

# A New Lexical Chain Algorithm Used for Automatic Summarization

Edgar González   Maria Fuentes

TALP Research Center

22 October 2009

# Lexical Chains

- Cohesion links between parts of text
- Semantical relations holding between their words
  - Reiteration
  - Synonymy
  - Hypernymy/Hyponymy
  - Antonymy
  - Holonymy/Meronymy
  - ...

# Lexical Chains

---

En una trobada pública en la casa de campanya de Cárdenas i enfront dels representants dels mitjans, els tres candidats van discutir durant unes dues hores les seues propostes sobre el debat. El candidat del PAN va insistir reiteradament a celebrar esta mateixa nit esta discussió. . .

---

# Lexical Chains

---

En una trobada pública en la casa de campanya de Cárdenas i enfront dels representants dels mitjans, els tres **candidats** van discutir durant unes dues hores les seues propostes sobre el debat. El **candidat** del PAN va insistir reiteradament a celebrar esta mateixa nit esta discussió. . .

---

## Reiteration

# Lexical Chains

---

En una trobada pública en la casa de campanya de Cárdenas i enfront dels representants dels mitjans, els tres candidats van discutir durant unes dues hores les seues propostes sobre el **debat**. El candidat del PAN va insistir reiteradament a celebrar esta mateixa nit esta **discussió**...

---

## Hypernymy/Hyponymy

# Lexical Chains

---

En una trobada pública en la casa de campanya de Cárdenas i enfront dels **representants** dels mitjans, els tres **candidats** van discutir durant unes dues hores les seues propostes sobre el debat. El **candidat** del PAN va insistir reiteradament a celebrar esta mateixa nit esta discussió. . .

---

## Common Hypernymy

# Lexical Chains

---

En una trobada pública en la casa de campanya de Cárdenas i enfront dels **representants<sub>1</sub>** dels mitjans, els tres **candidats<sub>1</sub>** van discutir durant unes dues hores les seues propostes sobre el **debat<sub>2</sub>**. El **candidat<sub>1</sub>** del PAN va insistir reiteradament a celebrar esta mateixa nit esta **discussió<sub>2</sub>**...

---

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization



# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

Unable to recover from early errors

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

Able to recover from early errors

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

Able to recover from early errors

Exponential runtime cost

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

Able to recover from early errors

Linear runtime cost

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

Able to recover from early errors

Linear runtime cost

Based on local decisions

# Lexical Chains

- Applications
  - Information Retrieval
  - Natural Language Processing
    - Automatic Summarization
- Algorithms
  - Greedy
  - Non-Greedy
    - [Barzilay, 1997]
    - [Silber and McKoy, 2002]
    - [Galley and McKeown, 2003]

Able to recover from early errors

Linear runtime cost

Based on local decisions

Lack of a global optimization criterion

# Proposal

Use  
global function optimization  
methods



# Proposal

# Use Relaxation Labelling

# Proposal

## Use Relaxation Labelling

- Popular choice
- Already used for NLP tasks
  - [Padró, 1998]

# Contributions

- Propose RELAX
  - Lexical chain building algorithm based on Relaxation Labelling
- Integrate RELAX into a multilingual summarization system
  - Flexible, reusable and open source NLP components
- Make an online demo

# Outline

- 1 Lexical Chaining
- 2 Experiments
- 3 Implementation
- 4 Conclusion

# Lexical Chaining

# Setting

- Following [Barzilay, 1997]
- Iteration, Synonymy and Hypernymy/Hyponymy
  - Inclusion of Antonymy and Holonymy/Meronymy straightforward
  - Use of WordNet / EuroWordNet

# Problem

 $w_1$  $w_2$ 

...

