"Confortation"

a new qualitative category for analyzing biomedical texts

Delphine Battistelli[1], Antonietta Folino[2], Patricia Geretto[2], Ludivine Kuznik[2], Jean-Luc Minel[1], Florence Amardeilh[1]

[1] **MoDyCo-UMR 7114,** Université Paris X -CNRS, F92001 Nanterre, France

delphine.battistelli@paris-sorbonne.fr, jean-luc.minel@u-paris10.fr, florence.amardeilh@mondeca.com

[2] **Université Paris X,** F92001 Nanterre, France

antonietta.folino@unical.it, pgeretto@hotmail.fr, ludivine.kuznik@gmail.com

Contents

- **Context: Microbio Project**
- **Epistemic modality: state of the art, some approaches**
- **Methodology**
  - ✓ **Linguistic Analysis**
  - ✓ **Semantic notions map**
  - ✓ **Towards an automatic semantic annotation**
- **Conclusion: future research**

Context: Microbio Project

- **Biomedical research on miRNAs**
  - ✓ MicroRNAs: small RNA molecules encoded in the genomes of plants and animals
  - ✓ MiRNAs involved in development and inhibition of some diseases
  - ✓ Huge interest in the biology community
  - ➔ Dramatic increase in the amount of published information
- **Microbio project: 1/1/2008–12/31/2009 –Pasteur Institute, Uruguay**

Institut Pasteur de Montevideo, Uruguay
Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba, Argentina
INCO, **Facultad de Ingeniería,** Universidad de la República del Uruguay
Pontificia Universidade Católica do Rio Grande do Sul, Faculdade de Informatica, Brasil
Universidad de Concepción, Chili
MoDyCO *(Modèles, Dynamiques, Corpus)* UMR 7114 CNRS, Université Paris Ouest Nanterre la Défense, France
LORIA *(Laboratoire Lorrain de Recherche en Informatique et ses Applications),* Nancy, France

- ✓ Extraction information tool and ontology
- ✓ A need to ascertain the validity of new knowledge
- ✓ A collaboration between biologists, computer scientists and linguists

Epistemic modality
State of the art: some approaches

- **Thompson et al. 2008**
  - ✓ Level of certainty: absolute, high, medium, low
  - ✓ Point of view: authors' or others' ideas
  - ✓ Knowledge type: speculative, deductive, sensory and **demonstrative** (*find*, *demonstrate*, *show*, *confirm*, etc.)
- **Kilicoglu et al. 2008**
  - ✓ Unhedgers: *demonstrate*, *show*, *prove*... : strong degree of certainty in positive sentences and one of hedging in negative ones.
- **Eva Thue Vold 2008**

Epistemic modality markers are

*" Linguistic items that explicitly qualify the **truth value of a proposition** "*

Linguistic analysis: corpus

- **Corpus**
  - ✓ 3 corpora: successive steps
  - ✓ Articles on miRNAs from Pubmed (Medline)
  - ✓ Search on: « miRNAs and human »
  - ✓ 10, 20 and 13 articles on miRNAs
- **Study**
  - ✓ Linguistic analysis and manual annotations
  - ✓ Test of an automatic annotation: Gate platform
- **Articles**
  - ✓ Articles about biologists' experiments
    - ✓ Objectives, means, results and phases of the research process

Linguistic analysis: Observations

- *"we **found that** the 4-nt insertion (MBS-4ntin) produced the largest recovery of GFP expression among the decoys tested"*
- *"the 4-nt insertion (MBS-4ntin) **were found to** produce the largest recovery of GFP expression among the decoys tested"*
- *"the 4-nt insertion (MBS-4ntin) **produced** the largest recovery of GFP expression among the decoys tested"*

Markers

reveal, show, find, report, result, observe, determine…

Observations: data observations, results of an experiment

Linguistic analysis: " Confortation "

- *"These results **confirm** efficient nuclear export of the synthesized TuD RNA molecules"*
- *"the relative low density of miRNA-binding SNPs at the 3'-UTRs of human genes **supports** the important role of miRNA–target interaction."*
- *"Recently, a strong link between mirRNAs that are altered… and cancer biology **has been demonstrated**."*

Markers

demonstrate, validate, support, consistent with…

"Confortation": consolidation, strengthening, reinforcement

Linguistic analysis: General knowledge

- *"It is currently estimated that the expression of… is…"*

Markers

"it is well known", "currently"…

"RNA silencing (RNAi) is a new gene regulatory mechanism".

