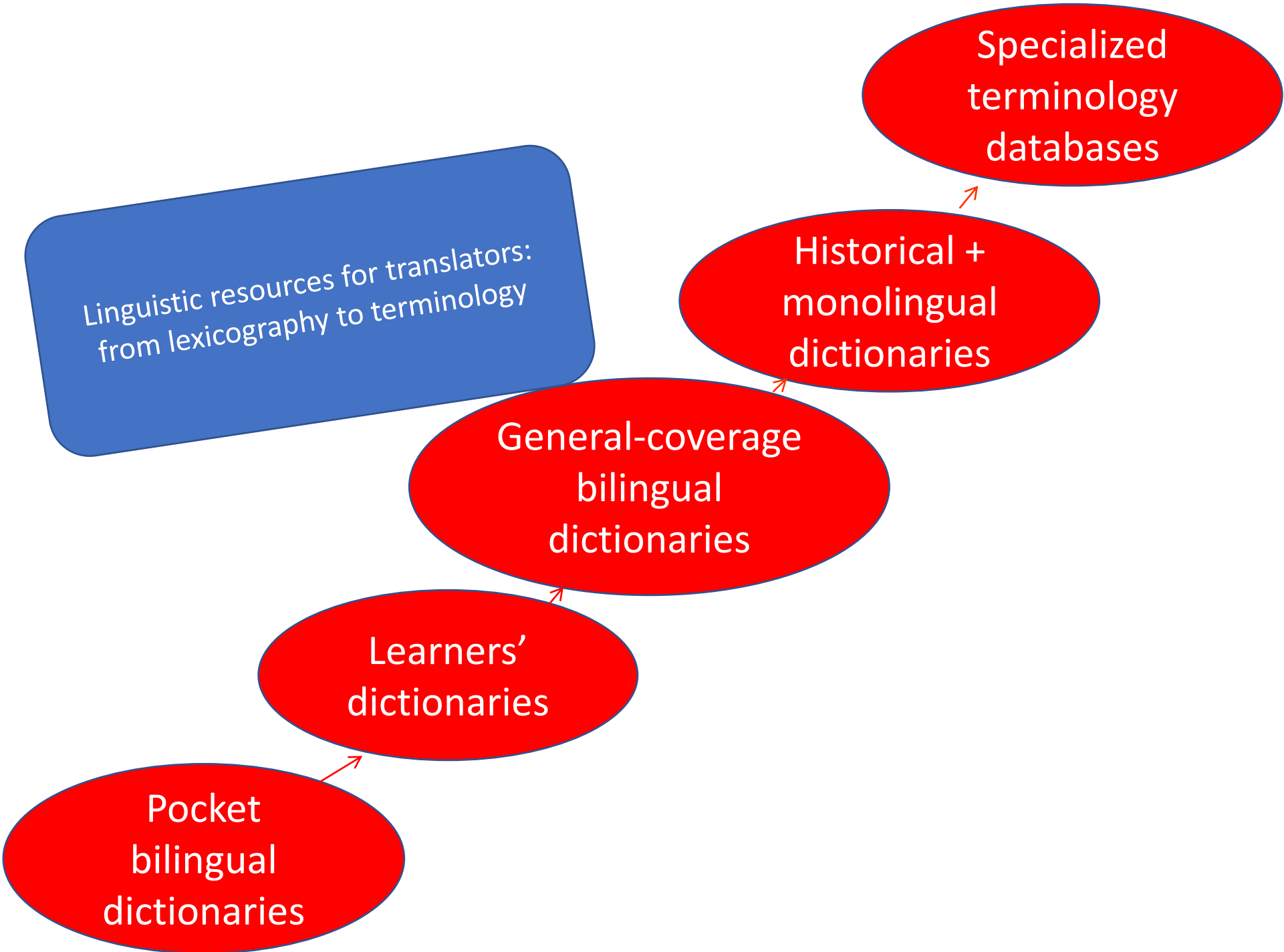
- Malaga (1990): [Automatic extraction of lexical-semantic relations from dictionary definitions](#)
- Tampere (1992): [Collocation acquisition from a corpus or from a dictionary: a comparison](#)
- Amsterdam (1994): [Using lexical functions to discover metaphors](#)
- Göteborg (1996) (2 papers):
 - [Ergativity, collocations and lexical functions](#)
 - [The DECIDE project: Multilingual Collocation Extraction](#) (Grefenstette, G., Heid, U., Schulze, B.M., Fontenelle, T. & Gérardy, C.)
- Liège (1998): [The semantic analysis of "of"-phrases for word sense disambiguation](#)
- Stuttgart (2000): [Extracting Phraseology for Content Analysis and Document Retrieval](#)
- Copenhagen (2002): [Using dictionary grammar codes to resolve attachment ambiguities](#) (Fontenelle, Th. & Kharrat, A)
- Lorient (2004): [Lexicalization for proofing tools](#)
- Torino (2006): [Developing a lexicon for a new French spell-checker](#)
- Barcelona (2008): [Lexicon Creator: A tool for building lexicons for proofing tools and search technologies.](#) (Fontenelle, T., Cipollone, N., Daniels, M. & Johnson, I.)
- Bolzano (2014): [From lexicography to terminology: a cline, not a dichotomy,](#)
- Tbilissi (2016): [Aspects lexicaux d'une langue inventée : la langue des Schtroumpfs](#)

A few publications...



Languages and multilingualism

- Teaching (EFL)
- Basic research (linguistics, NLP, computational lexicography)
- Applied research & development (computational linguist/lexicographer, NLP developer & program manager)
- Lexicography
- Terminology
- Interpreting
- Translation & translation management



Evolution of my own research/publications

- Lexical acquisition for NLP
 - From learners' dictionaries
 - From bilingual dictionaries
- Developing rule-based Machine Translation
- Developing proofing tools
 - Grammar checkers
 - Spell-checkers
 - Lemmatisation for search engines
- Terminology & translation

Main problem: the lexicon

- For many (most?) NLP applications, the lexicon is the core component
- Lexicon building = time-consuming, labor-intensive process

Should I say
“WAS” ?

AI
Machine Learning
Large Language Models
(LLMs)
ChatGPT
...

Lexical acquisition and electronic dictionaries (1980s & 1990s)

- Extraction of syntactic and semantic information for NLP
 - Monolingual dictionaries (English: LDOCE/COBUILD/CIDE/OALD...)
 - ➔ Building lexicons for:
 - * Machine Translation
 - * WSD (Word sense disambiguation)
 - * Information Retrieval (IR)
- [Boguraev/Briscoe 89; Wilks et al.96, Acquilex project...]