 $w_n$ Words  $W$

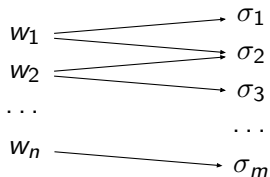
# Problem

$w_1$	$\sigma_1$
$w_2$	$\sigma_2$
$\dots$	$\sigma_3$
$w_n$	$\dots$
	$\sigma_m$

Senses  $S$

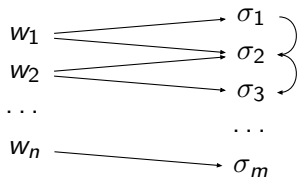


# Problem



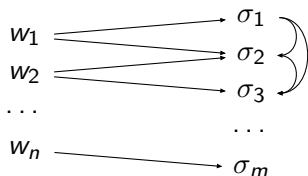
Meaning function  $\phi$

# Problem



Direct hypernymy relation  $\sigma_i \rightarrow \sigma_j$

# Problem



Hypernymy relation  $\sigma_i \xrightarrow{*} \sigma_j$

Strict hypernymy relation  $\sigma_i \xrightarrow{+} \sigma_j$

# Problem

$w_1$     $w_2$     $w_3$     $w_4$     $w_5$     $\dots$     $w_n$

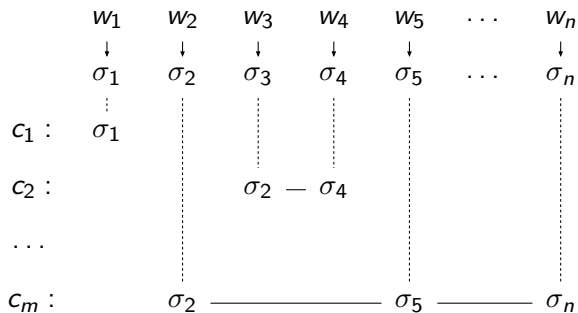
Document  $D$

# Problem

$$\begin{array}{ccccccc} w_1 & w_2 & w_3 & w_4 & w_5 & \dots & w_n \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow \\ \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \sigma_5 & \dots & \sigma_n \end{array}$$

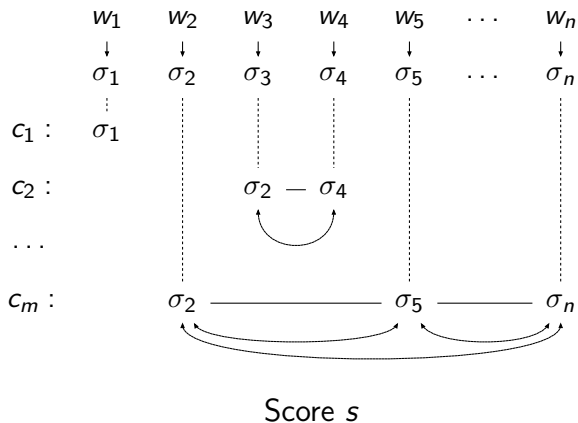
Interpretation function  $\psi$

# Problem



Lexical chaining  $C$

# Problem



# Problem

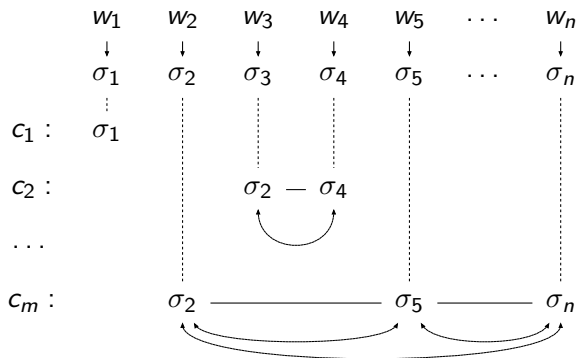
- *Optimal* interpretation  $\hat{\psi}$  and *optimal* lexical chaining  $\hat{C}$ :

$$\hat{\psi}, \hat{C} = \arg \max_{\psi, C} s(C | \psi)$$

- Maximum cardinality  $\hat{C}$



## Relaxation Labelling?



Interpretation  $\psi$   
 Lexical Chaining  $C$   
 Score  $s$

# Relaxation Labelling

- Variables  $V$
- Labels  $L_j$  for each variable
- Constraints  $R$
- Find a weighting of the labels for each variable, optimizing *Support*

