

Grammar development in human resource linguistics

Munich, July 30, 2012

# Recent functionality for constructing local grammar graphs

Éric Laporte

Université Paris-Est Marne-la-Vallée

Universidade federal do Espírito Santo

# Outline

- Objectives
- Sequences-to-graph
- Visual tools
- Ambiguity tool
- Repetition tool

# Objectives

## **Comfort of graph authors**

Partial automation of graph construction

## **Interoperability with other tools**

Use of graphs in a process chain involving  
systems that do not handle graphs

# Sequences-to-graph

Select a list of sequences from a corpus

Convert it to a graph

Seq2Grf

# Sequences

(...)

*according to announcements*

*according to another newspaper*

*according to Anthony Bolton, manager*

*according to army officials*

*according to author*

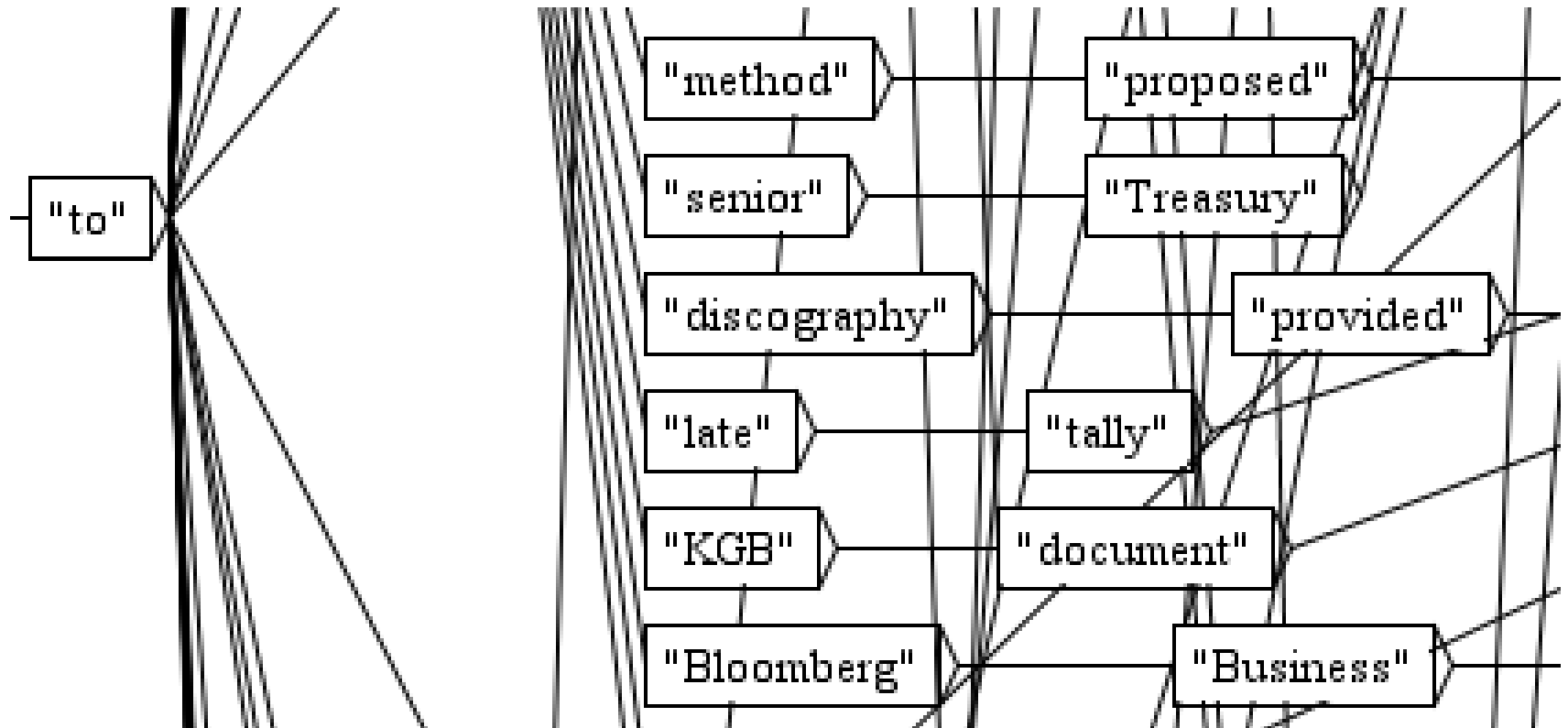
*According to available data*

*according to biographer*

(...)

180 sequences  
including duplicates

# Generated graph



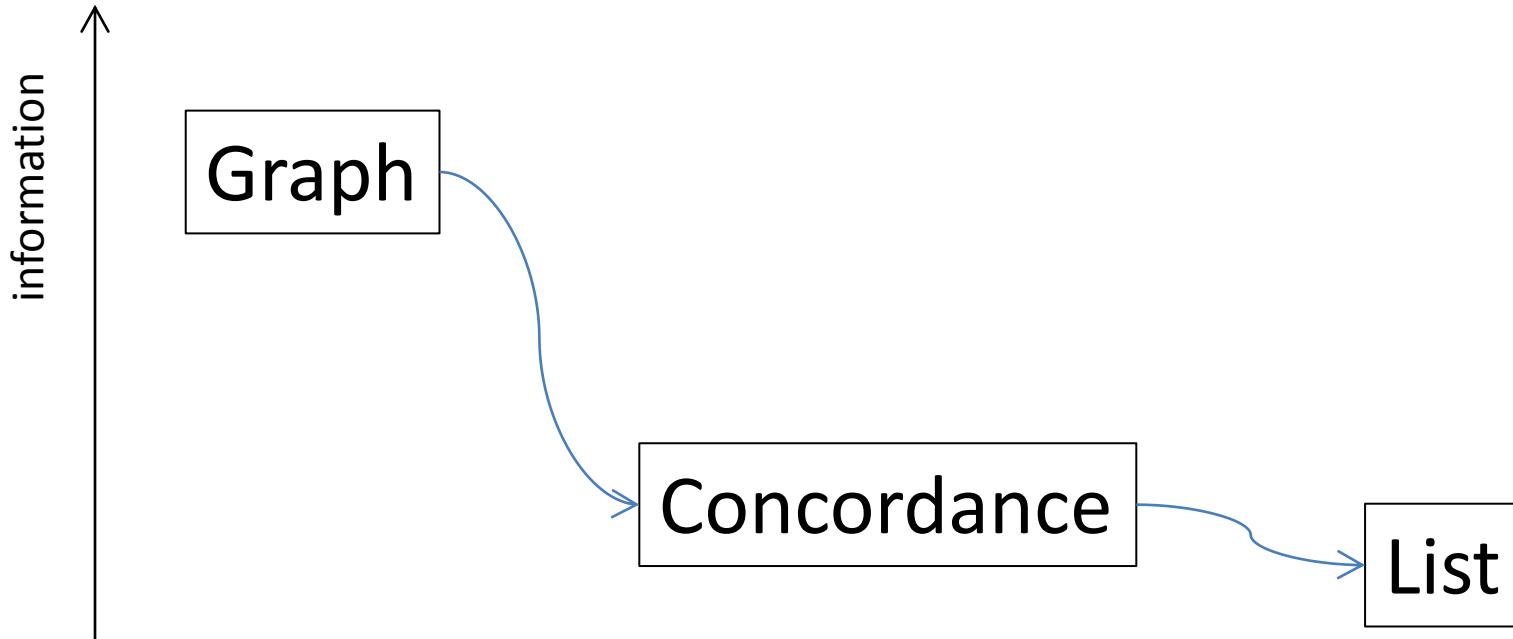
Equivalent graph: 198 states

# Generated graph

o land, according to Anatoli Adamishin, a first deputy for  
nesday, according to announcements by France's state rail  
e city, according to another newspaper, {S} Novaya Gazeta,  
voided, according to Anthony Bolton, manager of Fidelity G  
States, according to army officials and independent analyse  
gorous, according to army officials, because the Environme  
ready, according to author David Landau. {S} Rabbi Moses I  
, where according to biographer Vera Pichel she fended off  
olence, according to British officials, political leaders

