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
land, according to Anatoli Adamishin, a first deputy for
next day, according to announcements by France's state rail
city, according to another newspaper, Novaya Gazeta,
voided, according to Anthony Bolton, manager of Fidelity G
States, according to army officials and independent analy-
gorous, according to army officials, because the Environ-
ready, according to author David Landau. Rabbi Moses T
, where according to biographer Vera Pichel she fended off
blence, according to British officials, political leaders
Sequences-to-graph

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

According to Anatoli Adamishin, a first deputy foreign
according to announcements by France's state rail system,
according to another newspaper, Novaya Gazeta, which
according to Anthony Bolton, manager of Fidelity Global
according to army officials and independent analysts.
(S) According to available data, the development of the firm
according to Beverage Marketing Corp., an industry rese
according to biographer Vera Pichel she fended off a pa
according to Birinyi Associates Inc. (S) UAL rose 27/
according to Bora. (S) "In 1990, I wasn't a professional
according to Brazilian soccer great Pelé, but that does
according to British officials, political leaders of bo

1 subst. or del.: 387 matches
## Approximate graphs

<table>
<thead>
<tr>
<th>Equivalent graph</th>
<th>States</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 insertion</td>
<td>398 states</td>
<td>150 matches</td>
</tr>
<tr>
<td>1 substitution</td>
<td>415 states</td>
<td>339 matches</td>
</tr>
<tr>
<td>1 deletion</td>
<td>402 states</td>
<td>339 matches</td>
</tr>
<tr>
<td>1 ins. or subst.</td>
<td>608 states</td>
<td>339 matches</td>
</tr>
<tr>
<td>1 ins. or del.</td>
<td>408 states</td>
<td>339 matches</td>
</tr>
<tr>
<td>1 subst. or del.</td>
<td>443 states</td>
<td>387 matches</td>
</tr>
</tbody>
</table>
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 1st or last word:
World Cup, according to Brazilian soccer great

Match with deletion of 1st or last word:
World Cup, according to Brazilian soccer great
Sequences-to-graph

Graph

Graph 0

Graph 1

Concordance 0

List 0

List 1
Lists as an interface format

Unitex

Graph 2
List 2
List processing

Tools without graphs

List 1
List 0
List generation

Graph 1
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
Visual tools

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
Repetition tool

Number of repetitions

$[1,9]$

according

<advert>
<bank>
<coach>
<diplomat>
<service>
<farmer>
<manager>
<historian>
<broadcaster>
<scientist>
<person>
<official>
<activist>
<expert>

<report>
<survey>
<newspaper>
<figure>
<study>
<tally>
<article>
<paper>
<circular>
<dispatch>
<discography>
<poll>
<method>
<rule>
<invoice>
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