# Relaxation Labelling

- Variables  $V$ 
  - $w_i$
- Labels  $L_i$  for each variable
  - $\phi(w_i)$
- Constraints  $R$ 
  - $\psi(w_i) = \sigma_i \wedge \psi(w_{i'}) = \sigma_{i'} \longrightarrow s(\sigma_i, \sigma_{i'})$
- Find a weighting of the labels for each variable, optimizing *Support*
  - Strongly related to  $s(C \mid \psi)$

# RELAX Algorithm

- 1 Generate relaxation labelling problem
- 2 Solve relaxation labelling problem
- 3 Take most weighted label to generate interpretation  $\phi$
- 4 Determine lexical chains  $c_i$

# Experiments

# Experimental Framework

- 120 Catalan news documents (EFE)
- Algorithms
  - BACKTRACK
  - SILBER
  - RELAX
- Metrics
  - Score ratio
  - Success rate
  - Runtime

## Results

	BACKTRACK	SILBER	RELAX
<b>Score Ratio</b>	100%	98.0%	99.6%
<b>Success Rate</b>	100%	54.1%	87.5%
<b>Runtime</b>	-	1.00	1.82

# Results (SILBER)

---

Melbourne (Austràlia), 24 may (EFA).- El Gran Consell de Caps va iniciar hui, dimecres, una segona reunió<sub>1</sub> en què participa un representant del colpista George Speight, per a solucionar la crisi<sub>2</sub> política que viu el país<sub>3</sub> des de l'intent<sub>4</sub> de colp d'estat<sub>4</sub> del passat divendres. Tevita Vakalalabure, designat Advocat General del Govern del colpista George Speight va acudir hui al quarter en què se celebra la reunió<sub>1</sub> dels caps<sub>5</sub> nadius. Es creu que la seua participació<sub>6</sub> té com a objectiu representar als rebels davant dels caps<sub>5</sub>, però esta suposició<sub>7</sub> no ha sigut encara confirmada de forma<sub>4</sub> oficial

...

---

Score: 80



# Results (RELAX)

---

Melbourne (Austràlia), 24 may (EFA).- El Gran Consell de Caps va iniciar hui, dimecres, una segona **reunió<sub>1</sub>** en què participa un representant del colpista George Speight, per a solucionar la crisi política que viu el **país<sub>1</sub>** des de l'**intent<sub>2</sub>** de colp d'**estat<sub>1</sub>** del passat divendres. Tevita Vakalalabure, designat Advocat General del Govern del colpista George Speight va acudir hui al quarter en què se celebra la **reunió<sub>1</sub>** dels **caps<sub>3</sub>** nadius. Es creu que la seua **participació<sub>4</sub>** té com a objectiu representar als rebels davant dels **caps<sub>3</sub>**, però esta **suposició<sub>5</sub>** no ha sigut encara confirmada de **forma<sub>2</sub>** oficial

...

---

Score: 98

# Implementation

# Architecture

- Architecture: FEMSUM
  - [Fuentes, 2008]
  - Flexible and highly modularized multitask summarization architecture
  - Parametrizable summarizers suitable to specific needs

# Implementation Framework

- Linguistic Analysis: FreeLing suite
  - [Carreras et. al., 2004]
  - Tokenization, tagging, NE detection, parsing. . .
  - Catalan, English, Galician, Italian, Spanish and Welsh
  - GNU General Public License
  - Available at:
    - <http://garraf.epsevg.upc.es/freeling>

# Implementation Framework

- Integration: General NLP Platform
  - [González, 2004]
  - Client/server architecture
  - Integration and sharing of heterogeneous, flexible, independent and reusable components
  - Free software

# Online Demo

- CGI frontend script
- Accessible at:
  - <http://nidhoggr.lsi.upc.edu/~demo/flsummary.php>

# Online Demo

**Summary Demo**

**Run**

**File**

**Language**  **Summary words**

**Lexical Chains**  Common Nouns  Proper Nouns  Verbs  Adjectives  Adverbs

**Hypernym depth**  **Heuristic**

**Summary Chains**  **Extra Options**  Remove Used Chains  
 Classify Named Entities  
 Segment-Local Resolution

**Functionality**

Currently on-line summarization of generic textual documents in Catalan, English and Spanish. Using as a preprocess the Reader 2.1 Open Source Suite of Language Analysis. Lexical Chains computed.