Lexical Acquisition for NLP

- ‘Lexical acquisition bottleneck’
(Boguraev & Briscoe ’89; Wilks et al ’96; Byrd et al. 1987, Calzolari-Pisa...)
- Source of data for building NLP Lexicons:
 - Acquisition of syntactic information (sub-categorisation, complementation...) from existing (bilingual/monolingual) dictionaries available in Machine-Readable form (MRDs)
 - Big textual corpora
- extraction of collocational knowledge from dictionaries and corpora (Church & Hanks ’89; Church *et al.* ’94; Fontenelle ’97; Grefenstette *et al.* 96)

Computational
Lexicography

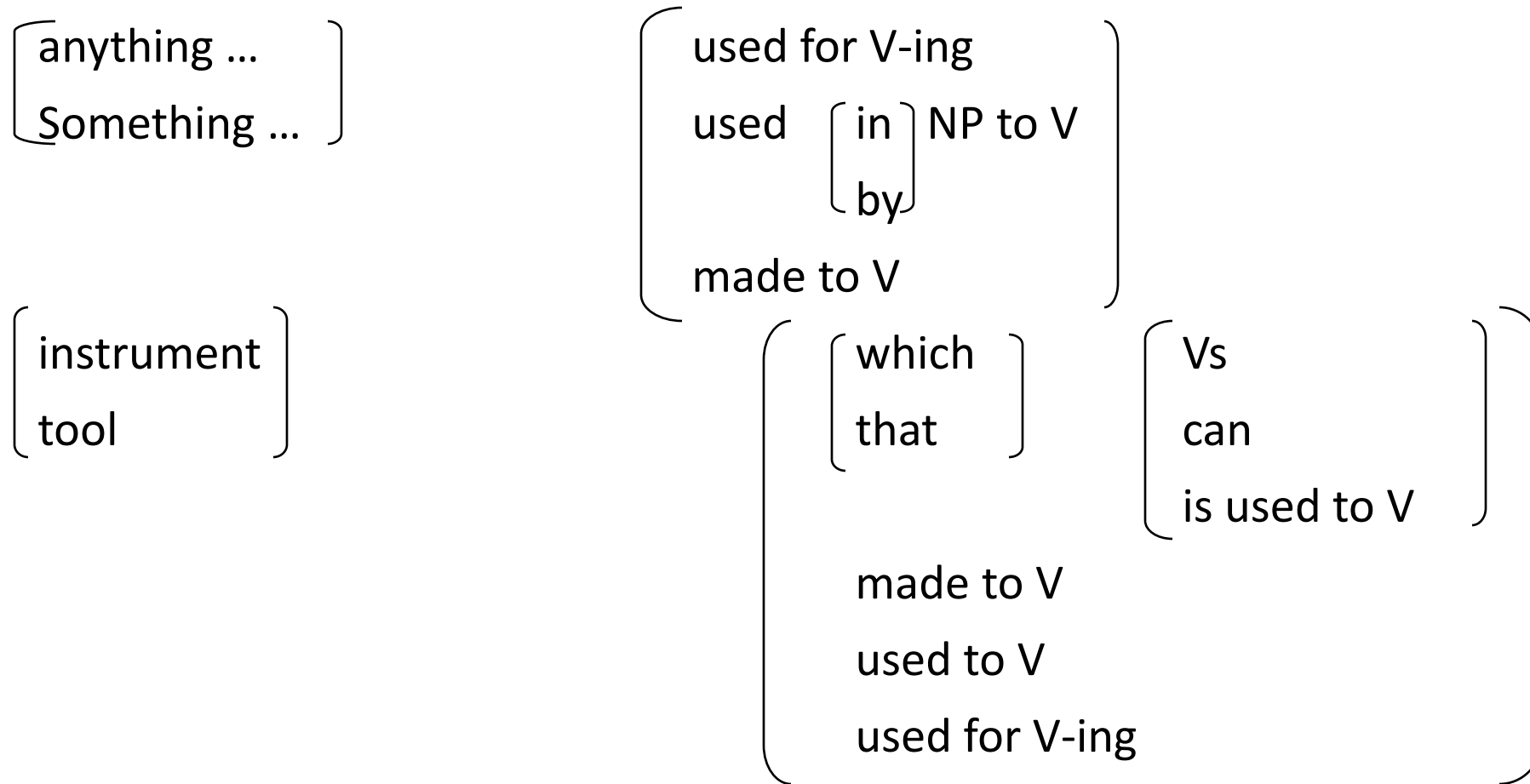
- 1980s/1990s: from MRD
(typesetting tapes) to LKB
- Today: from DB to print

Exploitation of machine-readable dictionaries (MRDs) transformed into lexical databases (LKBs)

- Initially with Merriam Webster's dictionary (Amsler 1980), then LDOCE 1973, probably the first truly computerized dictionary (Michiels 1982, then Boguraev & Briscoe 1989)
- LDOCE: used by research teams in Liège (Michiels 1982), Cambridge (Boguraev & Briscoe 1992), IBM Yorktown Heights (Byrd 1987, 1989)...
- Acquisition of syntactic information (from LDOCE grammar codes)
- e.g. acquisition of 'ergative' verbs from LDOCE (Fontenelle – DSNA 1989, Euralex 1990)
- 'ergative' causative-inchoative alternation (Levin)
- Automatic extraction of lexical-semantic information from dictionary definitions

Extracting taxonomies

- Amsler (1979), Michiels (1982), Boguraev & Briscoe (1989), Wilks et al. (1996)
- Locating genus terms + defining structures (Michiels 82 on LDOCE):



Typical instruments: exploiting definition patterns

LDOCE:

- **accumulator**: a box-like instrument in which electrical power can be stored
- **airgun**: any apparatus that uses strong air pressure to force out a material, esp. paint
- **alarm**: any apparatus, such as a bell, noise, flag, by which a warning is given

LDOCE → MindNet →
Microsoft grammar checker

Retrieving 'ergative' verbs from MRDs/LKBs

- Causative/inchoative alternation (Levin 1993)
- Typical of 'change-of-state' verbs (*grow, increase, improve, darken...*)
 - OPEC increased oil prices.
 - Oil prices increased.
 - John rang the bell.
 - The bell rang.
 - Mary opened the door.
 - The door opened.
- ≠ indefinite object alternation
 - John is eating an apple → John is eating
(but not: * an apple is eating)

Retrieving 'ergative' verbs from a lexical database

- Combination of
 - Definition patterns
 - to (cause to) Verb
 - to (allow to) Verb
 - to (help to) Verb
 - to make or become
 - to bring or come

bake 1 v [T1; I0] to (cause to) cook using dry heat in a special box
shorten v [T1; I0] to make or become short or shorter
fasten 2 v [L9; X9: (to, on)] to make or become firm in (a given state) or joined to (a given thing)

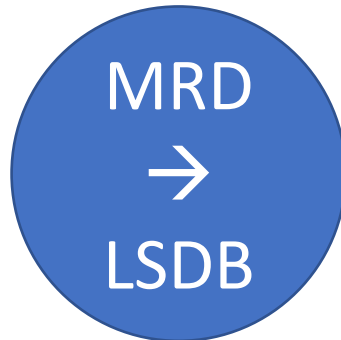
LDOCE (Procter, 1973)

