

# TexTec

## Text Processing Technologies

Activities 2005

## Research Topics 2005 (1)

1. *Textual Entailment*
  - Tree edit distance on dependency trees
2. *Information Extraction*
  - Relation extraction from biomedical literature (both supervised and unsupervised approaches)
3. *Text Categorization*
  - Supervised and Unsupervised
  - Cross Language TC
4. *Word Sense Disambiguation*
  - All Words, English
  - Latent Semantic Analysis
5. *Tools for Text Processing*
  - TextPro Platform
  - PoS tagger, EntityPro, SensePro

## Research Topics 2005 (2)

### 6. *Content Annotation*

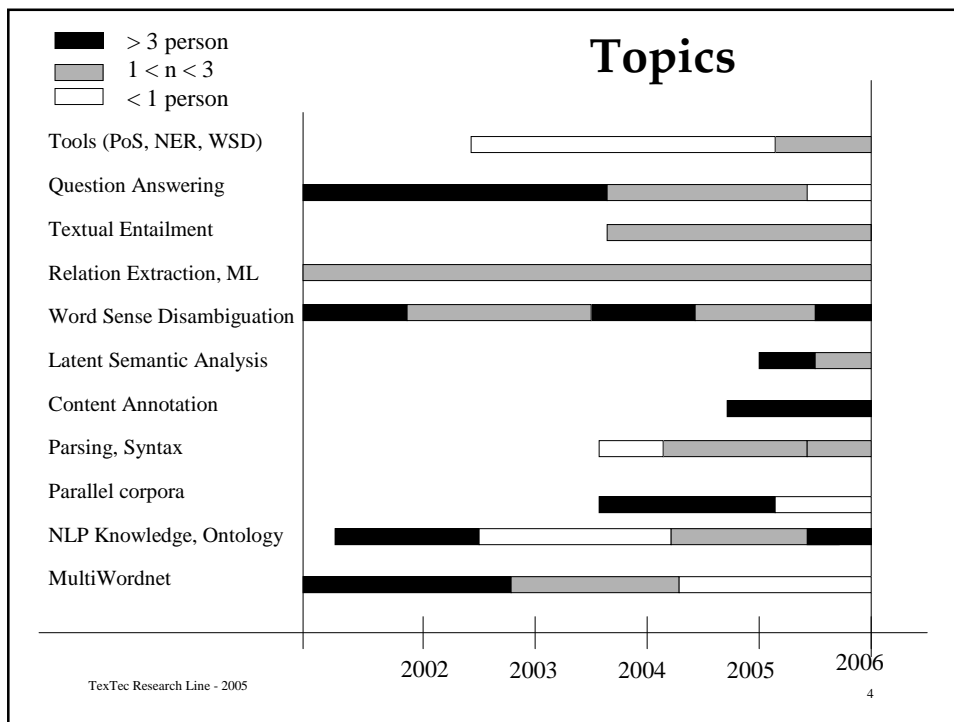
- Temporal expressions, Persons, Organizations
- Chronos (English, Italian)

### 7. *Syntax*

- Syntactic Network
- Dependency trees: Sense Discriminative Patterns
- Statistical parsing of Italian (both constituency and dependency)

### 8. *MultiWordNet*

- Sense coverage
- Images and examples
- OntoBrowser



## People

	2004	<b>2005</b>	2006
• Permanent:	4	<b>4</b>	4
• Temporary:	3	<b>3-</b>	4
• Technicians:	3	<b>3-</b>	2
• Consultant:	6	<b>4</b>	2
• Post doc:	1	<b>1</b>	1?
• PhD students:	5	<b>6</b>	4+
• Graduate students	2	<b>1</b>	1+
• Visiting students/researcher:	1	<b>4</b>	?
	<hr/>		
	26	<b>26--</b>	18+

TexTec Research Line - 2005

5

## Publications 2005

### *Books/ Proceedings*

- Ontology Learning
- CLEF proceedings (LNAI)
- Pascal proceedings (LNAI)

### *Journals*

- CL: --
- Applied AI: 1
- JNLE: 1
- Book review 1
- Signos 1

### *Program Committees*

Conferences: 9  
Workshops: 10

### *Conferences and Workshops*

- ACL	1	ws: 2
- EMNLP	2	
- Int. Conf. ML	1	
- RANLP	1!!	
- NAACL	--	ws: 1
- FLAIRS	1	
- CICLING	2	
- CLEF	2	
- CONLL	1	ws: 1
- Cognitive Science	1	
- Pascal WS	4	
- WS Meaning	1	
- Affective Comp.	1	
- Dagstuhl	1	
- SWAP	1	
	<hr/>	
Total	25	

TexTec Research Line - 2005

6

## Evaluations Campaigns (participation)

- **Recognizing Textual Entailment 2005**
  - Tree edit distance approach 7/17
  - Semantic approach (with Venice) 1/17 (with 62% coverage)
- **DUC-2005**
  - Multi-document summarization:
  - Linguistic quality 1/31 (manual)
  - Responsiveness 18/31 (manual)
  - Rouge 20/31
- **CLEF-2005**
  - Italian/Italian: 1/2
  - Italian/English: 1/1
  - Bulgarian/English 1/2
- **Information Extraction Evaluation (PASCAL)**

## Evaluations Campaigns (Organization)

- **CLEF-2005:**
  - In collaboration with CELCT
  - Multilingual QA: 24 participants (similar to TREC)
  - 9 languages
  - LNAI book
- **Recognizing Textual Entailment 2005 (PASCAL)**
  - 17 participants
  - LNAI book
- **Information Extraction Evaluation (PASCAL)**
  - ?? participants
- **2006**
  - RTE-2, in collaboration with CELCT
  - CLEF-2006, in collaboration with CELCT
  - Senseval ??