Equivalent graph: 149 matches

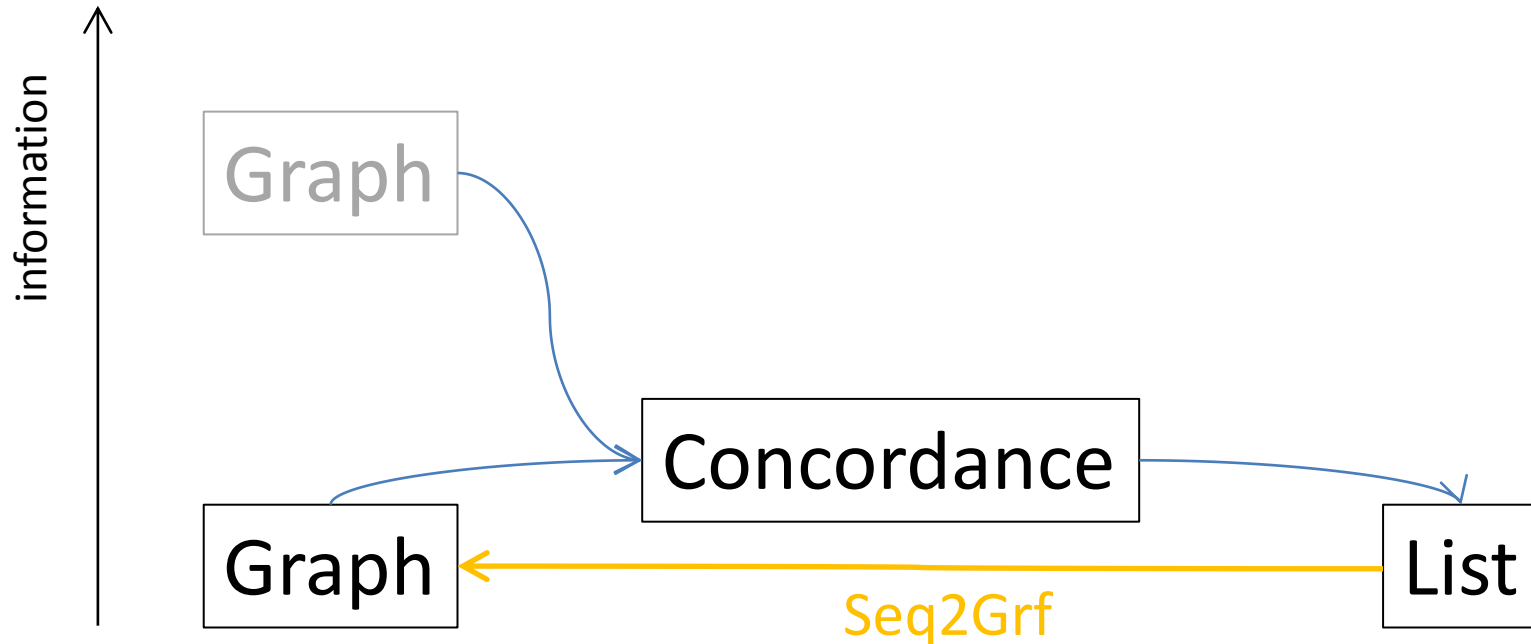
# Sequences-to-graph



Usually, a list automatically generated from a graph contains less information than the graph

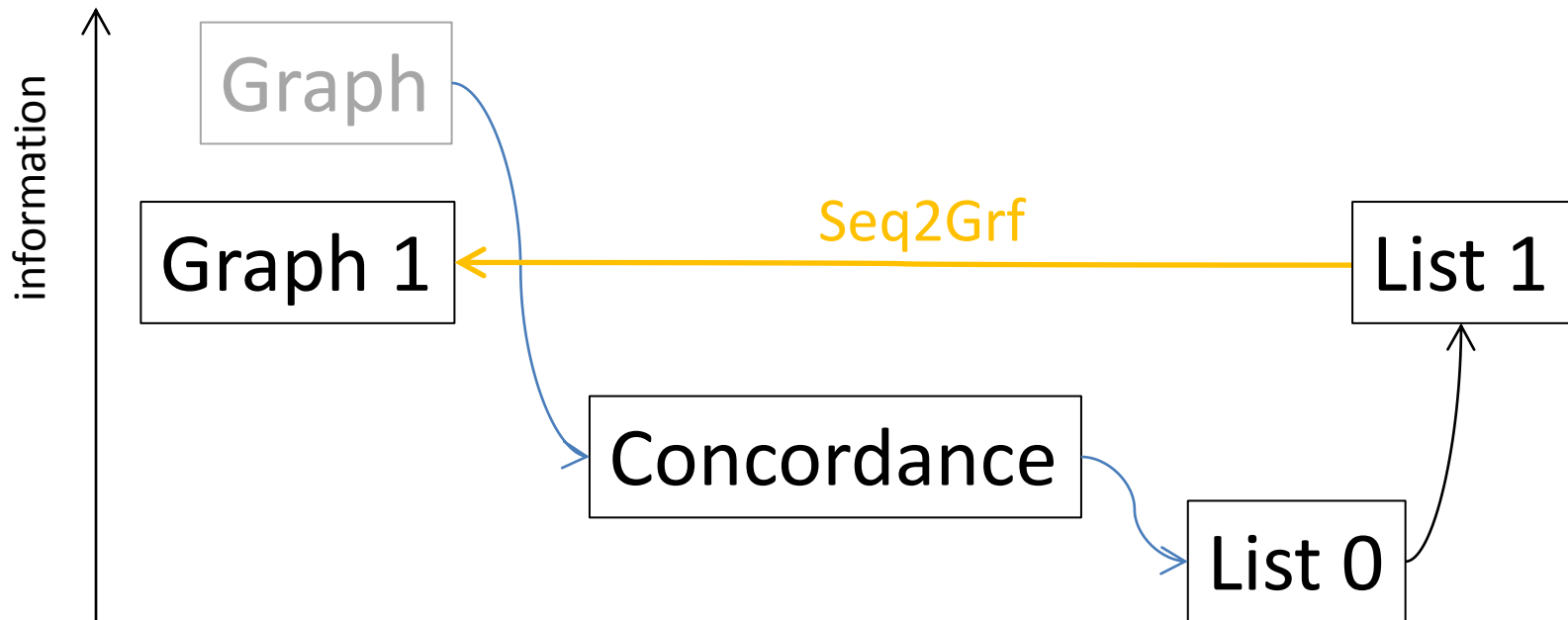


# Sequences-to-graph



A graph automatically generated from a list contains the same information

# Sequences-to-graph



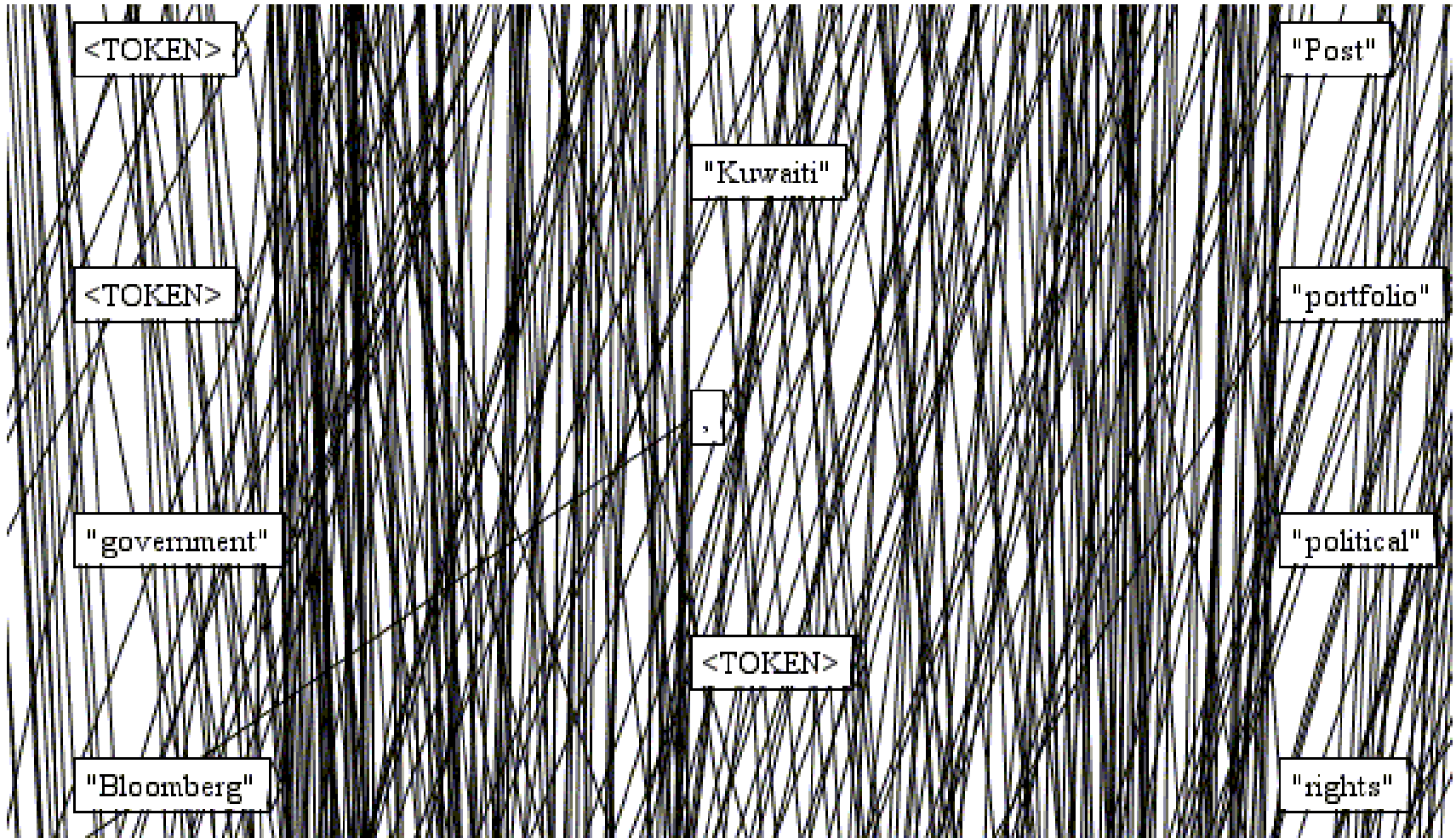
Elaborating information directly on the list,  
manually or with other tools

