

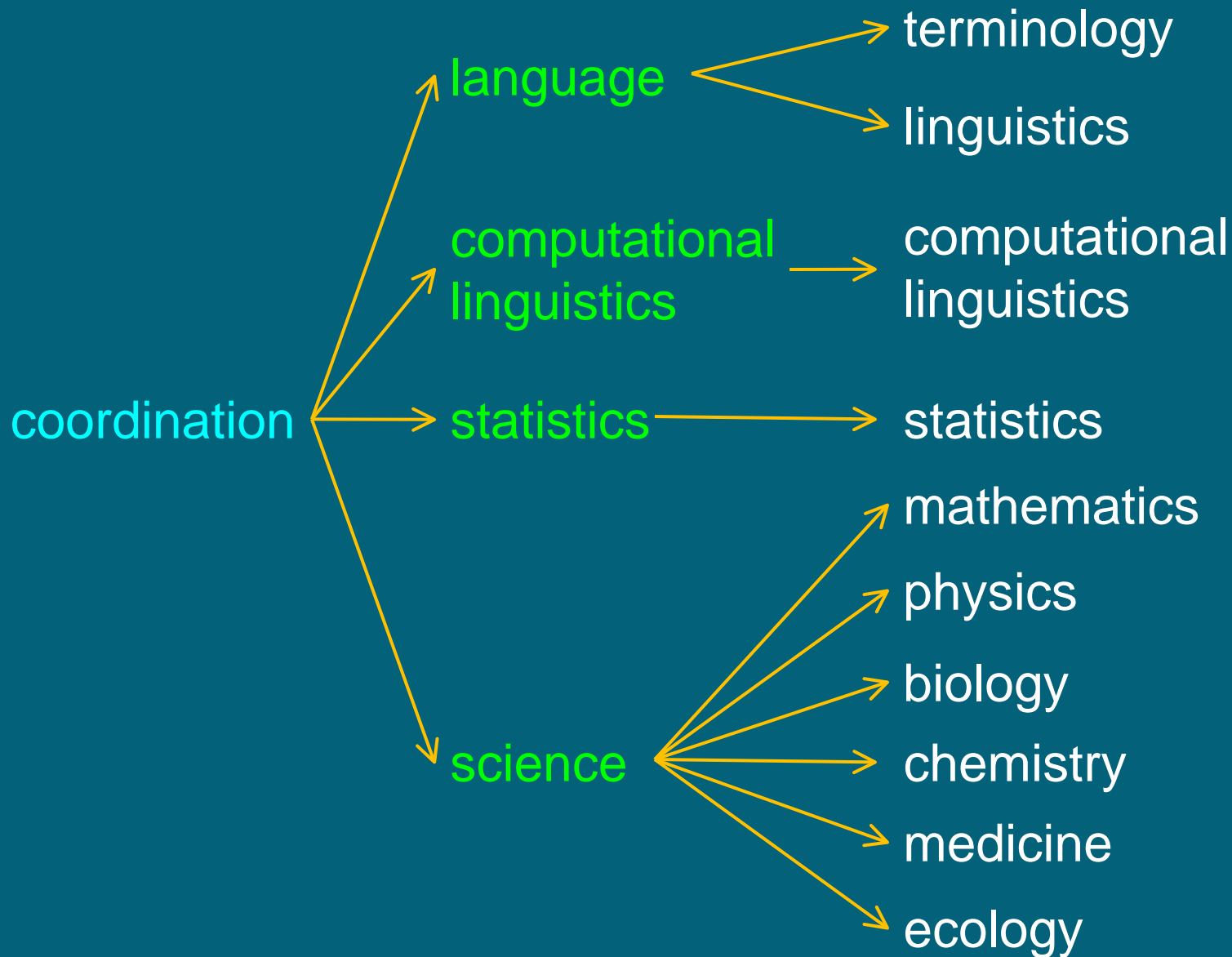
**IV Terminology Summit**  
**Gatineau, Québec, 7 – 8 October 2008**



**The basic scientific  
terminology of Mexico: a truly  
interdisciplinary project  
Spanish – English**

María Pozzi  
*El Colegio de México*

# Basic scientific terminology



# The process

1. Corpus construction
2. Lemmatisation of each word
3. Term extraction – canonical form
4. Quantitative analysis
5. Qualitative analysis
6. Finding English equivalents

# At a later stage...



- Publication of a basic dictionary of science Spanish – English
- Evaluation of how scientific concepts are being introduced to students

# Corpus of Mexican scientific texts



	Primary						Secondary			Preparatory			Total
	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	
Maths	*	*	*	*	*	*	*	*	*	*	*	*	12
Natural Sciences	*	*	*	*	*	*							6
Biology							*				*	*	3
Physics								*		*		*	3
Chemistry									*		*	*	3
Anatomy											*		1
Ecology						*						*	2
<b>Total</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>4</b>	<b>5</b>	<b>30</b>

# Corpus of Mexican scientific texts



- newspaper articles on science
- popular science articles
- popular medicine articles
- articles on environmental issues

# Corpus markup

- **text heading**: full bibliographic reference, science to which the text belongs, school grade to which the text belongs
- **text structure**: titles, subtitles, indexes, tables of content, glossaries, symbols, abbreviations, etc.
- **lemma** corresponding to each word and part of speech
- **term** (one word or multiword term) and part of speech

# Lemmatisation



- two computer programs are being used to lemmatise and assign the corresponding POS to each lemma
- manual disambiguation



# Term extraction



- two computer systems will be used to extract terms automatically
- manually complemented
- canonical form manually assigned
- semiautomatic assignment of POS to each term

# Quantitative analysis



- **absolute frequency**

number of occurrences in COCIEM of each lemma

- **relative frequency**

rate of occurrences of each lemma in relation to the total within each science and between sciences

# Quantitative analysis

- **corrected frequency**

measure that combines the frequency of a lemma with its dispersion between sciences, in such a way that for lemmas of similar absolute frequency, those with a more uniform dispersion have a higher corrected frequency

- **dispersion index**

index measuring how lemmas are distributed among the different sciences, taking into account the inequalities in the size of the sample for each science

# Quantitative analysis

These measures will allow us to know:

- the basic scientific vocabulary used in Mexico,
- the basic terminology of each science
- scientific terms shared by two or more sciences,
- the most important and the most common scientific terms,
- the scientific knowledge that should be known by a specific group of students.

# Qualitative analysis



- morphologic
- syntactic
- collocations, phraseology, etc.
- term formation processes
- semantic

# English equivalents

- each term will be assigned its corresponding English term(s)
- these will be the basis for the Spanish – English and English – Spanish indexes

# The basic vocabulary of science



- Mexican Spanish, with English equivalents
- in printed format
- multimedia Web-based
- user-oriented, i.e., geared towards the easier understanding of scientific concepts

# How concepts are introduced

- Mexico has a poor level of scientific education
- students are encouraged to learn definitions and formulae by heart
- identifying definitions and other concept descriptions will allow us an evaluation of how scientific concepts are introduced to students
- if necessary, a recommendation will be made to change textbooks



# Final considerations



- this is the first time ever that a multidisciplinary group of specialists has come together to think and do something about the state of scientific terminology in México
- having been made aware of the problems scientists are willing to participate in future work
- this means, hopefully, a future for terminology in México