General knowledge: a truth presented as shared by the scientific community.

Linguistic analysis: Objects of study

- *"**To address whether** …, **we first tested** …"*
- *"We next tested **whether** the regulation of ER by miR-206 occurred through direct targeting of the ER 3 -UTR."*
- *"**To test** the generality and specificity of the inhibitory effects of TuD RNA, **we constructed** a reporter cell system for miR-140-5p …"*

Markers

whether / if

infinitive constructions

Objects of study: object of the biologists' experiment, scope of their investigation

Linguistic analysis: Hypotheses

- *"these two findings **suggest** that miRs **should** be ..."*
- *" …dopaminergic neurons in midbrain **might** benefit from FGF20 for proliferation, differentiation, and even protection from stress assault. "*

Markers

Modal auxiliaries *may, might...*

Modal items (verbs, adjectives, adverbs, nouns) *suggest, believe, hypothesis, likely ...*

Hypotheses: deduction, speculative idea derived from the results of an experiment

Linguistic analysis: Lack of knowledge

- *"the biological functions of miRNAs are largely unknown"*
- *" the functional genetic variants remain to be determined"*

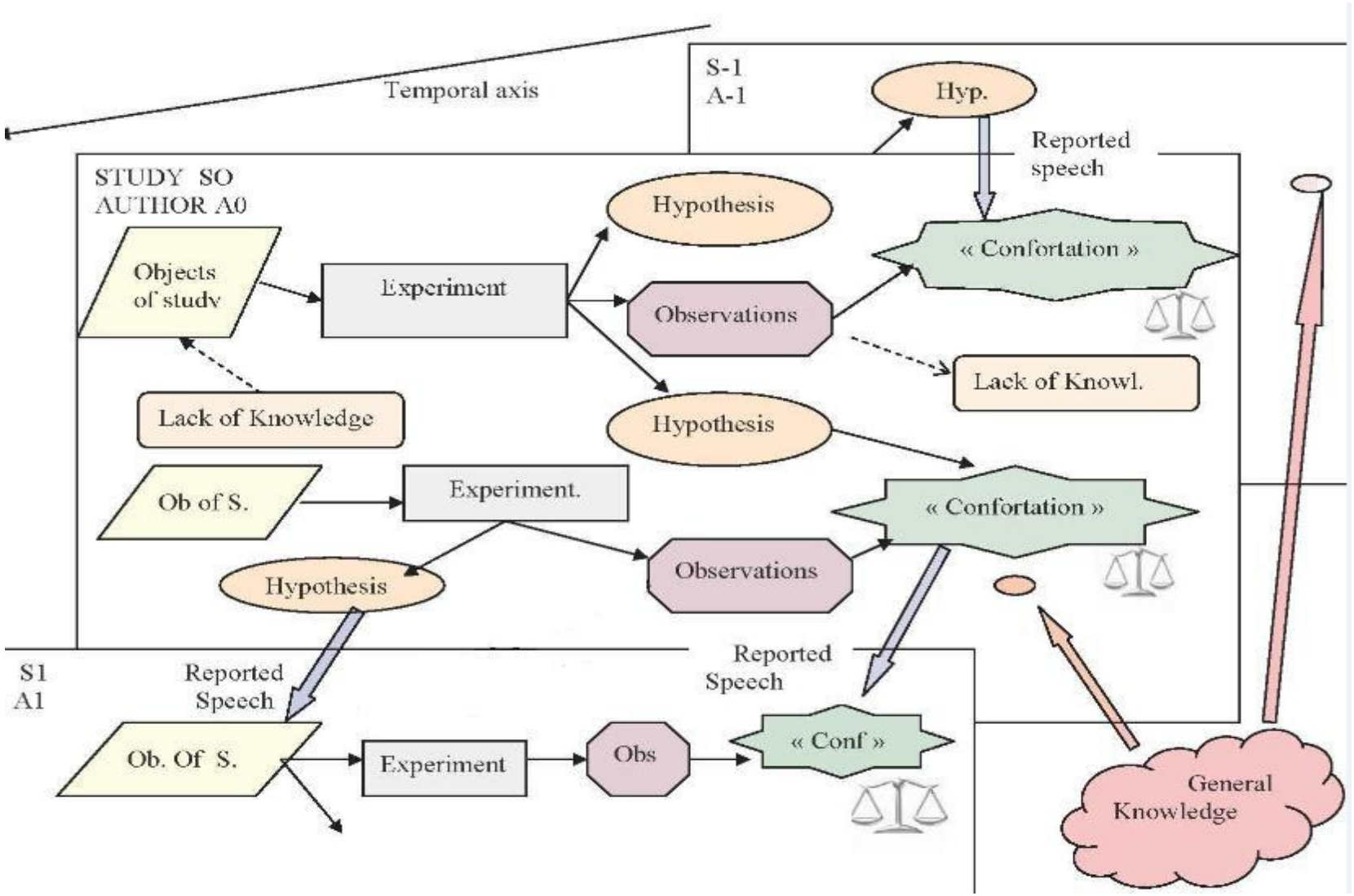
Markers