# Approximate graphs

Generate a graph more tolerant than the list

Extract new sequences from a corpus

# Approximate graph



Graph with 1 subst. or del.: 443 states

# Approximate graph

ced to land, [according to Anatoli Adamishin, a first deputy](#) foreign  
{S} Wednesday, [according to announcements](#) by France's state rail syste  
th the city, [according to another newspaper](#), {S} Novaya Gazeta, which  
l be avoided, [according to Anthony Bolton, manager](#) of Fidelity Global  
ited States, [according to army officials](#) and independent analysts. {S}  
and rigorous, [according to army officials](#), because the Environmental  
e was ready, [according to author](#) David Landau. {S} Rabbi Moses Teitel  
weapons. {S} "[According to available data](#), the development of the fir  
ited States, [according to Beverage](#) Marketing Corp., an industry rese  
apital, where [according to biographer](#) Vera Pichel she fended off a pa  
ket indexes, [according to Birinyi](#) {S} Associates Inc. {S} UAL rose 27/  
o the gospel [according to Bora](#). {S} "In 1990, I wasn't a professional  
e World Cup, [according to Brazilian](#) soccer great Pel?, but that does  
ant violence, [according to British officials, political leaders](#) of bo

1 subst. or del.: 387 matches

# Approximate graphs

Equivalent graph	198 states	149 matches
1 insertion	398 states	150 matches
1 substitution	415 states	339 matches
1 deletion	402 states	339 matches
1 ins. or subst.	608 states	339 matches
1 ins. or del.	408 states	339 matches
1 subst. or del.	443 states	387 matches

# Approximate graphs

## Same number of matches

1 substitution	415 states	339 matches
1 deletion	402 states	339 matches
1 insertion or substitution	608 states	339 matches
1 insertion or deletion	408 states	339 matches

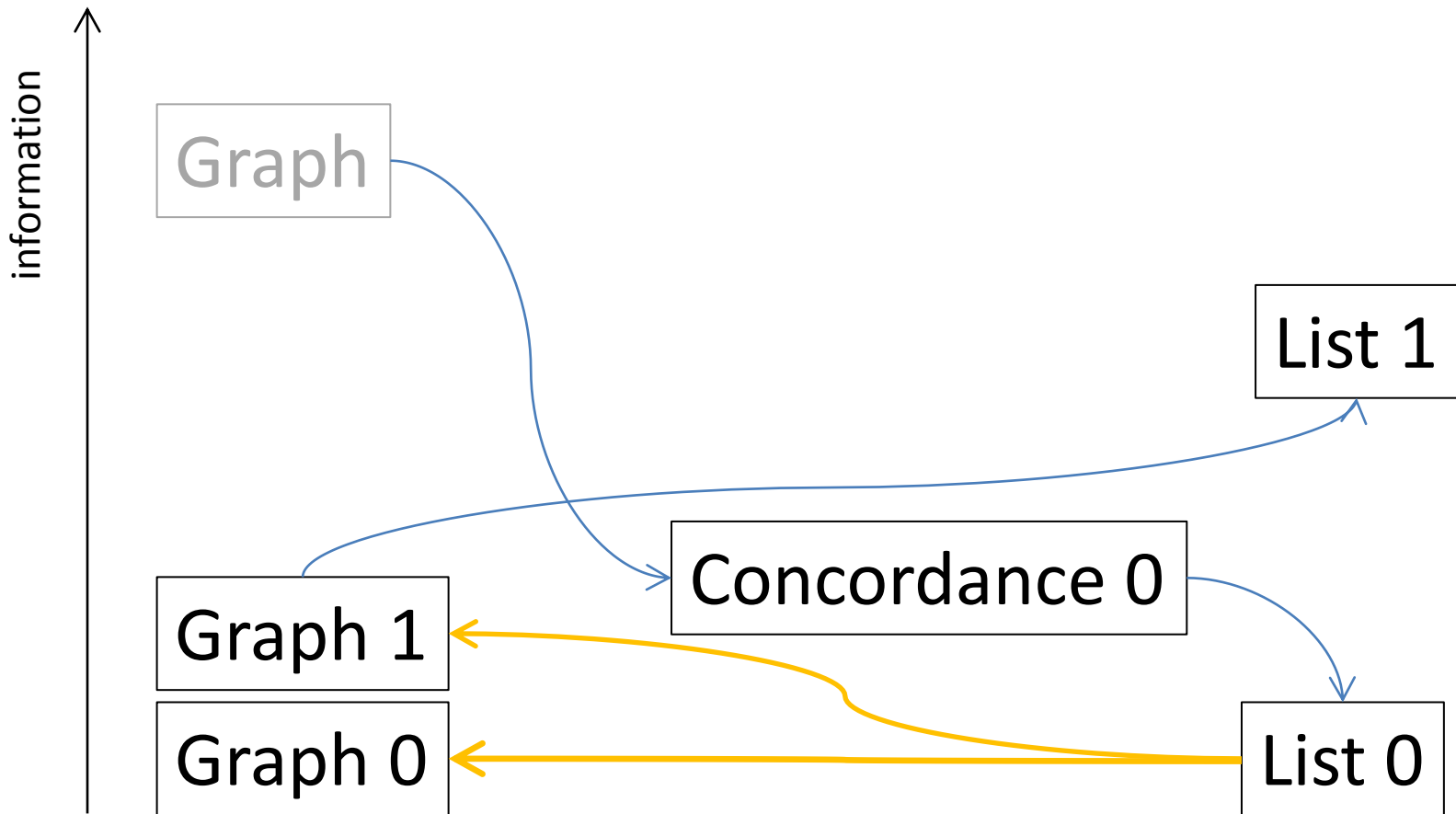
Match with substitution of 1<sup>st</sup> or last word:

*World Cup, according to Brazilian soccer great*

Match with deletion of 1<sup>st</sup> or last word:

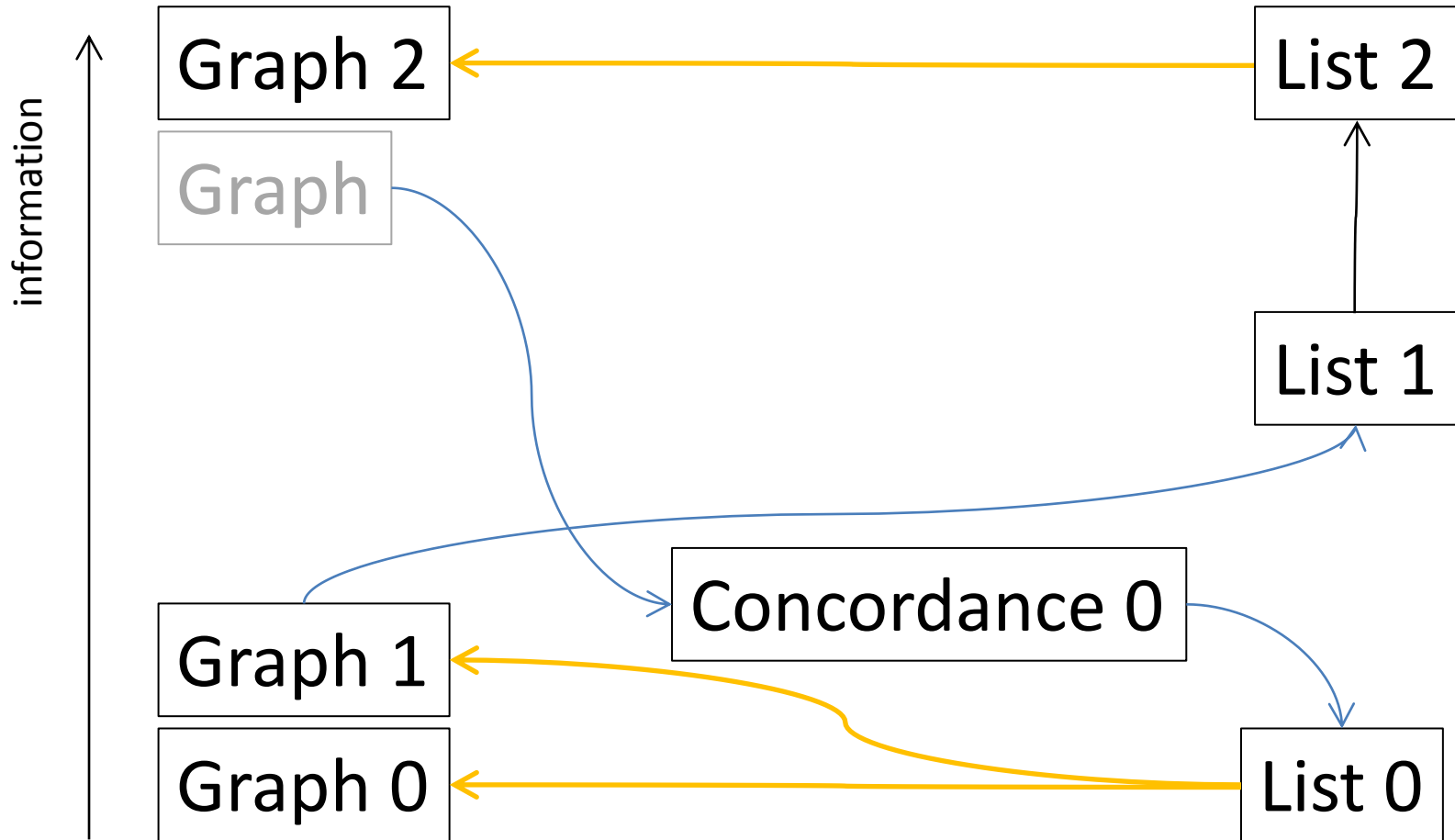
*World Cup, according to Brazilian soccer great*