And

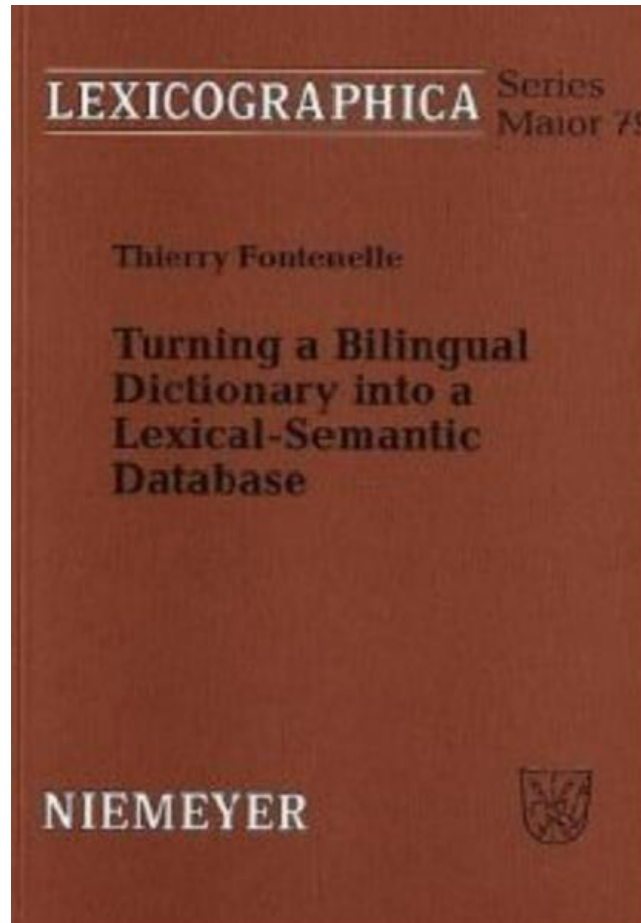
- Grammar codes:
 - T1: transitive verb
 - I0: intransitive verb
- Fontenelle, T. & Vanandroye, J.: 'Retrieving ergative verbs from a lexical database', *Dictionaries: Journal of the Dictionary Society of North America*, Vol.11, 1989, pp.11-39
- Fontenelle, Th. [Automatic extraction of lexical-semantic relations from dictionary definitions](#) (Euralex 1990 Proceedings)

Typesetting tape: Collins-Robert English-French dictionary (Atkins & Duval 1978)

>u1< abandon >u155< [>u11< >u18< b >u43< nd >u11< n] >u2< 1
>u6< vt >u2< (a) >u8< >u6< forsake >u9< >u6< person >u5<
abandonner, quitter, de>u129<laisser. >u8< >u6< fig >u9< >u4< to
>u40< o.s. to >u5< se livrer a>u128< . S'abandonner a>u128< , se
laisser aller a>u128< , >u7< >u3< (b) >u8< >u6< Jur etc >u5< :
>u6< give up >u9< >u6< property, right >u5< renoncer a>u128< ;
>u6< action >u5< se de>u129<sister de, >u7< >u3< (c) >u5< faire
(acte de) de>u129<laissement de, >u7< >u3< 2 >u6< n >u8< U >u9<
>u5< laisser-aller >u6< m >u5< , abandon >u6< m >u5< ,
rela>u132<chement >u6< m. >u4< with (gay) >u40< >u5< avec (une
belle) de>u129<sinvolture >u7< >u1< abandoned >u155< [>u11<
>u18< b >u43< nd >u11< nd] >u6< adj >u2< (a) >u8< forsaken
>u9< >u6< person >u5< abandonne>u129< , de>u129<laisse>u129<
>u6< place >u5< abandonne>u129< . >u2< (b) >u8< dissolute
>u9< de>u129<bauche>u129< , >u7<



Transforming the Collins-Robert dictionary into a lexical-semantic database



- Collins-Robert English-French dictionary (Atkins & Duval 1978)
- Creation of a collocation database
- Exploiting the metalinguistic information of the dictionary
- ‘Augmenting’ the dictionary with lexical-semantic relations
- Theoretical framework: Mel’čuk’s Lexical Functions (Meaning-Text Theory and Explanatory Combinatory Dictionaries)
- *Turning a bilingual dictionary into a lexical-semantic database*, Lexicographica Series Maior 79, Max Niemeyer Verlag, Tübingen, 1997

DECIDE project

- *Designing and Evaluating Extraction Tools for Collocations in Dictionaries and Corpora* - "Multilingual Action Plan" (MLAP/93-19), DG XIII/E, European Commission, Luxembourg, 1994-1996.
- Partners:
 - University of Liège
 - University of Stuttgart
 - Xerox Research Centre Europe
 - Adviser: B.T.S. (Sue) Atkins
- [The DECIDE project: Multilingual Collocation Extraction](#) (Grefenstette, G., Heid, U., Schulze, B.M., Fontenelle, T. & Gérardy, C.), Euralex 96 Proceedings, Göteborg, 1996.

Collocations and selection restrictions in the Collins-Robert dictionary

Collocations and selection restrictions in the Collins-Robert dictionary

- Typical subjects of a verb between square brackets []

Collocations and selection restrictions in the Collins-Robert dictionary

- Typical subjects of a verb between square brackets []
- Nouns typically used as a complement of another noun: between square brackets

Collocations and selection restrictions in the Collins-Robert dictionary

- Typical subjects of a verb between square brackets []
- Nouns typically used as a complement of another noun: between square brackets
- Direct objects of a verb / nouns modified by an adjective: unbracketed, unparenthesised

Collocations and selection restrictions in the Collins-Robert dictionary

- Typical subjects of a verb between square brackets []
- Nouns typically used as a complement of another noun: between square brackets
- Direct objects of a verb / nouns modified by an adjective: unbracketed, unparenthesised
- adjectives/verbs/adverbs modified by an adverb : unbracketed, unparenthesised

Collins-Robert: Typical subjects

- **glance off** *vi* [*bullet etc*] ricocher, dévier; [*arrow, sword*] dévier
- **zing 2** *vi* [*bullet, arrow*] siffler **the bullet ~ed past his ear** la balle lui a sifflé à l'oreille; **the cars ~ed past** les voitures sont passées dans un bruit strident.

Collins-Robert: Typical objects

- **discharge** **1 vt (a)** *ship, cargo* décharger; [*bus etc*] *passengers* débarquer; *liquid* déverser; (*Elec*) décharger.
(b) *employee* renvoyer, congédier; (*Mil*) *soldier* rendre à la vie civile; (*for health reasons*) réformer; (*Jur*) *prisoner* libérer, mettre en liberté, élargir...
(c) *gun* décharger, faire partir; **arrow** décocher
- **shoot** **3 vt (fire)** *gun* tirer or lâcher un coup de (*at sur*); **arrow** décocher, lancer, tirer (*at sur*); *bullet* tirer (*at sur*); *rocket, missile* lancer (*at sur*)
- **barb** **2 vt** **arrow** garnir de barbelures, barbeler; *fish hook* garnir de barbillons.

Collins-Robert: N+N combinations

- **cloud 1 n a** (*Met*) nuage, nuée (*liter*); [*smoke, dust etc*] nuage; [*insects, arrows*] nuée; [*gas*] nappe...
- **quiver n** (*for arrows*) carquois
- **shaft n (a)** (*stem etc*) [*arrow, spear*] hampe; [*tool, golf club*] manche; [*feather*] tuyau; [*column*] fût; [*bone*] diaphyse...
- **sheaf n, pl sheaves** [*corn*] gerbe; [*papers*] liasse; [*arrows*] faisceau.

Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

Lexical functions & Mel'čuk's explanatory combinatorial dictionaries (ECD)

- Lexical function : **$f(x)=y$**

“A lexical function (f) is used together with a keyword to signify a set of either phraseological combinations related to the keyword or those words which can replace the keyword under certain conditions”.

(Steele & Meyer 1990:41)

Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

Lexical functions & Mel'čuk's explanatory combinatorial dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge

Lexical functions & Mel'čuk's explanatory combinatorial dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy

Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride

Lexical functions & Mel'čuk's explanatory combinatorial dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride
- Sing (dust) = fleck, speck, mote
Sing (salt) = grain

Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride
- Sing (dust) = fleck, speck, mote
Sing (salt) = grain
- Liqu (law) = abrogate, repeal
Liqu (file) = delete, erase
Liqu (subscription) = cancel

Lexical functions & Mel'čuk's explanatory combinatorial dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride
- Sing (dust) = fleck, speck, mote
Sing (salt) = grain
- Liqu (law) = abrogate, repeal
Liqu (file) = delete, erase
Liqu (subscription) = cancel
- CausPredMinus (price) = decrease, drop, lower, reduce...

Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride
- Sing (dust) = fleck, speck, mote
Sing (salt) = grain
- Liqu (law) = abrogate, repeal
Liqu (file) = delete, erase
Liqu (subscription) = cancel
- CausPredMinus (price) = decrease, drop, lower, reduce...
- IncepPredPlus (price) = boom, jump, rise, skyrocket, soar

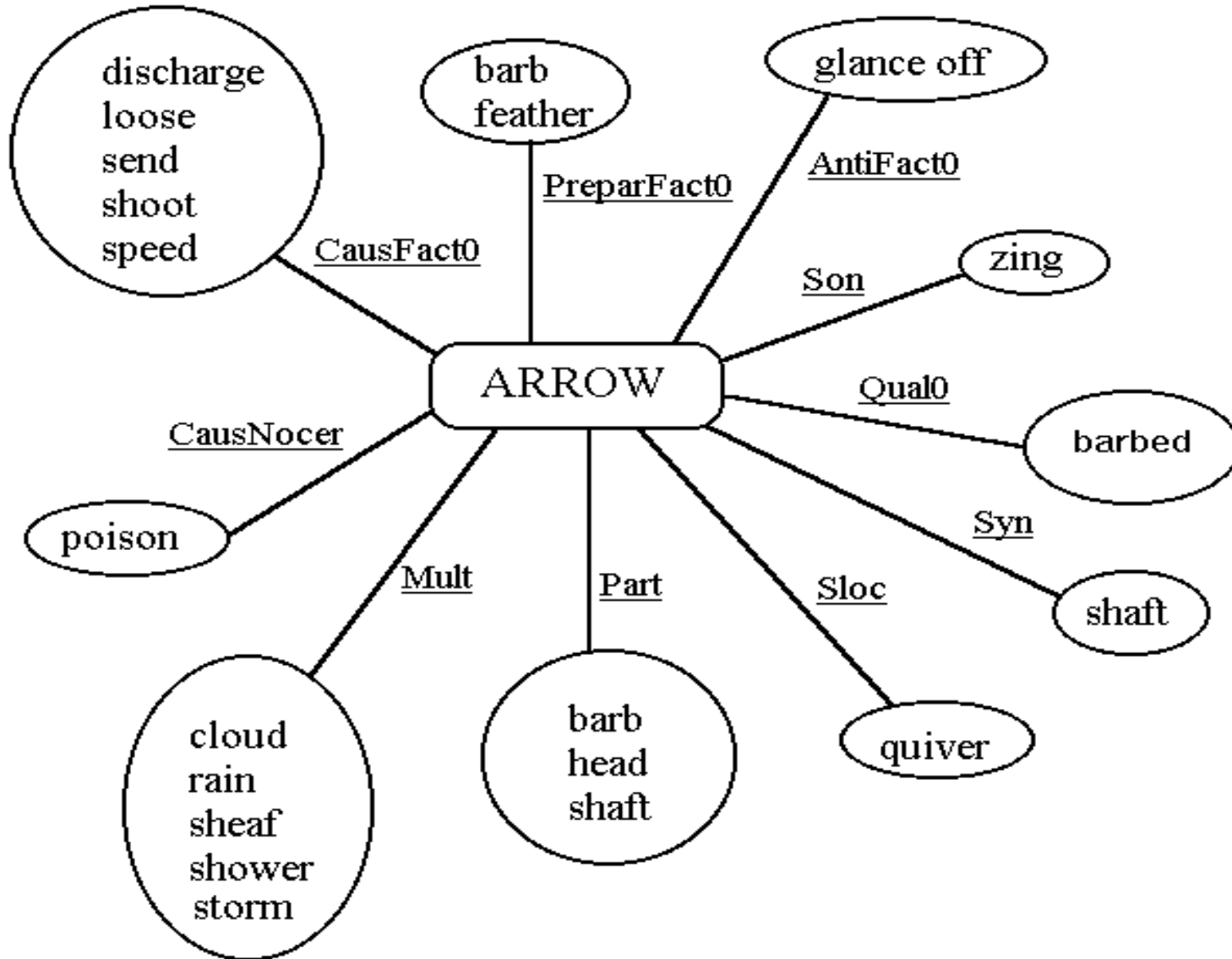
Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride
- Sing (dust) = fleck, speck, mote
Sing (salt) = grain
- Liqu (law) = abrogate, repeal
Liqu (file) = delete, erase
Liqu (subscription) = cancel
- CausPredMinus (price) = decrease, drop, lower, reduce...
- IncepPredPlus (price) = boom, jump, rise, skyrocket, soar
- S_{loc} (fox) = earth, hole, kennel
S_{loc} (bee) = hive
S_{loc} (lion) = den

Lexical functions & Mel'čuk's explanatory combinatory dictionaries (ECD)

- Oper₁ (influence) = exert
Oper₁ (attention) = pay
Oper₁ (suicide) = commit
Oper₁ (complaint) = lodge
- Magn (thirst) = unquenchable
Magn (liar) = arrant
Magn (bachelor) = confirmed
Magn (smoker) = heavy
- Mult (fish) = school, shoal
Mult (dog) = pack
Mult (lion) = pride
- Sing (dust) = fleck, speck, mote
Sing (salt) = grain
- Liqu (law) = abrogate, repeal
Liqu (file) = delete, erase
Liqu (subscription) = cancel
- CausPredMinus (price) = decrease, drop, lower, reduce...
- IncepPredPlus (price) = boom, jump, rise, skyrocket, soar
- S_{loc} (fox) = earth, hole, kennel
S_{loc} (bee) = hive
S_{loc} (lion) = den
- Son (bee) = drone
Son (bell) = ring
Son (elephant) = trumpet

Semantic network



What is the sound made by an arrow?

