

CONTEXT AS A
NON-ONTOLOGICAL
DETERMINANT OF SEMANTICS

Simone Santini and Alexandra Dumitrescu

Escuela Politécnica Superior

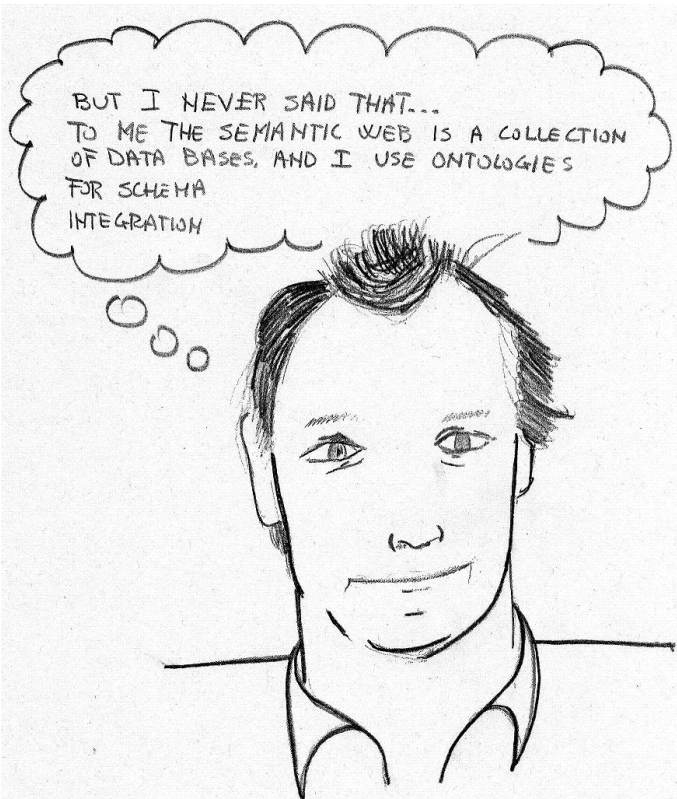
Universidad Autónoma de Madrid

THE SEMANTIC WEB: FORMAL ANNOTATION AS A WAY TO MEANING

↑
(if only!)

In general, semantics is the study of meaning. (The word "semantic" comes from the Greek *semantikos*, or "significant meaning," derived from *sema*, or "sign.") Semantic Web technologies help separate meanings from data, document content, or application code, using technologies based on open standards. If a computer understands the semantics of a document, it doesn't just interpret the series of characters that make up that document: it **understands the document's meaning**.

(IBM, "The future of the Web is Semantic", published on-line, my emphasis)



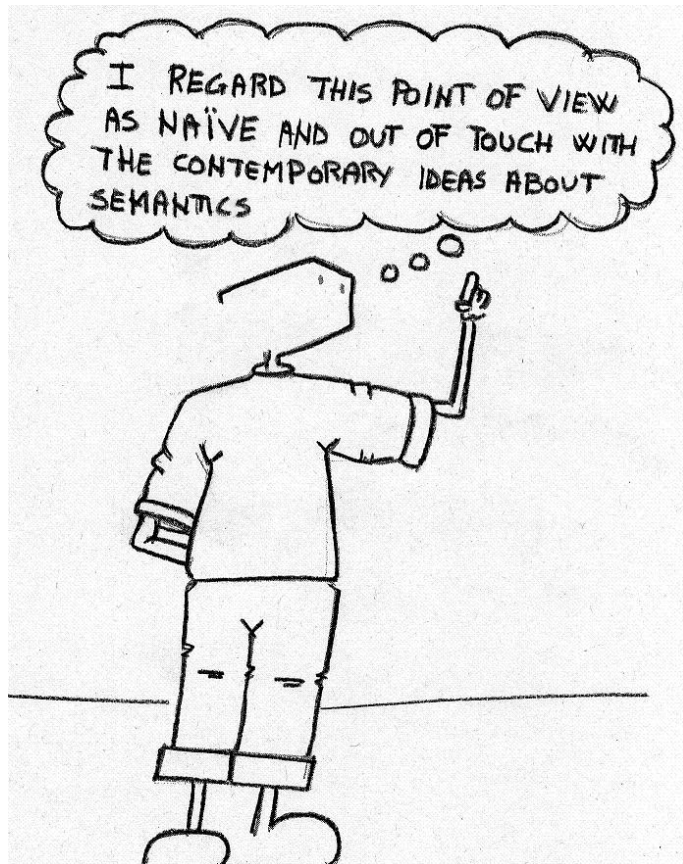
The crucial element for the definition of meaning in the semantic web is represented by **ontologies**.