# Sequences-to-graph

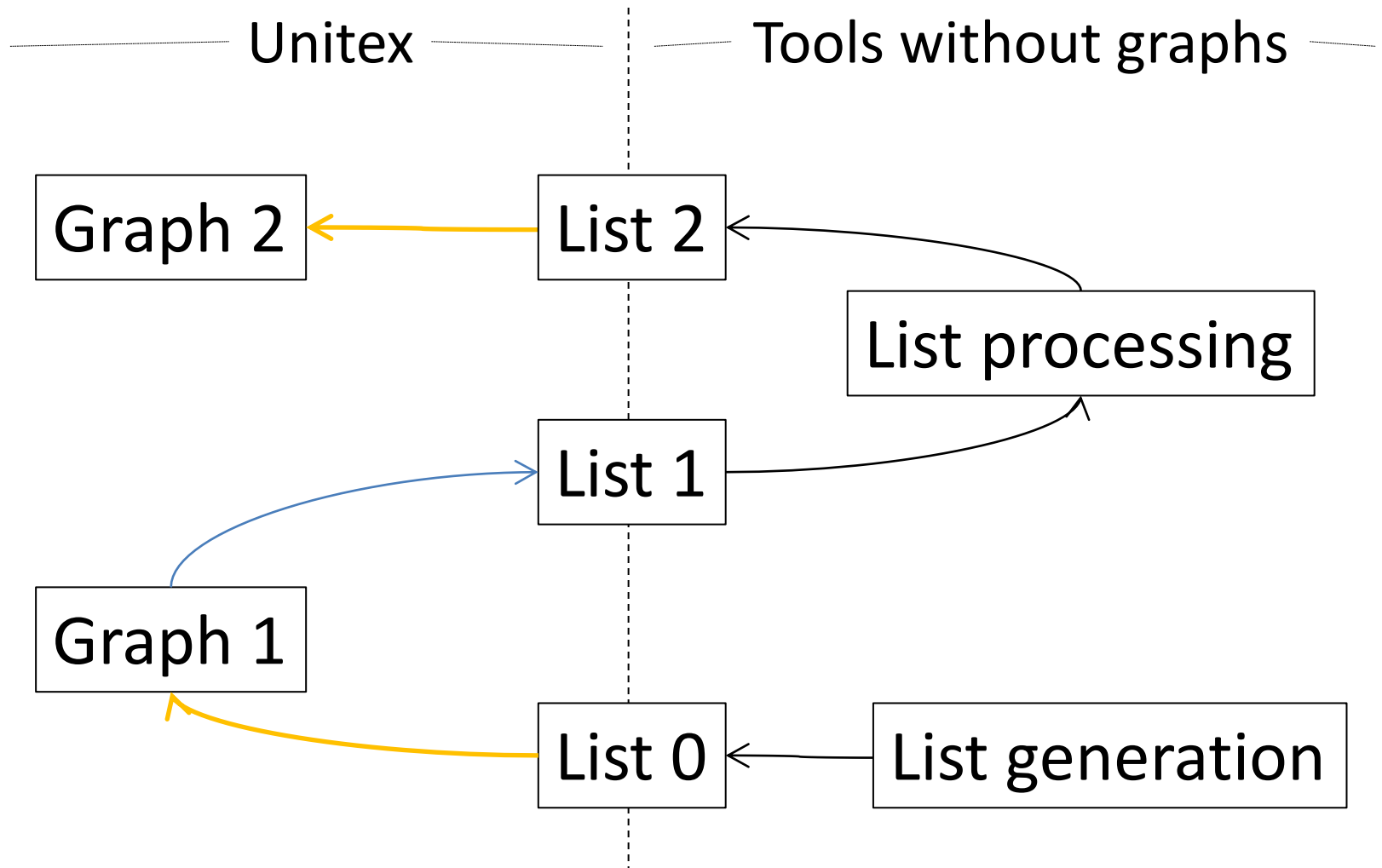




# Sequences-to-graph



# Lists as an interface format

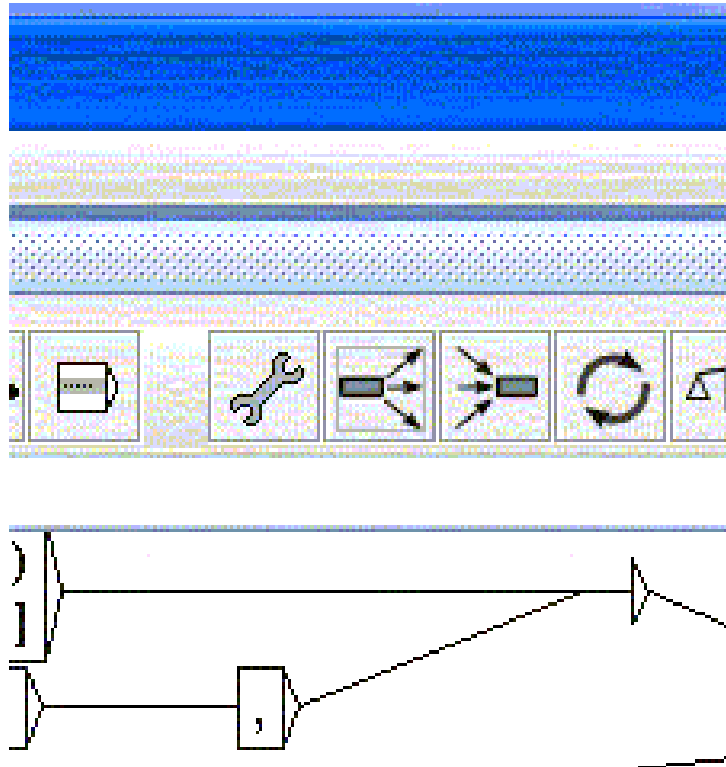


# Visual tools

Visualize the structure of a local grammar

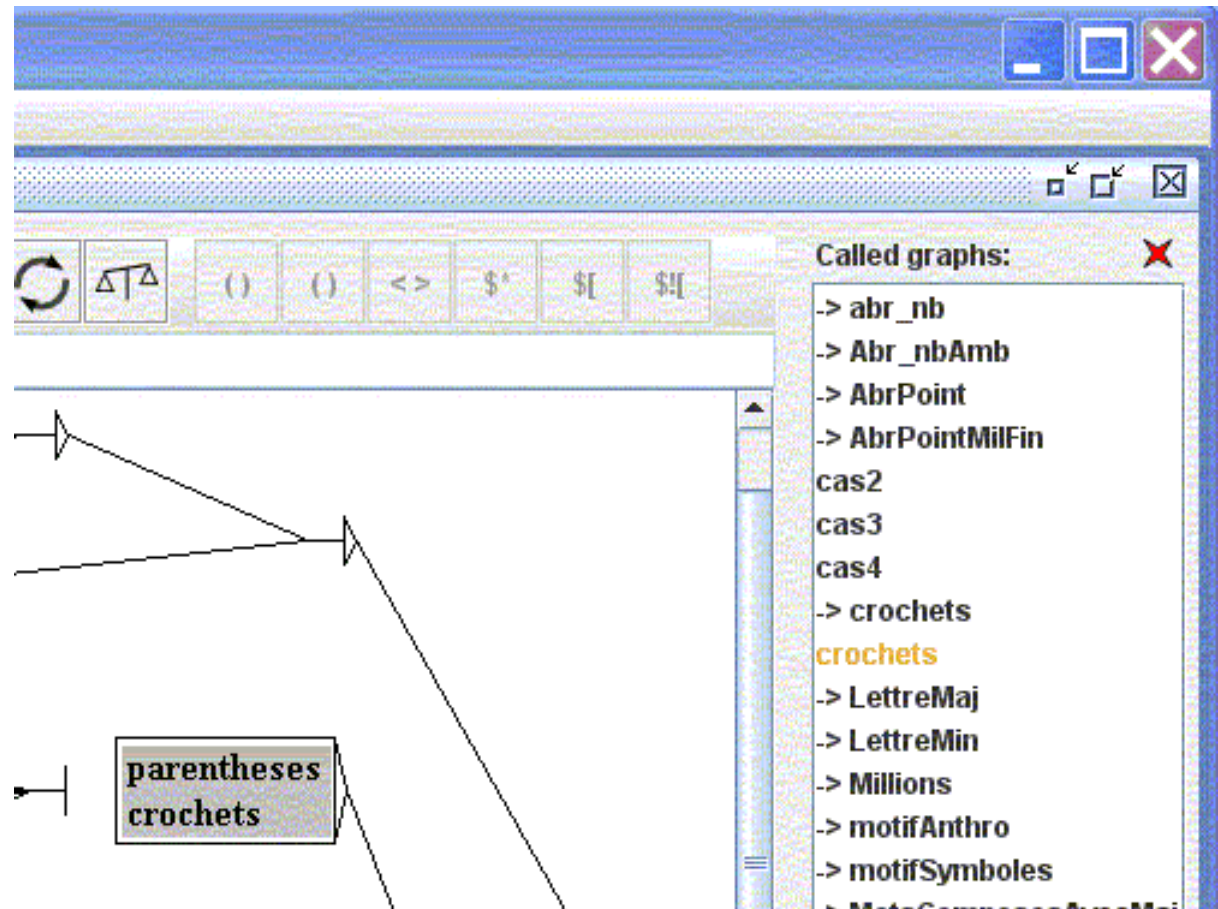
Callees

- direct
- indirect