"be unknown...", "remain unclear...", "remain to be determined"

- *"however, their biological functions… remain largely undefined and experimentally untested".*

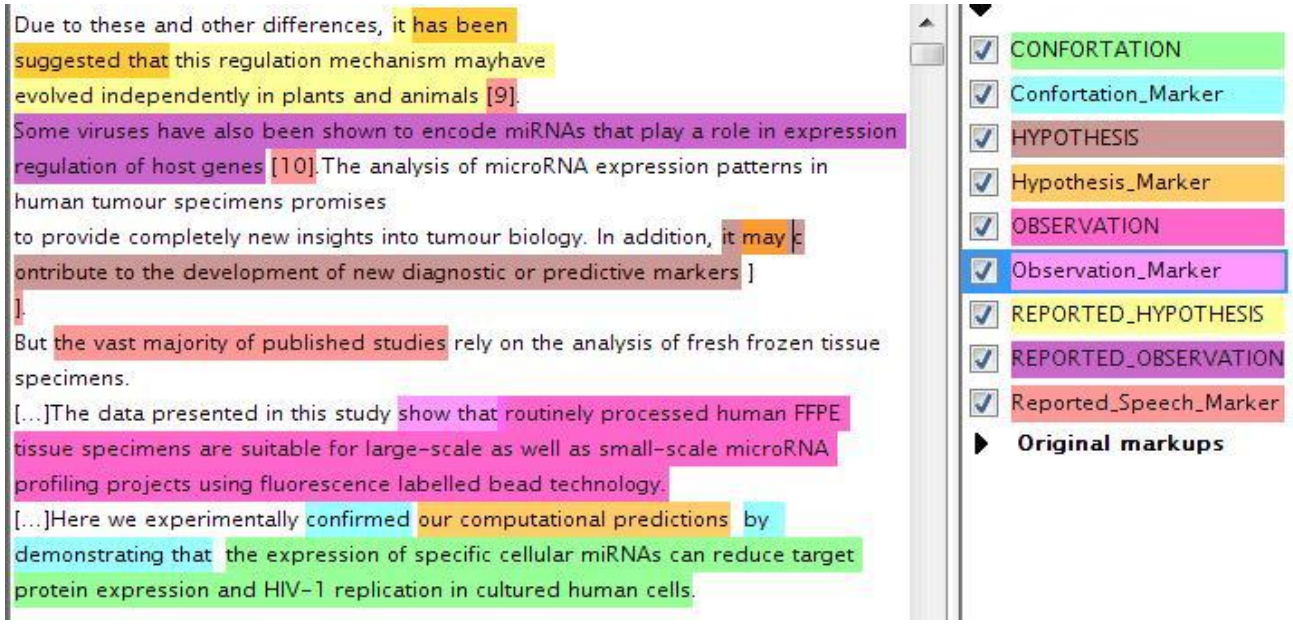
Lack of knowledge: but not uncertainty

Semantic notions map



Towards automatic semantic annotation

- **Our automatic annotation tool is still under development.**
- **The main limitation: right scope**
  - ✓ for sentences with *show*, detection of the proposition is needed;
  - ✓ for reported speech, the whole sentence is necessary;
  - ✓ in some cases such as for example anaphoric expressions in which *this* or *these* is used, the context before the sentence is essential.



Conclusion: Future research

- *"Confortation"* of observations and hypotheses: used by the biologists

➔ what is finally the weight of truth-value for each observation and hypothesis put forward?

- "Confortation" expressed by the authors on their own observations and hypotheses put forward in one article

➔ the truth value weight for an observation or hypothesis can change in the course of an article

Future investigation

Sentences without markers

Methods in biomedical articles

Verbs : tenses, negative modal verbs and passive voice

Thank you for your attention