We are interested in the conditions (and the limits) under which computing science can be involved in *true* semantics, not the semantics of the "Berner-Lee's semantic web", as Yorick Wilks put it.

TWO ALLEGED CHARACTERISTICS OF FORMAL ANNOTATION

↑
(if only!)

annotation represents the **pure meaning** of the data, distilling it from the superstructure and uncertainty of language



formality allows one to make the same semantic assumptions that one does in data bases, that is, that semantics can be defined using more or less the same methods that one uses in data bases (i.e. that **tarskian** semantics is enough).

I see so many problems with this it ain't even funny. Briefly... two orders of problems:

- i) pragmatic;
- ii) theoretical.

Human nature part I: **PEOPLE LIE**

From: "Caja Madrid" <kk74374bsh33@terra.es>
To: Unknown
Fecha: Tue, 29 Apr 2008 18:05:49 +0200
Asunto: Acceso a su cuenta ha sido limitado

CAJA MADRID - AVISO DE SEGURIDAD

Estimado/a cliente,
Recientemente hemos revisado su cuenta de Caja Madrid, y necesitamos ms informacin con el fin de proporcionar un servicio eficaz. Por esta razn su acceso a la cuenta ha sido limitado. Con su colaboracion deseamos restaurar su acceso tan pronto como sea posible. Disculpe las molestias.

From: Chew Ong <health.dynamics@bigpond.com>
To: undisclosed-recipients:
Fecha: Sat, 10 May 2008 5:06:45 +1000
Asunto: URGENT ATTENTION !!!

Good Day,

I am Chew Ong, an attorney at law. A deceased client of mine, who hereinafter shall be referred to as my client, died as the result of a heart-related condition on the 11 November 2001. His heart condition was due to the death of all the members of his family in the Gulf Air Flight Crashes in Persian Gulf Near Bahrain Aired August 23, 2000 - 2:50 p.m. ET as reported on:<http://transcripts.cnn.com/TRANSCRIPTS/0008/23/bn.08.html>
I have contacted you to assist in distributing the money left behind by my client before it is confiscated or declared unserviceable by the bank where this deposit valued at TWELVE MILLION EIGHT HUNDRED THOUSAND UNITED STATES DOLLARS. (\$12,800.000.00 USD). is lodged.

From: "Lee Mackey" <a-b-c-d-123@abichandani.com>
To: s.sanchez@uam.es
Fecha: Mon, 28 Apr 2008 18:33:45 +0100
Asunto: luck is one step ahead

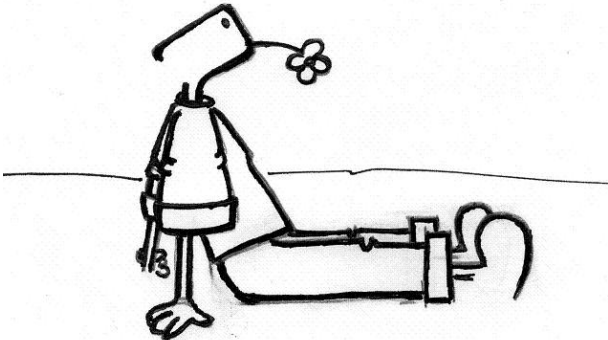
Hello! I am bored this afternoon. I am nice girl that would like to chat with you. Email me at Lena@likeihape.cn only, because I am using my friend's email to write this. To see some pictures of me.