## Online Demo

Summary Demo - Mozilla Firefox

Eitxer Edita Visualitza Historial Adreces d'interés Eiges Ajuda

http://nidhoggr.lsi.upc.edu/~demo/cgi-bin/flsummary.cgi

Més visitades - Gmail - Safata d'en... Facebook | Home akkalabeth's Music ...

## Summary Demo

### Active Word

Form

Tag

Lemma

### Active Chain

Number

Type

Strength

### Summary

Melbourne ( Austràlia ) , 24 may ( EFA ) . - El Gran Consell de Caps va iniciar hui , dimecres , una segona reunió en què participa un representant de el colpista George Speight , per a solucionar la crisi política que viu el país des de l' intent de colp d' estat de el passat divendres .

### Process

Melbourne ( Austràlia ) , 24 may ( EFA ) . - El Gran Consell de Caps va iniciar hui , dimecres , una segona **reunió** en què participa un representant de el colpista **George Speight** , per a solucionar la crisi política que viu el **país** des de l' intent de **colp d' estat** de el passat divendres .

Tevita Vakalalabure , designat Advocat General de el Govern de el colpista **George Speight** va acudir hui a el quarter en què se celebra la **reunió** de els **caps** natius .

Es creu que la seua participació té com a objectiu representar a els rebels davant de els **caps** , però esta suposició no ha sigut encara confirmada de forma oficial .

Mentrestant **George Speight** ha despedit el **Comandant Bill** de el edifici de el Parlament on es troben els aproximadament 30 ostatges segrestats per els colpistes , després de declarar -se ahir com " cervell " de el **colp** .

Timoci Silatolu , vice primer ministre de el **govern** proclamat per els colpistes , va indicar a un periodista local que l' **expulsió** de **Bill** es va deure a el **fet** que la seua Intervenció davant de els mitjans de comunicació posava en perill la seguretat de el **grup** rebel .

El **Comandant Bill** , va dir ahir que els colpistes " matarien a els rebels abans que algú intentara rescatar-los " .



# Conclusions

# Contributions

- Propose RELAX
  - Lexical chain building algorithm based on Relaxation Labelling
- Integrate RELAX into a multilingual summarization system
  - Flexible, reusable and open source NLP components
- Make an online demo

# Contributions





- Propose RELAX
  - Lexical chain building algorithm based on Relaxation Labelling
  - Show it to outperform other popular algorithms in chain score
- Integrate RELAX into a multilingual summarization system
  - Flexible, reusable and open source NLP components
- Make an online demo

# Contributions




- Propose RELAX
  - Lexical chain building algorithm based on Relaxation Labelling
  - Show it to outperform other popular algorithms in chain score
- Integrate RELAX into a multilingual summarization system
  - Flexible, reusable and open source NLP components
- Make an online demo :)

# Gràcies!

# Bibliography I

-  R. Barzilay, *Lexical Chains for Summarization*, Ms Thesis, Ben-Gurion University of the Negev, 1997.
-  M. Galley, K. McKeown, *Improving Word Sense Disambiguation in Lexical Chaining*, Proc. of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03), 2003.
-  H.G. Silber, K.F. McKoy, *Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization*, Computational Linguistics, 28(4), pp. 487–496, 2002.
-  L. Padró, *A Hybrid Environment for Syntax-Semantic Tagging*, PhD Thesis, Dep. Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, 1998.

# Bibliography II

-  M. Fuentes, *A Flexible Multitask Summarizer for Documents from Different Media, Domain and Language*, PhD Thesis, Universitat Politècnica de Catalunya, 2008.
-  X. Carreras, I. Chao, L. Padró, M. Padró, *FreeLing: An Open-Source Suite of Language Analyzers*, Proc. of the 4th International Conference on Language Resources and Evaluation, 2004.
-  E. Gonzàlez, *Un Sistema Genèric de Cerca de Resposta*, Degree Thesis, Facultat d'Informàtica de Barcelona, Universitat Politècnica de Catalunya, 2004.