# Visual tools

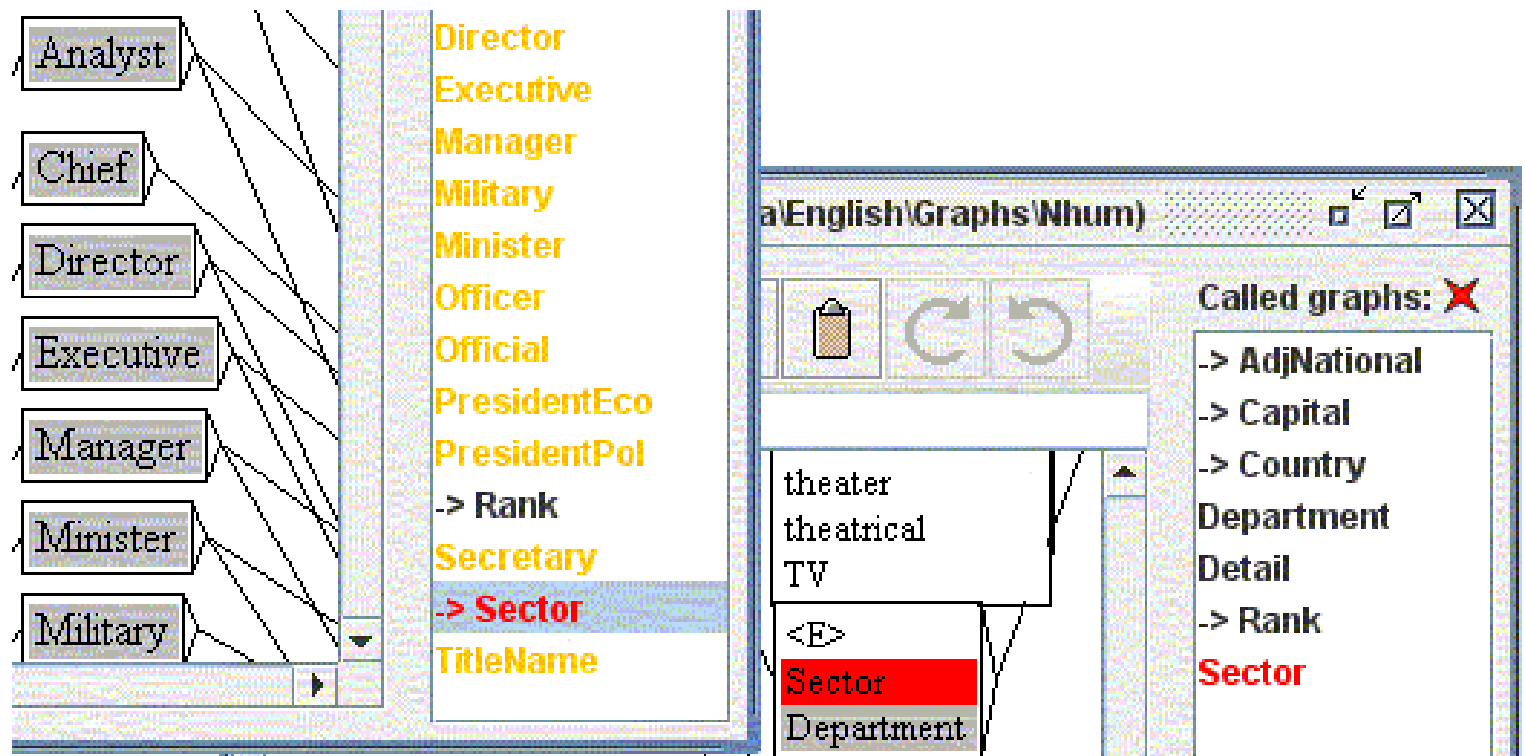
Callees



Ineffective calls: (i) in invalid paths

# Visual tools

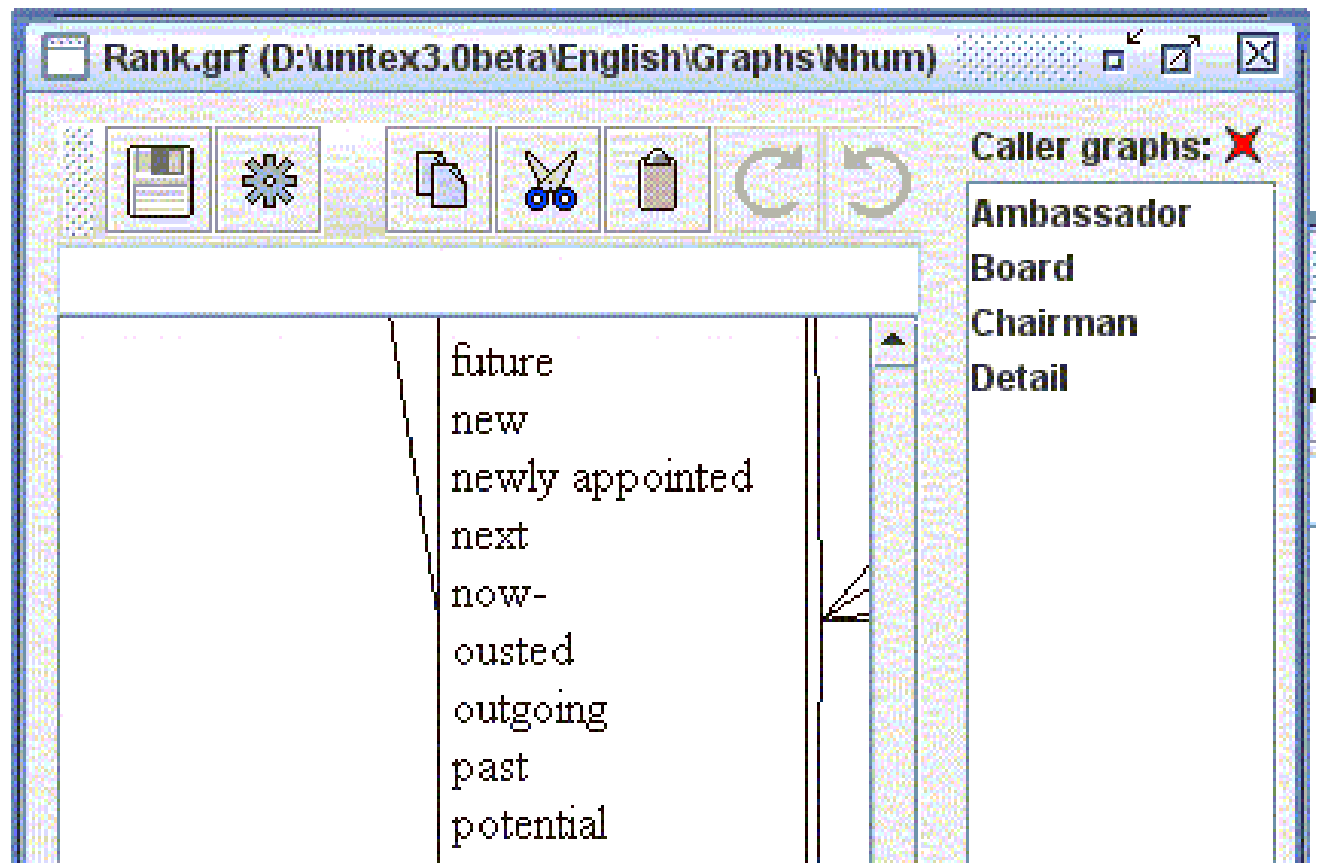
## Callees



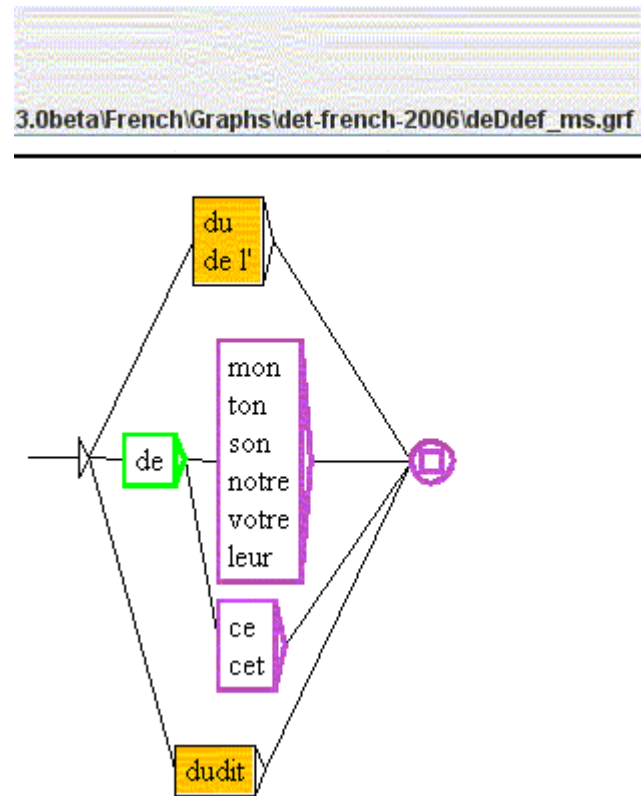
Ineffective calls: (ii) graph not found

# Visual tools

Callers

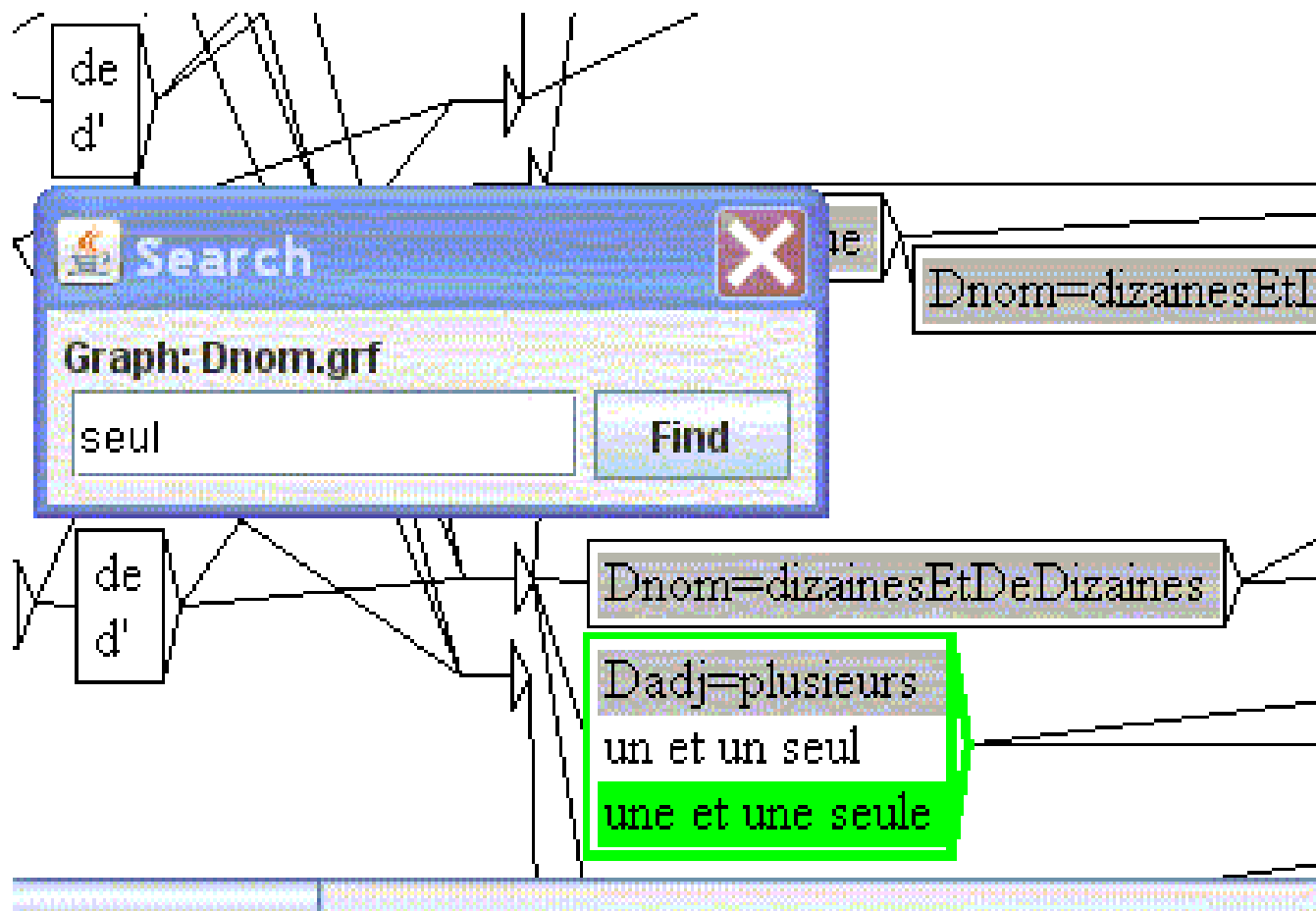