Human nature part II: **PEOPLE ARE LAZY**

How many send you
emails without subject
line?

How many of your
friends have files
called *untitled.doc*?

How many send you
emails with a subject
line not at all related
to the content of the
message?



How many keep most of
their files on the
desktop?

Human nature part III: **PEOPLE ARE STUPID**

I made some searches on ebay, a place where it is in your own interest that the things you sell be spelled right

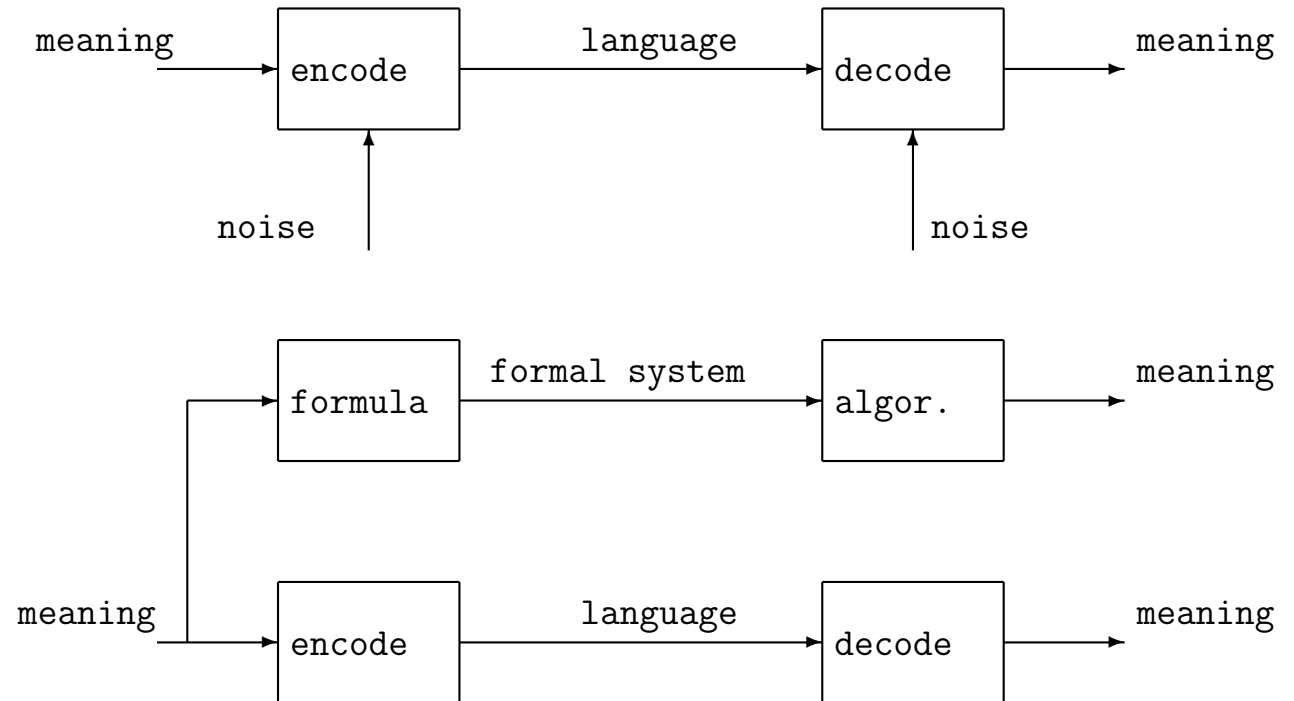
search word	items found
plam	26
imb	16
morotola	8
aple	134
micosoft	4
bealtes	4
Tostoy	2
panting	21 (*)
leonado	13

(*) one of them actually referred to the verb *to pant*: it was a T-shirt with the picture of a panting wolf.