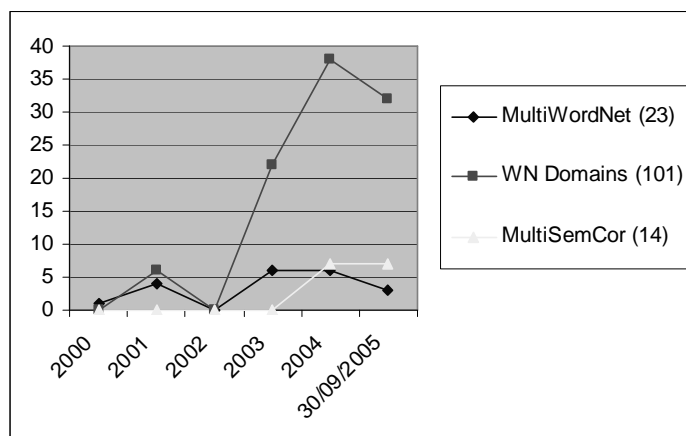
## Organization of Events

- **MEANING final workshop**
  - Trento, February 2005
  - 70 participants
- **PASCAL-RTE**
  - Southamton (UK), 11-13 April 2005
- **Workshop on Cross-Language Knowledge Induction**
  - Euroalan, Cluj, August 2005, 40 participants
- **CLEF 2005** (Vienna, September 2005)
  - Workshop on multilingual question answering
- **EACL 2006** (local organization)
  - Trento, 3-7 April 2006
  - 250 participants expected
  - Organization of three workshops
    - Multilingual Question Answering
    - Cross Language Knowledge Induction
    - Adaptive Text Extraction and Mining

TexTec Research Line - 2005

9

## Resource Distribution



**Multiwordnet:** research 250 euros

**WordNet Domains:** research: free; commercial: 10000 (1 license)

**Multisemcor:** research: free

**TIES:** 2 licenses

TexTec Research Line - 2005

10

## Collaborations

- Bar Ilan University (Ido Dagan)
- Haifa University (Shuly Wintner)
- ISTI-CNR (Fabrizio Sebastiani)
  
- CELCT: Centre for the Evaluation of Language and Communication Technologies (Trento)
  
- University of Rome Tor Vergata (Roberto Basili)
- Other collaborations
  - University of Alicante
  - DFKI

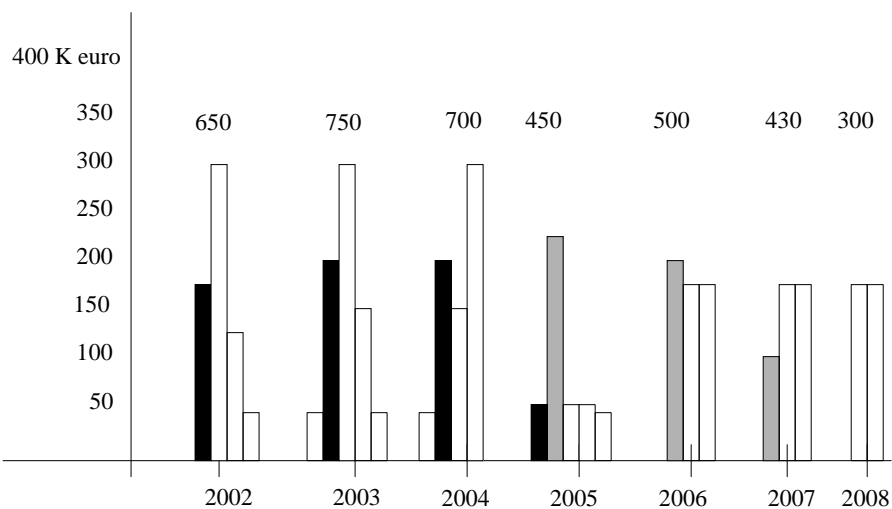
## Fund Raising (1)

- IST program
  - **Knowledge Technologies** (March 2005)
    - UBIQUA: strep on question answering (rejected)
    - EPOS: strep on opinion mining (rejected)
    - ADAPT-EU: IP on text processing for e-learning (rejected)
    - XMEDIA: IP on knowledge extraction (**accepted!**)
    - PATEXPART: strep on text mining for patents (**accepted!**)
  - **Multimodal Interfaces** (September 2005)
    - QALL-ME: strep on question answering (???)
    - MEANING+: strep on knowledge acquisition (???)
    - MIAMI: IP on text mining for the medical domain (???)
  - **Marie Curie**
    - KLUB: funding for organizing events (rejected)
    - M-SALT: semantic annotations in several languages (???)

## Fund Raising (2)

- **MIUR** (June 2005)
  - Furb with University of Pisa (???)
  - PROCURE on Knowledge Management with the University of Trento (???)
- **Italy-Israel Program** (June 2005)
  - A task on Semantic driven Information Retrieval (???)
- **E-Content Plus** (November 2005)
  - CULT: strep on intelligent access to multimedia archives (with Rome, Tor Vergata)
  - Multilingual Language technologies to access scientific publications: with the University of Trento
  - Multilingual portal for social science in Central and South Eastern Europe: with the Province of Trento

## Projects



## Education

- **PhD School in Trento:** Foundations of Human Language Technologies
  - 10 PhD students
  - In collaboration with the SSI division
- **Master in Language and Communication Technology** (University of Bolzano)
  - Text Processing: 6 students, lectures and projects
- **Eurolan 2005 Summer School**
  - Multilingual Question Answering
  - Exploiting Parallel Corpora
- **Tutorial at RANLP 2005** on Question Answering
  
- **2006:** PhD school, LCT Master, ESSLLI