

Computational Linguistics Lab: Introduction

Elena Cabrio, FBK-Irst
mail: cabrio@fbk.eu

Outline:

- ▶ Lab Info
- ▶ Program
 - ▶ Labs
 - ▶ Lessons
 - ▶ Grading
- ▶ Critiques & Projects suggestions
- ▶ Something about you...

Labs info:

- ▶ **Time:** Wednesday 18-19
 - ▶ Other possibilities?
- ▶ **Office hours:** all academic year, by prior arrangement via e-mail
- ▶ **Labs material:**
 - ▶ Slides
 - ▶ Readings

- ▶ **Info & slides:**

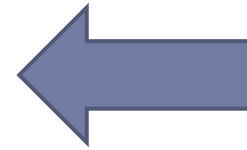
<http://www.inf.unibz.it/~bernardi/Courses/CompLing/10-11.html>

Program: LABS (first part)

- ▶ **Lab 1** (06/10/2010): Intro
- ▶ **Lab 2** (13/10/2010)
 - ▶ How to review a paper?
- ▶ **Lab 3** (20/10/2010)
 - ▶ Corpus linguistics
- ▶ **No lab on 27/10/10!**
- ▶ **Lab 4** (03/11/10)
 - ▶ Exercises of text annotation exploiting WN synsets
 - ▶ Reading group: WordNet vs Wikipedia
- ▶ **Lab 5** (10/11/10)
 - ▶ Reading group on Textual Entailment
- ▶ **Lab 6** (11/11/10) (Thu 10.30-12.30): **Critiques presentation**

Grading:

▶ **Critiques and Projects (50%)**



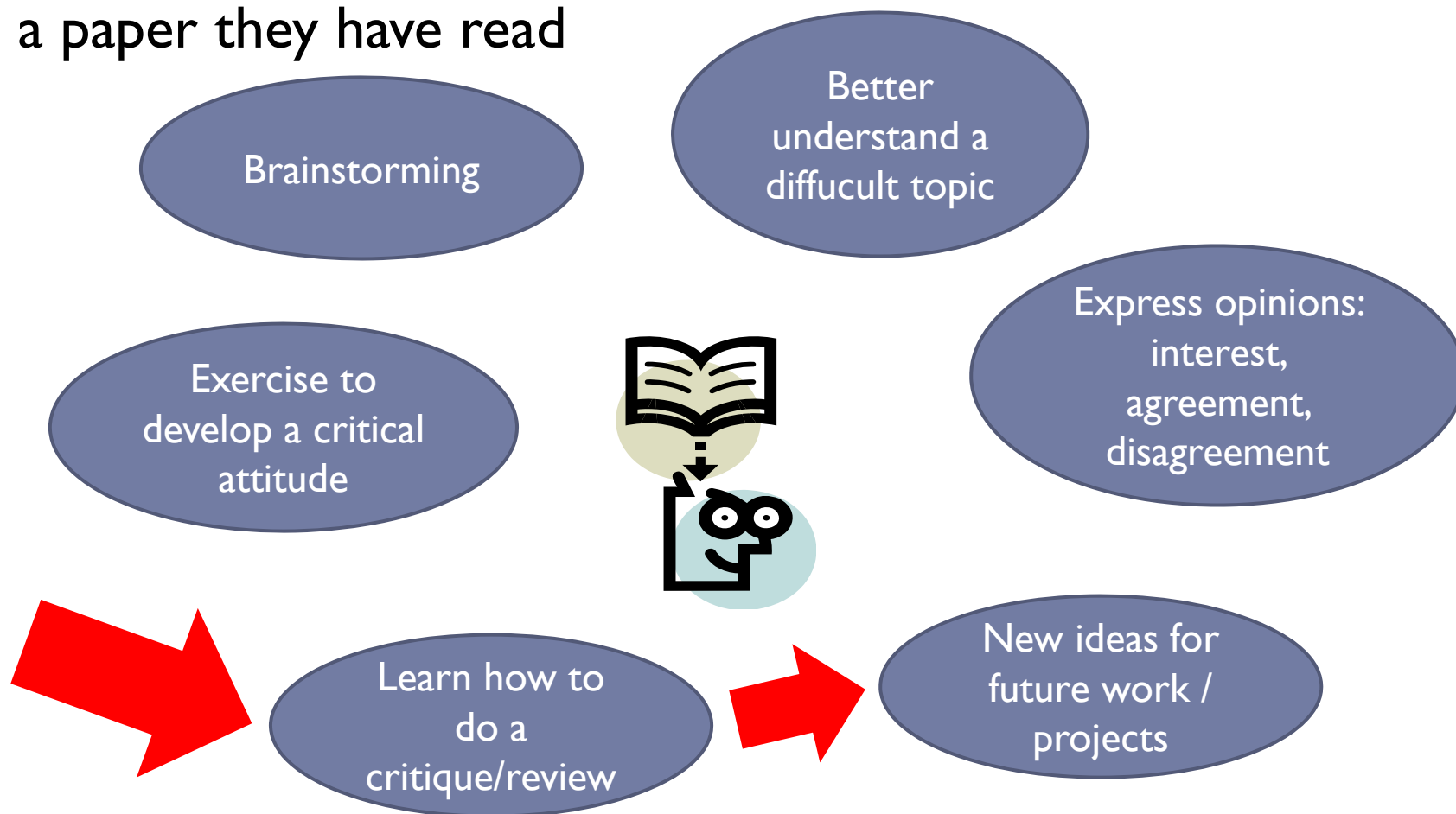
▶ **Final Exam (50%)**

By the way: what is a reading group?