THE THEORETICAL BASIS OF THE SEMANTIC WEB

An ontology can be used (among other things) to **represent** the meaning of a document of a fairly general nature, such as those found on the web.

Meaning is an *essence* that originates from the author; the reader does not participate in the creation of meaning: he is a passive receptor of a *commodity* that is exchanged through the currency of language.



Linguistics meaning is not codified by language, but **created** by it;
meaning is unstable, because so is the language;
the reader changes the frame of reference in which the text is interpreted,
forcing a reinterpretation;
interpretation is a dialectic process, and meaning is not a property of the
text, but of this process.

The nature of classification: **NO CLASSIFICATION IS NEUTRAL OR INNOCENT**

A classification imposes a set of values, and using the classification imposes those values on the user.

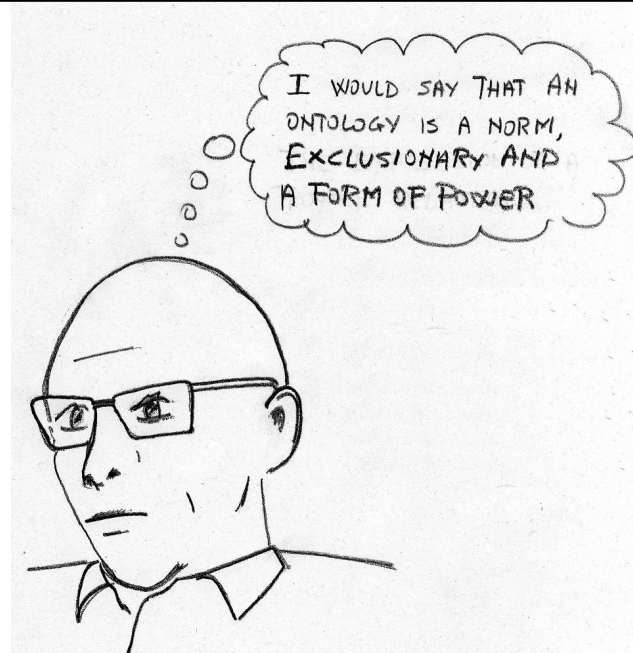
```
Car:
|
+---- sedan ---- full size
|             |--- midsize
|             |--- compact
|
+---- SUV ---- full size
|             |--- midsize
...
```

```
Car
|
+---- Low mileage ---- full size sedan
|                               |--- SUV -- full size
|                               |-- midsize
|
+---- Mid range ---- midsize sedan
|
+---- High mileage --- compact
|                               |-- hybrid
```