Son (arrow) = ? → zing

Accessing collocations via lexical functions

Q: Which verbs can be used to mean “delete a law”?

Liqu (law) =?

robcol -i law -lex liqu

Output:

abolish (vt) : law => abroger (loi,liqu)

annul (vt) : law => abroger (loi,liqu)

do away with (vt fus) : law => supprimer (loi,liqu)

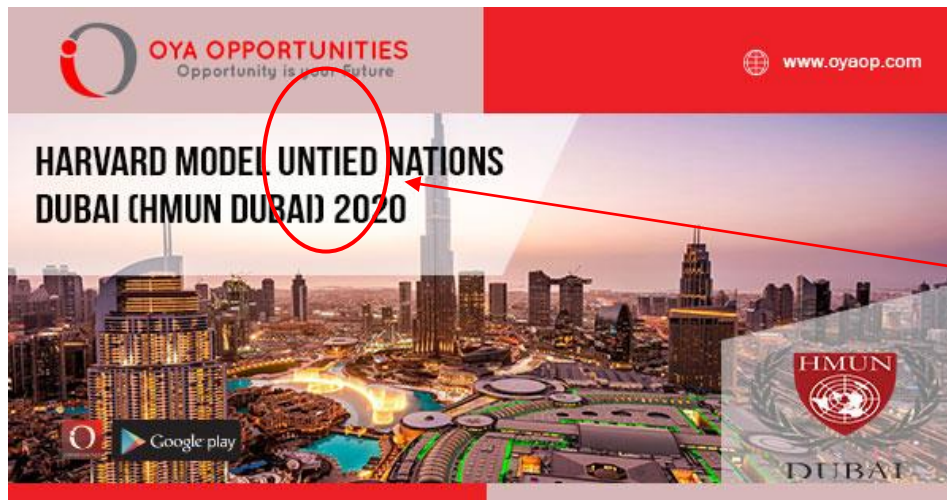
repeal (vt) : law => abroger (loi,liqu)

rescind (vt) : law => abroger (loi,liqu)

revoke (vt) : law => rapporter (loi,liqu)

Developing linguistic technologies for Microsoft Office

- Spell-checkers
- Grammar Checkers
- Context-sensitive spellers (a.k.a. contextual spellers)
- Spelling reforms
- Thesauri



Why the world needs spell-checkers 😊

Grammar checkers

Subject agreement

He think I should work on it tomorrow.

The chair welcome all the participants.

Determiners (a/an)

I work for an European institution.

A hourly wage of €20 seems most reasonable.

Comparatives

This book is more cheaper than that one.

My brother is more tall than me.

Lexicon-driven syntactic parser

The screenshot shows a grammar checker interface. On the left, a context menu is open over the word "taller", listing options such as "Grammar", "Cut", "Copy", "Paste Options", "Search 'tall'", "Read Aloud", "Translate", "Link", and "New Comment". On the right, a correction suggestion box is displayed with the title "Double-check the way this adjective makes a comparison". The word "taller" is highlighted in a grey box, and the suggestion list includes "Ignore Once", "Don't check for this issue", "Options for 'Grammar'", and "See More".

Context-sensitive spell-checker

This car is to expensive for me.

President Biden addresses the Untied Nations General Assembly.

You've got nothing to loose.

People say your hole life flashes before your eyes when you die.

Life insurance plays a very important part in our every day life.

How do your two roles compliment each other?

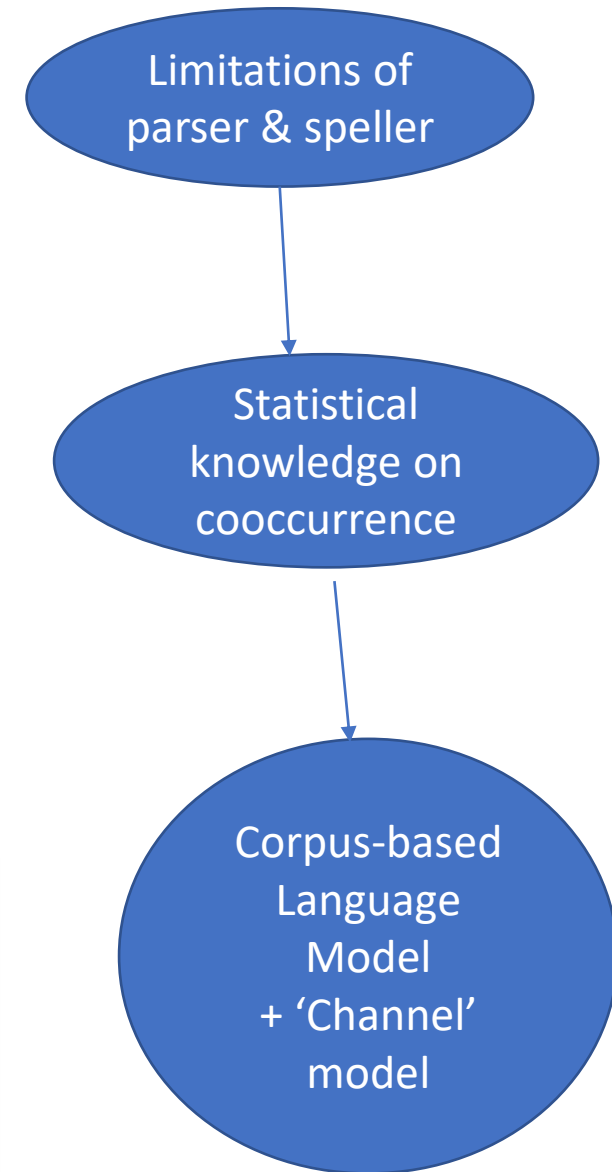
She bought a nice pear of shoes.

There is an international organisation who's role is to keep peace and stability on the continent.

Grammar
Some words are similar but are used differently

whose ...

⊘ ...



Building lexicons for proofing tools

- Grammar checkers
 - Lexicons for parsers (subcategorization information, semantic information, morphological information...)
- Spell-checkers: full-form lexicons (challenge for morphologically-complex languages
 - → creating “conjugators” (Fontenelle, Cipollone, Daniels & Johnson’s Euralex 2008 paper on [Lexicon Creator: A tool for building lexicons for proofing tools and search technologies.](#))
 - Spelling reforms (FR, NL...)
- Context-sensitive spellers (large language models, n-grams)

Building lists of “restricted” (taboo) words

- Application-specific lists (spell-checker ≠ grammar checker ≠ ‘did you mean’ speller in search engines)
- Blocking suggestions (spell-checkers...)
- Censorship in specific markets

- Challenge: defining taboo words, obscenities, profanities (beyond the “Seven ‘dirty’ words you can’t say on television” – Pinker 2008)

From lexicography to translation & terminology

Information isn't information if you cannot understand it.

Stella Hodkin Paris (Translators without Borders, September 2003)

- Managing teams of translators
- Multilingual communication
- Harnessing the power of language technologies (CAT tools, MT, term bases...)



