Local Grammars 1

Éric Laporte
Outline

Concordance
Local grammar of dates
Invoking a subgraph
Lexical masks
Dictionaries of a text
Concordance of a word

rman Fowler, said Mr. Major's job was safe. {S} "He policy changes."{S} Mr. Major's hold on power may hir RITAIN--PARLIAMENT{S} MAJOR, JOHN{S} CHIDGEY, DAVII lots at SAS and other major companies were not taki th insurance took two major strides forward in the health advocates as a major setback for the tobacc yamira of Burundi and Major General{S} Juv?nal Hab? an War{S} Headline 2: Major Jean-Guy Plante, left,rror show continues," Major General Romeo Dallaire vy fighting along the major highways.{S} General D forces is one of the major problems which the arm?
ypes, Thomas found the commanders a bit disturbed. "And, mon colonel," the officer said, "everything on Omaha Beach, there as a veteran. A Captain Joseph Dawson, introduced Captain Joseph Dawson, introducing the feat of the 101st Airborne commander, the day’s accomplishments. Lieutenant Colonel James E. Rudder and his Colonel James E. Rudder and his General Orwin Clark Talbott, who commander, Lam Horm, and his men were generals, one of 2,000 in the 140 generals, generals selling food,
Concordance: sequences

Text: (S) A senior government official charged Wednesday who was 10 years old said. (S) Key Word: government official (U.S. --ARMED GOVERNMENT OFFICIAL) (S) U.S.--ARMED government official from southern Rwanda said Sunday. (S) Until then government official said. (S) Until then government official. (S) He said Washington government official said. (S) "We want..."
Objective
Explore a linguistic phenomenon
Explore a text

Query language
A major criterion of quality
Allow diversity and sequences at the same time: graphs
Concordance

Local grammar of dates

Invoking a subgraph

Lexical masks

Dictionaries of a text
Dates in English

Hong Kong after the People candidate for the up of drivers for the easily blurred, as the has been as bad as the which organizes the chairmen have set the airlines have put the been mandated by the worked, helping make began a hunger strike July 1, 1997, handover to Chinese
April 1995 presidential poll was made until the July 3 race would be made until the June 6 anniversary showed (even the April 26 crash of the Taiwan-based June 26 regional elections, said the July 4 recess as the deadline to July 4th weekend on sale for travel June 26 election, the Group of Sen June 6, 1944, the biggest D-Day of May 25 after two of them were fir
Query for dates in English
Finite automata and regular expressions

Equivalent regular expression:

$$ha(ha)^*!$$
Imagine
the regular expression for this one

Source: Éric Laporte, 2006
Now for this one

Graphs take more advantage of the two dimensions of a page

Source: Maurice Gross

01/09/2014

Local Grammars 1
Beyond concordances

Such graphs are also used in NLP applications:
indexation
information extraction
annotation
How to edit a graph with Unitex-Gramlab

An open-source corpus processor based on language resources
- generation of concordances
- automatic annotation
- dictionary management
- dictionary-based morphological analysis
- graphical grammar editor

Now 22 languages
Runs on Linux, Windows, OS X
Paumier (2002-2014)
http://igm.univ-mlv.fr/~unitex
How to edit a graph with Unitex-Gramlab

So that you can play around if I'm boring
Menu FSGraph > New

Create a box
Ctrl-click where you want to create the box
Fill in the field above the graph
Validate with Enter

Create a transition
Click once on the source box
Click once on the target box
How to edit a graph with Unitex-Gramlab

Select a box
Click once on it
It becomes blue

Unselect a box
Click on the white background of the graph

Double-clicking on a box
Same as clicking twice
Creates a transition from the box back to it (loop)

Select several boxes
Draw a rectangle around them
You can move, delete, copy, paste them together
How to edit a graph with Unitex-Gramlab

Delete a box
Select it
Press Delete
Validate with Enter

Initial and final boxes
You cannot remove them
You cannot create new ones

Delete a transition
Click once on the source box
Click once on the target box
Outline

Concordance

Local grammar of dates

Invoking a subgraph

Lexical masks

Dictionaries of a text
.1 rising, it is probably less of a problem politically than it was. The new guru was less of a showman and more of a technician who be & "and there is much less of a feeling of impending crisis here th said, "There is less of a sense among juries that it is other peop sun proved to be more of a menace than soccer hooligans at the firs change would be more of a threat.\{S\} But the prevailing belief in \{S\} K mart is more of a runway scraper; its stock is off 40 perc on, he would be more of a provocation to the British, officials sa r individual is still more of an art than a science," the former C
Invoking a subgraph from a node

The adv node invokes the adv.grf graph
Equivalent to substituting the graph for the node

adv.grf:
Invoking a subgraph from a node

Objectives
Use the same graph in various contexts
Distribute a description in several graphs
Invoking a subgraph from a node

A graph can invoke itself
Local grammars

A graph and its subgraphs describing a set of expressions little dependent from the rest of syntax

Local grammar graphs (LGG)

A type of language resources

Used to recognise and process the expressions

Can be included in a production chain
A vintage area of NLP


( [http://books.google.fr/books?id=q4URKd5XKo0C&pg=PA329](http://books.google.fr/books?id=q4URKd5XKo0C&pg=PA329) )


( [http://halshs.archives-ouvertes.fr/docs/00/27/83/19/PDF/BELG.pdf](http://halshs.archives-ouvertes.fr/docs/00/27/83/19/PDF/BELG.pdf) )


An empty node can serve as an intermediate step between two others.
Forms in parallel in a node

A graphical convention
Select the box, type the content of the first line, type "+"
Type the content of the next line, etc.
Validate with Enter
Outline

Concordance

Local grammar of dates

Invoking a subgraph

Lexical masks

Dictionaries of a text
Inflection and derivation

The same meaning conveyed by words with different suffixes

*collect*  *collects*  *collected*  *collecting*  *collector*  *collection*

Increase the recall automatically
Inflection and derivation

Inflection

*collect* *collects* *collected* *collecting*

*<collect>* automatically includes the inflected forms

Derivation

*collection* *collector* *collective* *collectivism*

Automatically including derived forms would be more adventurous
Lexical masks

<collect> recognises automatically any inflected form of collect
<MOT> any simple word
<pre> any simple word with capitalized first letter
<NB> any contiguous sequence of digits
<V> any word marked as V in the lexicon (verbs)
Lexical masks

Key that recognises a set of linguistic forms on the basis of formal features
Can be used in graphs and regular expressions
Outline

Concordance

Local grammar of dates

Invoking a subgraph

Lexical masks

Dictionaries of a text
Dictionaries of a text

Lexical entries for the simple words in the text

Lexical entries for the multiword units occurring in the text
Dictionaries of a text

The words in the text not found in the dictionaries
Thanks

CONTACT
ÉRIC LAPORTE
00 +33 (0)1 60 95 75 52
ERIC.LAPORTE@UNIV-PARIS-EST.FR