# Graph diff



# Visual tools

Find word in graph: only in Gramlab





# Gramlab

A more professional interface

Slow or experimental features disabled

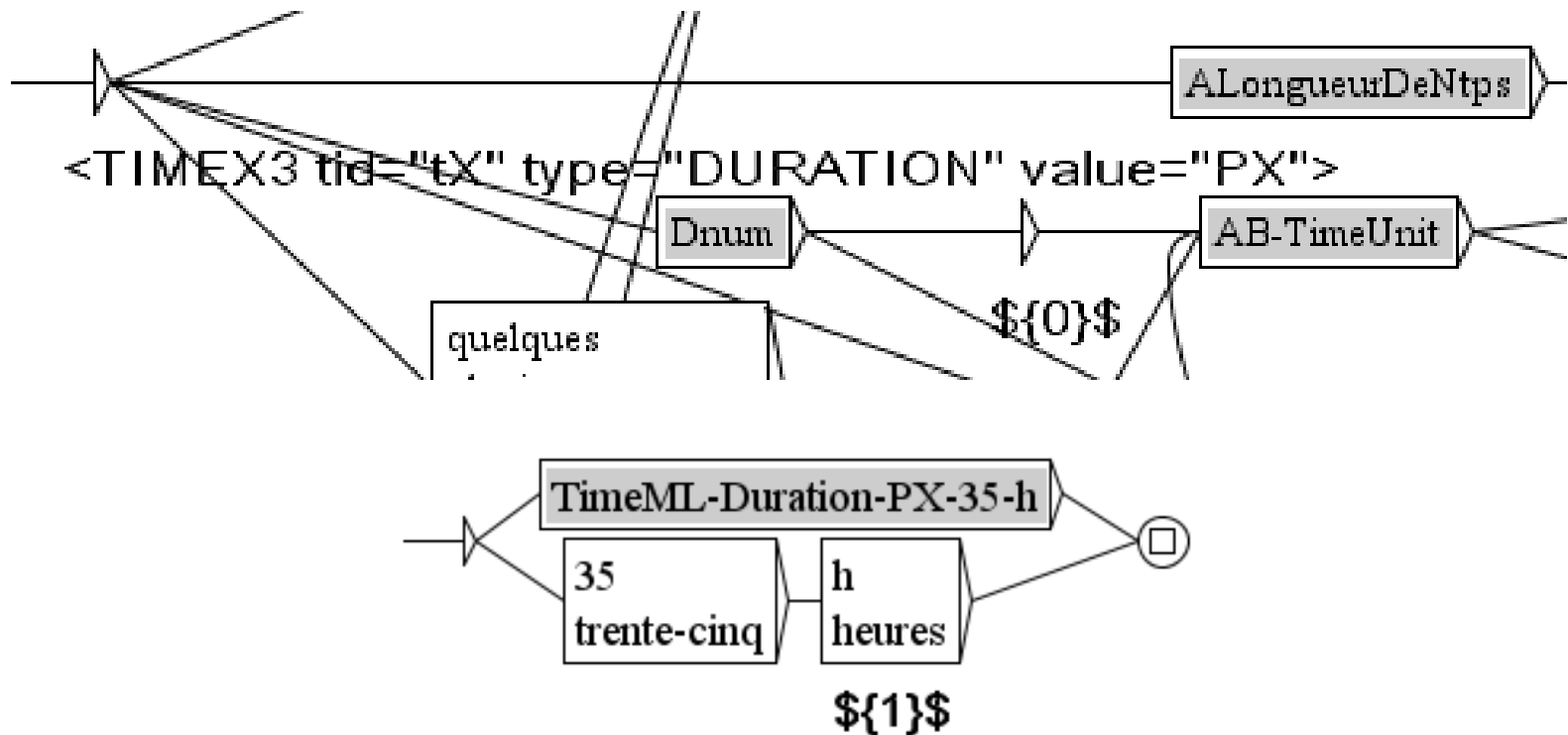
New features for efficiency

A project financed by the European Regional  
Development Fund (ERDF)