Translation Centre for the
Bodies of the EU



- *Terminology is the DNA of Knowledge*

Kara Warburton

(Head of the Terminology Management @ IBM)

- *There is no Knowledge without Terminology*

(TermNet.org)

Chunking: pre-assembled multi-word units

- Words and terms = prefabricated building blocks of language (chunks of language)
- Writers/speakers store, retrieve, process language very largely in chunks (ready-made sequences)
- Terminology vs. phraseology

Chunking: pre-assembled multi-word units

- Words and terms = prefabricated building blocks of language (chunks of language)
- Writers/speakers store, retrieve, process language very largely in chunks (ready-made sequences)
- Terminology vs. phraseology
 - Bovine spongiform

Chunking: pre-assembled multi-word units

- Words and terms = prefabricated building blocks of language (chunks of language)
- Writers/speakers store, retrieve, process language very largely in chunks (ready-made sequences)
- Terminology vs. phraseology
 - Bovine spongiform
 encephalopathy
 - Mad cow...

Chunking: pre-assembled multi-word units

- Words and terms = prefabricated building blocks of language (chunks of language)
- Writers/speakers store, retrieve, process language very largely in chunks (ready-made sequences)
- Terminology vs. phraseology
 - Bovine spongiform
 encephalopathy
 - Mad cow...
 disease
 - Maladie de la vache...

Language is a dynamic and living organism

- New terms are created every day (e.g. *3D printer; 3D scanner; selfie, VR headsets, drone, LLM, NFT, metaverse, cryptocurrency, blockchain...*)
- Some terms fall into oblivion or become deprecated (remember *floppy disk*?)
- Experts often have terminological preferences
 - Geopolitical factors
 - *Mexican flu vs. A(H1N1) virus*
 - *Trump's 'Chinese' virus vs. Covid-19*
 - Types of audience / publications
 - Scientists vs general public
 - Academic report vs. newspaper article
 - *BSE/Bovine Spongiform Encephalopathy vs. Mad Cow Disease*
 - *Coronavirus vs. COVID-19*

Is lexicography/terminology still relevant?

Do users still need dictionaries?

- We have Google
- We have corpora
- ...

Do translators still need specialised dictionaries / term bases?

- They have Google
- They have corpora
- They have translation memories
- They have MT
- ...

Translation Memories vs. Terminology

TRANSLATION MEMORIES	TERMINOLOGY
How something <u>was</u> translated	How something <u>should be</u> translated
From the past	For the future
Variety	Consistency
Segment level (phrase)	Subsegment level (term)

Eurodicautom – native language query

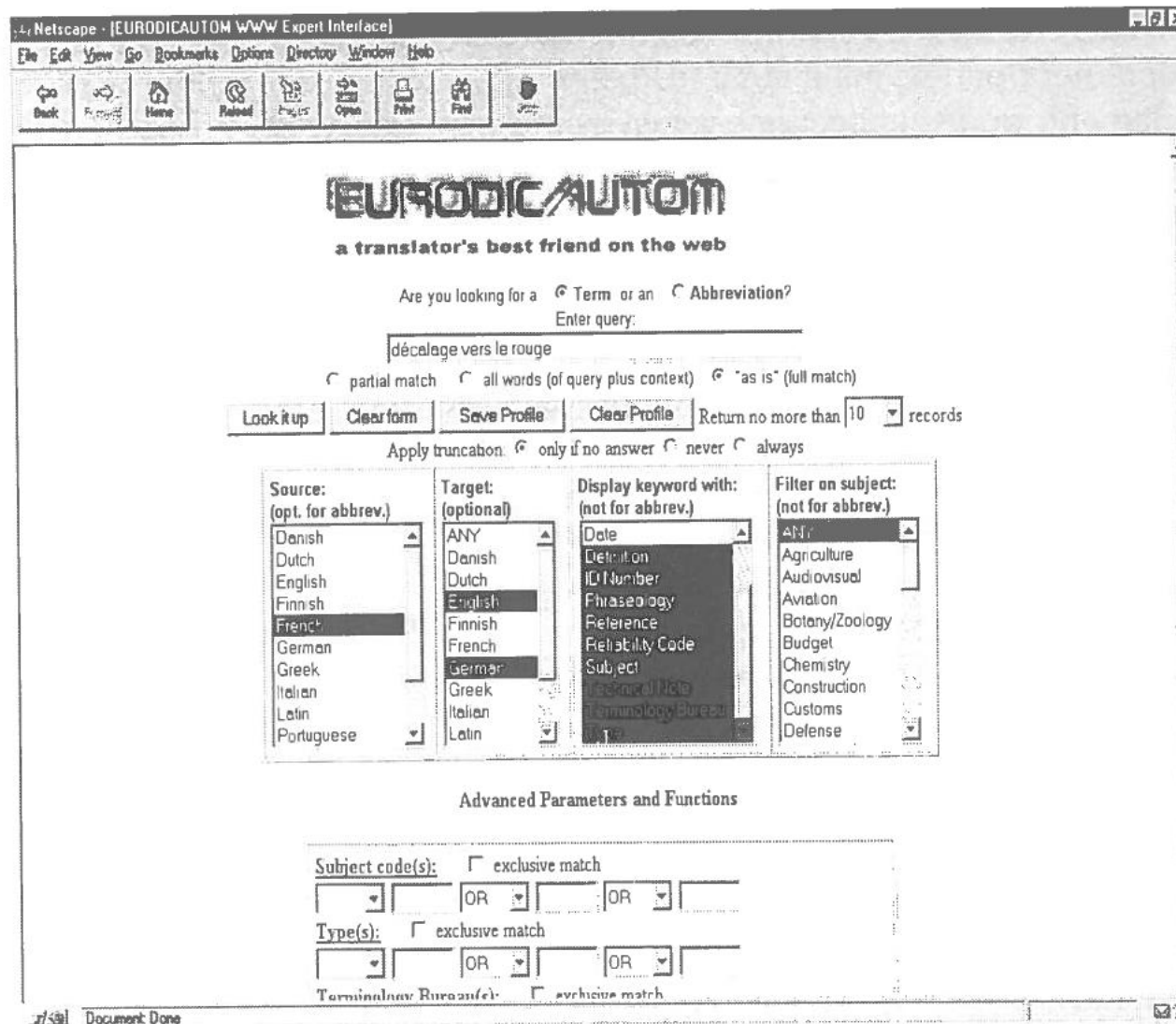
(European Commission terminology database)

(1980s – early 1990s)

```
q décalage vers le rouge@@; L FR EN DE; CF,A,9; TN; MC=N; CM=Y;  
NT=Y; PS=N; AB=N; AU=N; NI=Y; BE=Y; PH=Y; RF=Y; CF=Y;  
SHOWCF=Y; DF=Y; VE=Y; TY=Y; DATE=N; C=10
```

Pour le non initié, cette requête pourrait être paraphrasée comme suit: affiche les fiches terminologiques contenant le terme français (FR) 'décalage vers le rouge', donne les équivalents anglais (EN) et allemand (DE). Considère toutes les sources (CF, A, 9), c'est-à-dire EURODICAUTOM à proprement parler, mais aussi les bases TIS, du Conseil, et EUTERPE, du Parlement, ou les collections générées par les traducteurs. N'applique pas de troncature (TN) et considère exclusivement les termes dans lesquels la chaîne de la question est reprise textuellement et entièrement, à l'exclusion de tout autre matériau supplémentaire (@@)². Affiche uniquement les champs suivants (Y): code matière (CM), note (NT), bureau émetteur (BE), collection (TY), numéro d'identification (NI), phraséologie (PH), référence (RF), code de fiabilité (CF), définition (DF), vedette (VE), pour autant que ces éléments existent³; et n'affiche pas (N) le mot clé (MC), l'auteur de la fiche (AU), les restrictions géographiques (PS), l'abréviation (AB) ou la date. Affiche un maximum de 10 fiches (C = 10).