By the way: what is a reading group?

- ▶ It is a group of students/researchers who meet to discuss a paper they have read



Critiques & Projects:

▶ **CRITIQUES:**

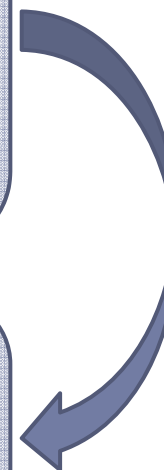
- ▶ Write a critique about a paper that presents/ describes some interesting aspects of a topic (related to the arguments presented during the course/labs)

▶ **Presentation: 11/11/2010**

▶ **PROJECTS:**

- ▶ On the bases of the reviewed paper, propose and carry out a project

▶ **Presentation: 20/01/2011**



Critiques & Projects: suggestions...

1. TEXT STATISTICS

- ▶ PAPER: William H. Fletcher, "Making the Web More Useful as a Source for Linguistic Corpora"
- ▶ PROJECT: Given a sample of a book, calculate number of types, number of tokens, type/token ratio, possible collocations + data analysis

2. COLLOCATION EXTRACTION

- ▶ PAPER: Dekang Lin, "Extracting Collocations from Text Corpora"
- ▶ PROJECT: Investigate various methods of extracting collocations from a corpus and finding relations between them

3. AUTOMATIC KEYPHRASE EXTRACTION FROM SCIENTIFIC ARTICLES

- ▶ PAPER: Ken Barker and Nadia Corracchia, "Using noun phrase heads to extract document keyphrases", or choose a paper from the SemEval 2010 task proceedings (<http://aclweb.org/anthology-new/S/S10/>)
- ▶ PROJECT: Implement a simple system to automatically extract keyphrases from scientific articles (a simplified version of the SemEval 2010 task)

Critiques & Projects: suggestions... (cont.)

4. AUTOMATIC DISCOVERING OF WN RELATIONS

- ▶ PAPER: Marti A. Hearst, "Automatic Acquisition of hyponyms from large text corpora"
- ▶ PROJECT: Automatic discovery of WN relations (decide on a lexical relation, e.g. meronymy; pick a list of word pairs in WN in which the relation holds; extract sentences from large corpora in which this terms occurs; find the commonalities about lexical and syntactic patterns and hypothesize patterns)

5. CORPUS-BASED WORD SENSE IDENTIFICATION

- ▶ PAPER: Claudia Leacock, Martin Chodorow: "Combining Local context and WordNet similarity for Word Sense Identification"
- ▶ PROJECT: Exploiting local context for word sense identification (experiment)

6. EXTENDING MULTIWORDNET TO A NEW LANGUAGE

- ▶ PAPER: Bernardo Magnini, Carlo Strapparava, Giovanni Pezzulo, Alfio Gliozzo, "Comparing Ontology-Based and Corpus-Based Domain Annotation in WordNet"
- ▶ PROJECT: Investigating techniques to extend MWN to a new language

Critiques & Projects: suggestions... (cont.)

7. PRECISION-ORIENTED TE MODULES

- ▶ Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer, Christopher Manning, "Learning to recognize features of valid textual entailments"
- ▶ Implement TE modules (within the EDITS architecture) to handle specific linguistic phenomena relevant to inference

Other proposals are welcome!

Arguments should be chosen by 27th October.

You are strongly encouraged to present your critiques to the other students on 11/11/10 and projects on 21/01/2010!

Something about you...

Background?

Interests?

Next lab: How to review a paper?

- ▶ **We'll review together the paper:**

Aurlien Max, Guillaume Wisniewski, *“Mining Naturally-occurring Corrections and Paraphrases from Wikipedia’s Revision History”*

(http://www.lrec-conf.org/proceedings/lrec2010/pdf/827_Paper.pdf)

- ▶ **Please, read it carefully by next Wednesday!**

- ▶ **Focus on:**

- ▶ **Points of strengths/limitations of the approach**
 - ▶ **Is the approach clearly described? Well evaluated? Applicable to real scenarios?**

Program: LESSONS

All lessons (and the second part of the labs) will be taught by Raffaella Bernardi, except:

- ▶ **28/10/10**

- ▶ **Lexical Semantics**

- ▶ **4/11/10**

- ▶ **Textual Entailment**