# Ambiguity tool

Control precedence between graph paths

Control transducer outputs



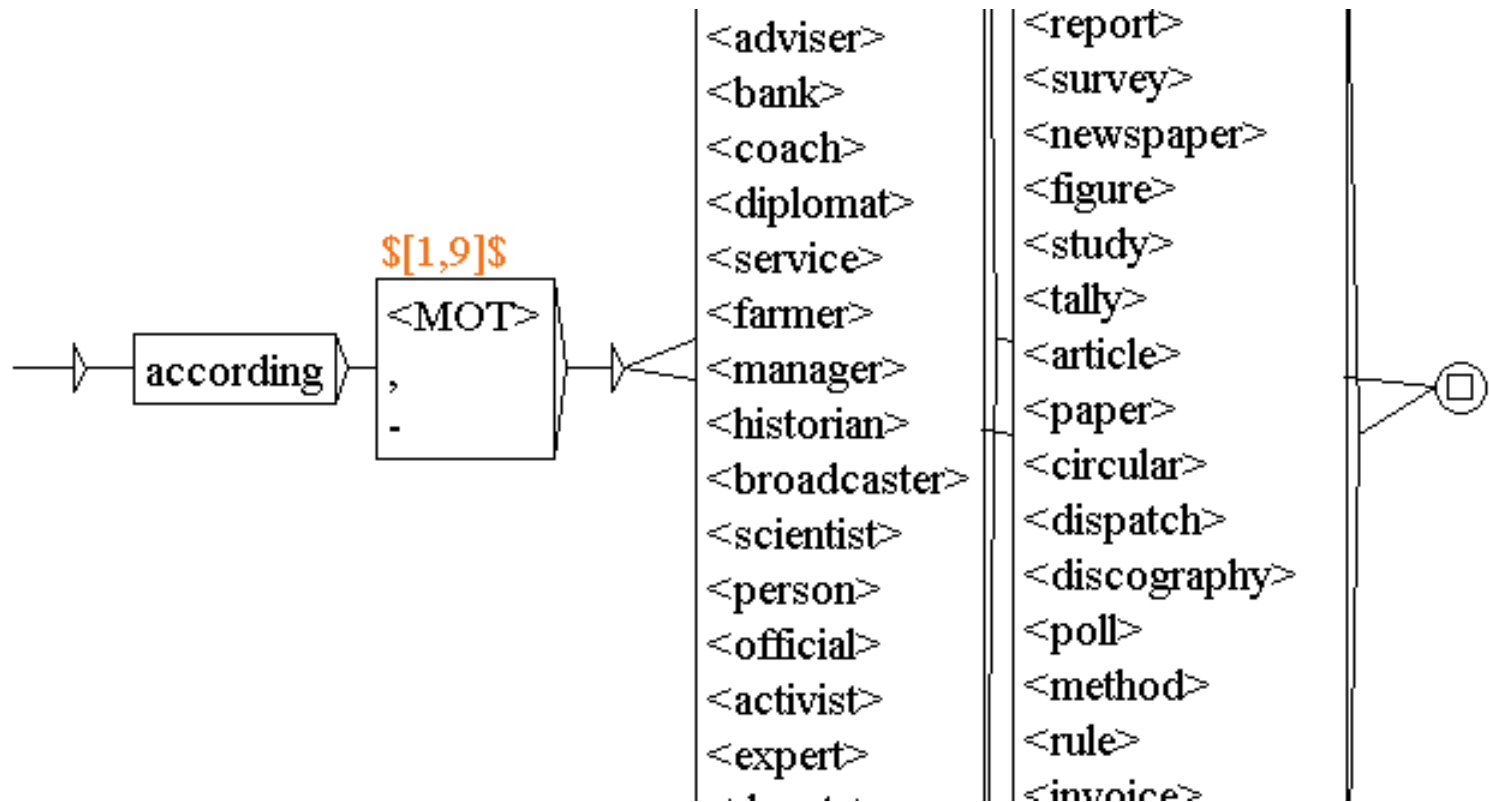
# Ambiguity tool

The path with the highest weight has precedence

vois tous les <TIMEX3 tid="tX" type="DURATION" value="PX"><TIMEUNIT>j  
et pratique d' <TIMEX3 tid="tX" type="DURATION" value="PX"><NUMWORD>un  
proposant des <TIMEX3 tid="tX" type="DURATION" value="PX"><TIMEUNIT>j  
t être suivi d' <TIMEX3 tid="tX" type="DURATION" value="PX"><NUMWORD>un  
laient ce même <TIMEX3 tid="tX" type="DURATION" value="PX"><TIMEUNIT>j  
la semaine de 35 heures (32 pour les services de nuit), une hausse sa  
es services de <TIMEX3 tid="tX" type="DURATION" value="PX"><TIMEUNIT>n  
s dimanches et <TIMEX3 tid="tX" type="DURATION" value="PX"><TIMEUNIT>j  
promotions.{S} <TIMEX3 tid="tX" type="DURATION" value="PX">A longueur  
eprise.{S} Ces <TIMEX3 tid="tX" type="DURATION" value="PX">quelques<TI  
lômés, ils ont <TIMEX3 tid="tX" type="DURATION" mod="LESS\_THAN" value=

# Repetition tool

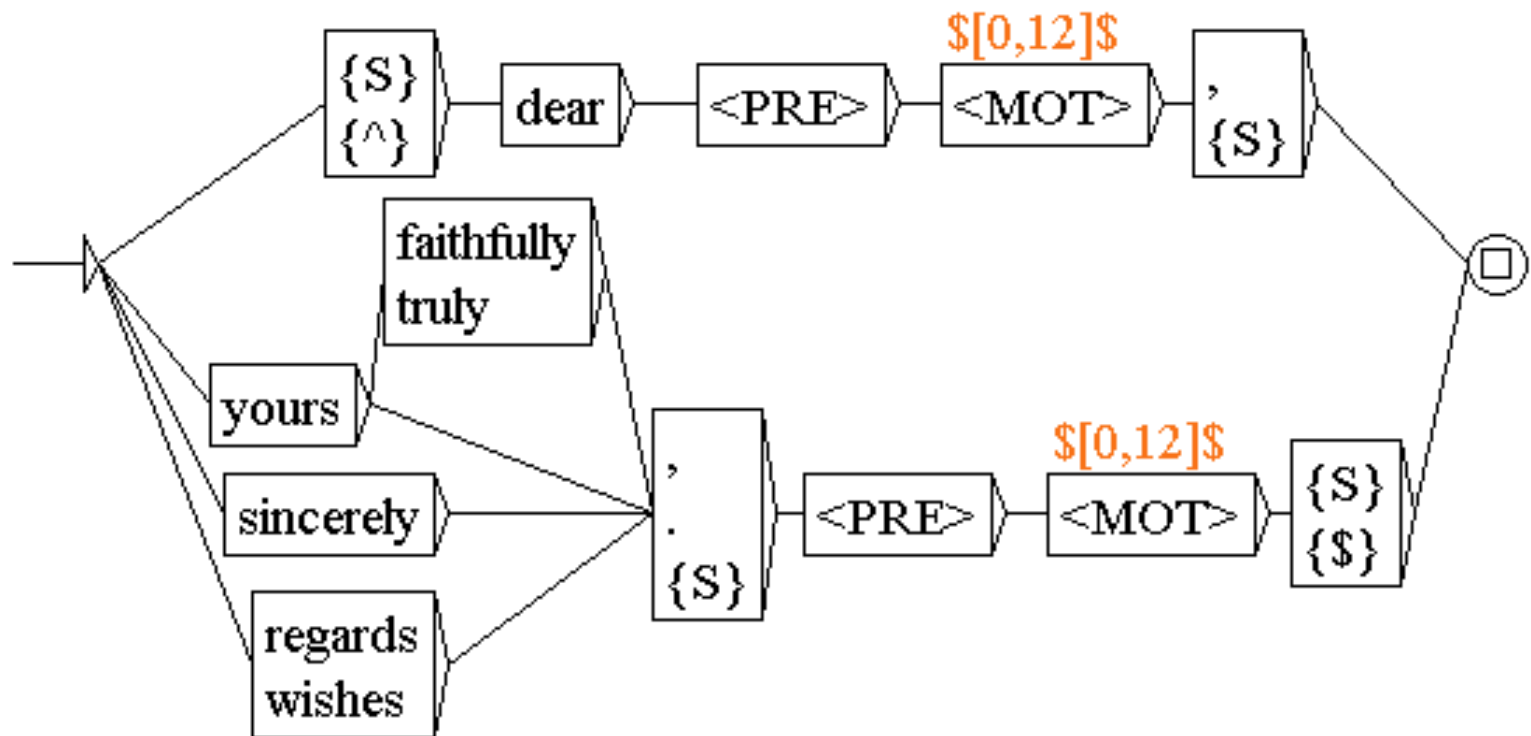
Number of repetitions



# Miscellaneous

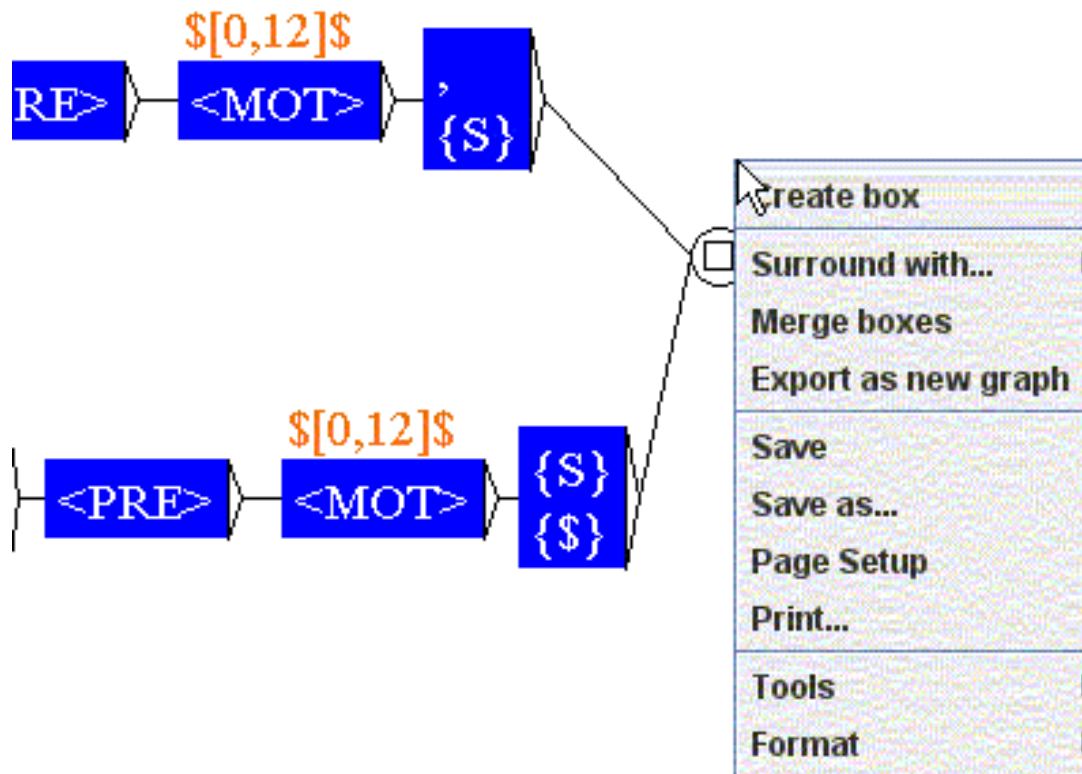
New lexical masks

Recognise beginning and end of text



# Miscellaneous

## Contextual menu of the graph editor



# Conclusion

Practical, technical ease of use

A key to the persistence of Unitex

# Soon

The Unitex-Gramlab Foundation

A project of the Fondation Louvain

Université catholique de Louvain-la-Neuve

Facilitate the communication between users and  
developers

Organize workshops

Tax advantages for donors

Transnational Giving Europe (TGE) network