Web interface – Eurodicautom (1998...)



The IATE Project

The IATE Project

- IATE = “InterActive Terminology for Europe”

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)
 - around 7 million terms

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)
 - around 7 million terms
- **Project Partners:** EC, Council, EP, Court of Auditors, European Economic and Social Committee/Committee of the Regions, Translation Centre for the Bodies of the European Union (CdT), SCIC, Court of Justice, European Investment Bank, European Central Bank (+ Publications Office as an observer)

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)
 - around 7 million terms
- **Project Partners:** EC, Council, EP, Court of Auditors, European Economic and Social Committee/Committee of the Regions, Translation Centre for the Bodies of the European Union (CdT), SCIC, Court of Justice, European Investment Bank, European Central Bank (+ Publications Office as an observer)
- **CdT = Project Initiator & Project Manager** since 2003

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)
 - around 7 million terms
- **Project Partners:** EC, Council, EP, Court of Auditors, European Economic and Social Committee/Committee of the Regions, Translation Centre for the Bodies of the European Union (CdT), SCIC, Court of Justice, European Investment Bank, European Central Bank (+ Publications Office as an observer)
- **CdT = Project Initiator & Project Manager** since 2003
- The development phase started in January 2000.
IATE has been operational since summer 2004.

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)
 - around 7 million terms
- **Project Partners:** EC, Council, EP, Court of Auditors, European Economic and Social Committee/Committee of the Regions, Translation Centre for the Bodies of the European Union (CdT), SCIC, Court of Justice, European Investment Bank, European Central Bank (+ Publications Office as an observer)
- **CdT = Project Initiator & Project Manager** since 2003
- The development phase started in January 2000.
IATE has been operational since summer 2004.
- IATE was **opened to the public** on 28 June 2007: <http://iate.europa.eu>
IATE 2 opened to the public on 12 November 2018.

The IATE Project

- **IATE = “InterActive Terminology for Europe”**
 - 890 000 entries (September 2023)
 - around 7 million terms
- **Project Partners:** EC, Council, EP, Court of Auditors, European Economic and Social Committee/Committee of the Regions, Translation Centre for the Bodies of the European Union (CdT), SCIC, Court of Justice, European Investment Bank, European Central Bank (+ Publications Office as an observer)
- **CdT = Project Initiator & Project Manager** since 2003
- The development phase started in January 2000.
IATE has been operational since summer 2004.
- IATE was **opened to the public** on 28 June 2007: <http://iate.europa.eu>
IATE 2 opened to the public on 12 November 2018.
- Requests from **200+ countries**
25 million queries in 2022 (Public) & **19.5 million** queries (internal)

IATE - <https://iate.europa.eu/home>

iate European Union terminology

en Log on

Last queries Standard view Interpreters' view

venture capital

Reset Search settings

Open expanded search

Source language

bg cs da de el **en** es et fi fr ga hr hu it
lt lv mt nl pl pt ro sk sl sv la mul

Target language

bg cs **da** **de** el en **es** et fi **fr** ga hr hu it
lt lv mt **nl** pl pt ro sk **sl** sv la mul

Searches last week	889 872	Number of terms	7 053 891
Number of entries	703 395	Modifications last week	50 965
New terms last week	3 825	Terms deleted last week	22 150

About IATE

IATE (Interactive Terminology for Europe) is the EU's terminology management system. It has been used in the EU institutions and agencies since summer 2004 for the collection, dissemination and management of EU-specific terminology. The project was launched in 1999 with the aim of providing a web-based infrastructure for all EU terminology resources, enhancing the availability and standardisation of the information.

About IATE FAQ

Download IATE IATE brochure

Discover the IATE search

Standard view Interpreters' view

Search by collection + Download IATE

The added value of term bases: metadata and user preferences

- Metalinguistic labels to capture preferred terminology & community/company- or institution-specific preferences
- Preferred / Obsolete / Deprecated
- Reliability codes
- Definitions / contexts...

★ 3504403	
medical science [SOCIAL QUESTIONS > health]	
en	
novel flu virus	★★★ reliable
novel influenza virus	★★★
novel influenza virus A(H1N1)	★★★
swine influenza virus A(H1N1) deprecated	★★★
Mexican influenza virus deprecated	★★★
influenza A(H1N1)v virus preferred	★★★

TERMINOLOGY CHALLENGES:

TERMINOLOGY CHALLENGES:

- Acquisition and representation

TERMINOLOGY CHALLENGES:

- Acquisition and representation
- Equivalence in EU languages (**24** today)

TERMINOLOGY CHALLENGES:

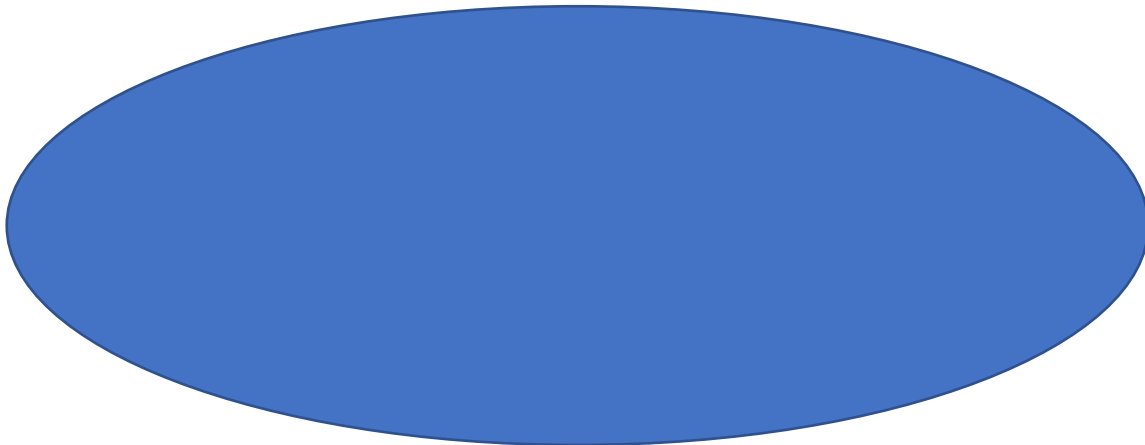
- Acquisition and representation
- Equivalence in EU languages (**24** today)
- Dissemination

TERMINOLOGY CHALLENGES:

- Acquisition and representation
- Equivalence in EU languages (**24** today)
- Dissemination
- Exploitation during translation process (term checkers...)
(Document-specific terminology extraction and recognition)

TERMINOLOGY CHALLENGES:

- Acquisition and representation
- Equivalence in EU languages (**24** today)
- Dissemination
- Exploitation during translation process (term checkers...)
(Document-specific terminology extraction and recognition)



TERMINOLOGY CHALLENGES:

- Acquisition and representation
- Equivalence in EU languages (**24** today)
- Dissemination
- Exploitation during translation process (term checkers...)
(Document-specific terminology extraction and recognition)

Terminological consistency/accuracy
vs.
the “dictatorship” of terminology

TERMINOLOGY CHALLENGES:

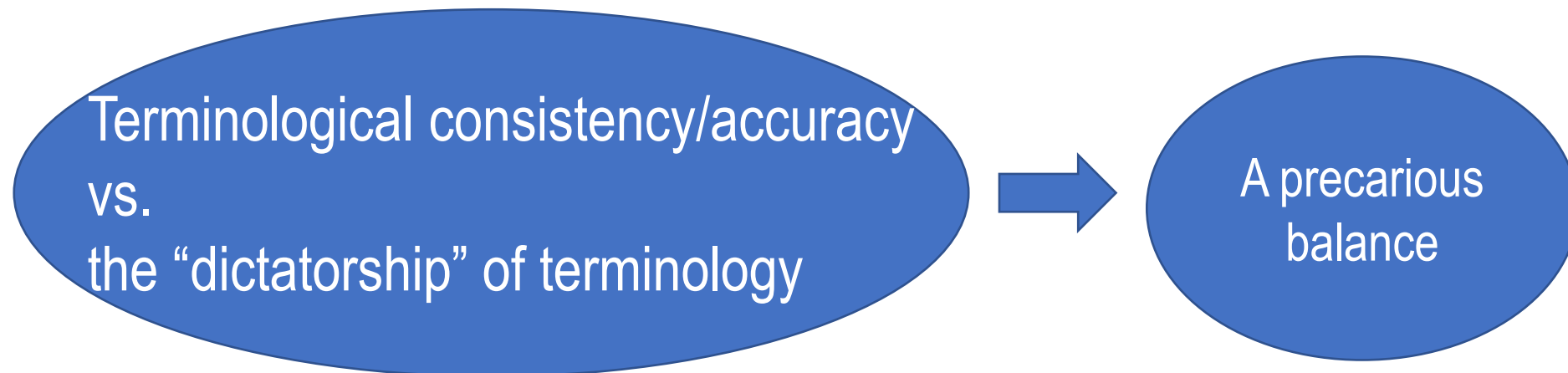
- Acquisition and representation
- Equivalence in EU languages (**24** today)
- Dissemination
- Exploitation during translation process (term checkers...)
(Document-specific terminology extraction and recognition)

Terminological consistency/accuracy
vs.
the “dictatorship” of terminology



TERMINOLOGY CHALLENGES:

- Acquisition and representation
- Equivalence in EU languages (**24** today)
- Dissemination
- Exploitation during translation process (term checkers...)
(Document-specific terminology extraction and recognition)



How much (I/we all) owe to Sue Atkins

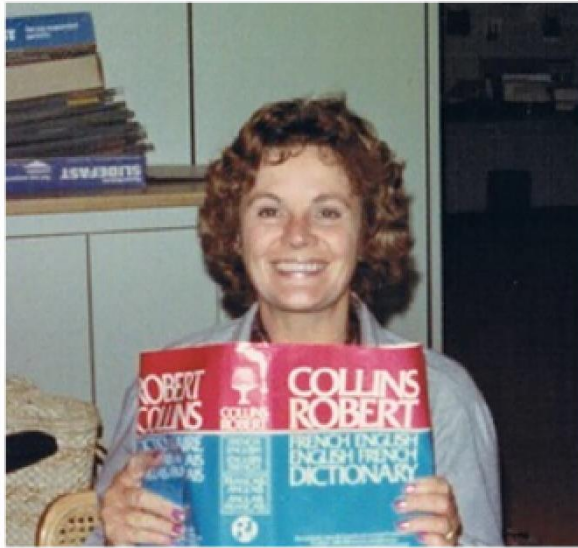
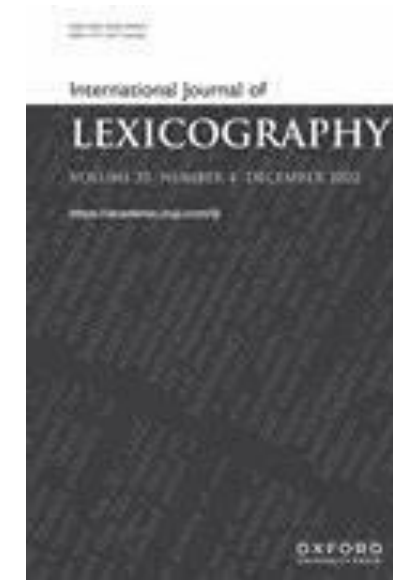
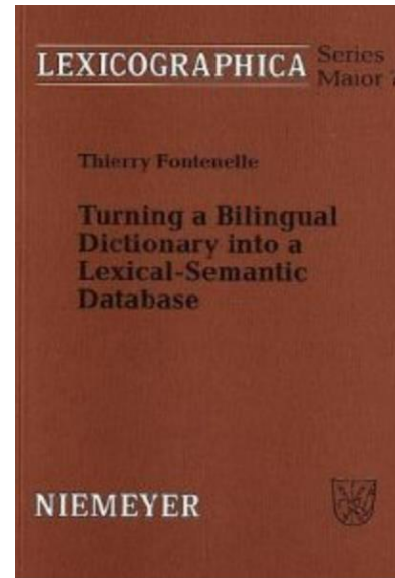


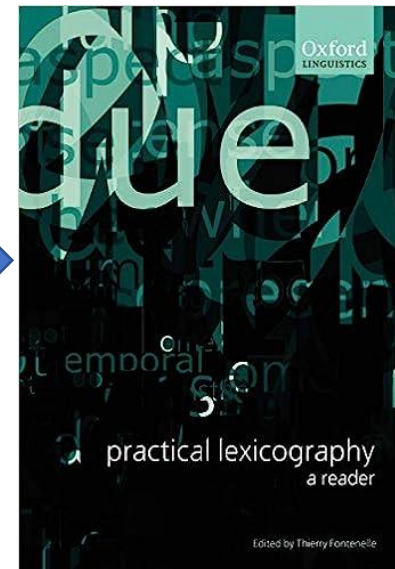
Figure 3. Sue with her first dictionary: publication of the Collins-Robert dictionary, 1978.



IJL Special issue on FrameNet & frame semantics (Vol 16/3, 2003)



Figure 1. Launching the Oxford Guide to Practical Lexicography in Barcelona, at Euralex 2008: Michael Rundell, Sue Atkins, Tony Cowie, and Thierry Fontenelle.



LSA Summer School, Stanford, 1987
Pisa Summer School in computational lexicography, 1988
Euralex 1998 (Liège)
DECIDE project (1994-1996)
IJL Sept. 2003
OUP Reader (2008)
...

Q&A



- Questions?