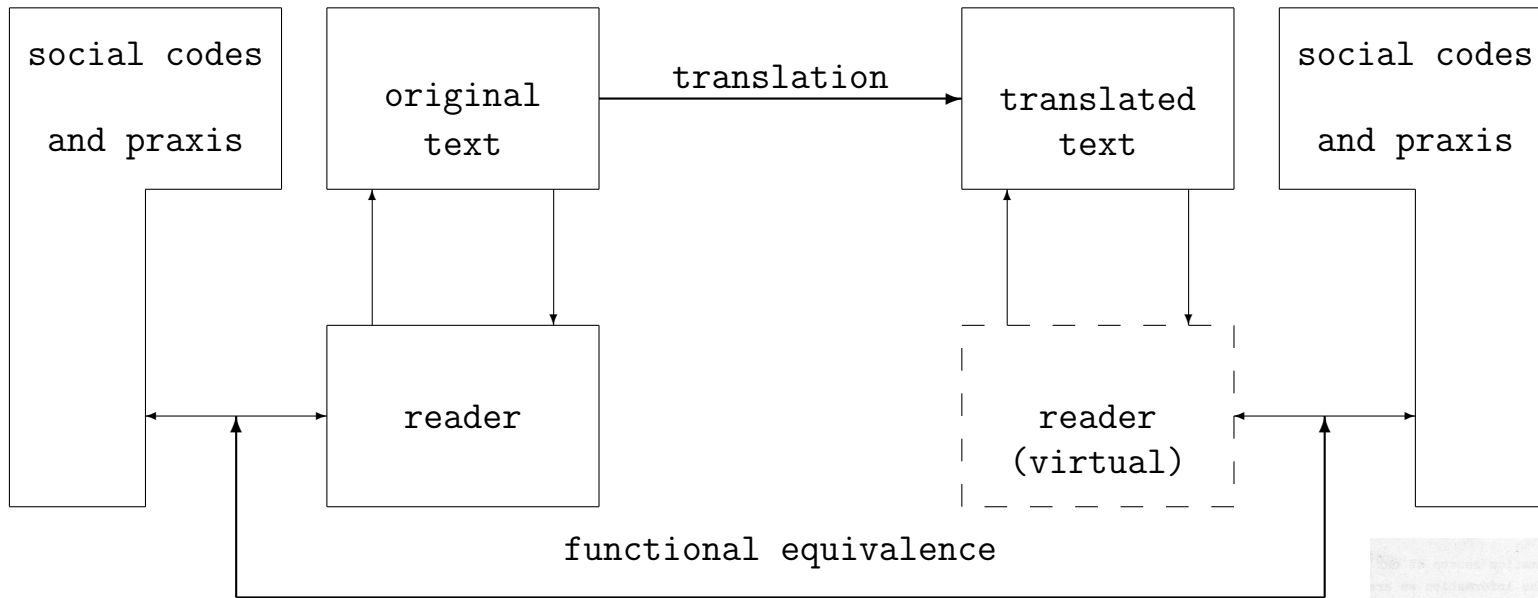
A classification always represents an ideological commitment that is forced upon whoever uses it.

It is not a matter of wanting to impose or not: it is in the nature of classification that this will happen.

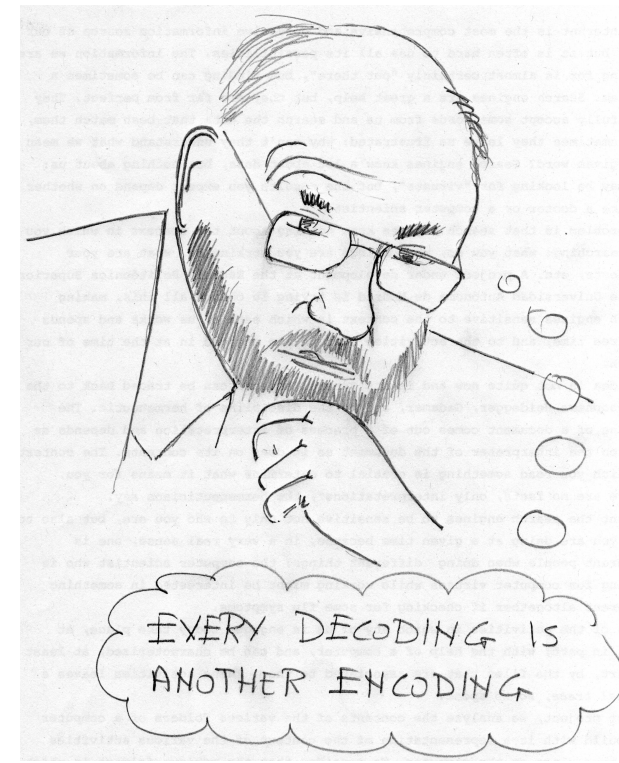
A classification is always, inevitably, an act of power, independently of whether the people who do the classification intend this to happen or not.



TRANSLATION AS FUNCTIONAL EQUIVALENTE (EUGENE NIDA)

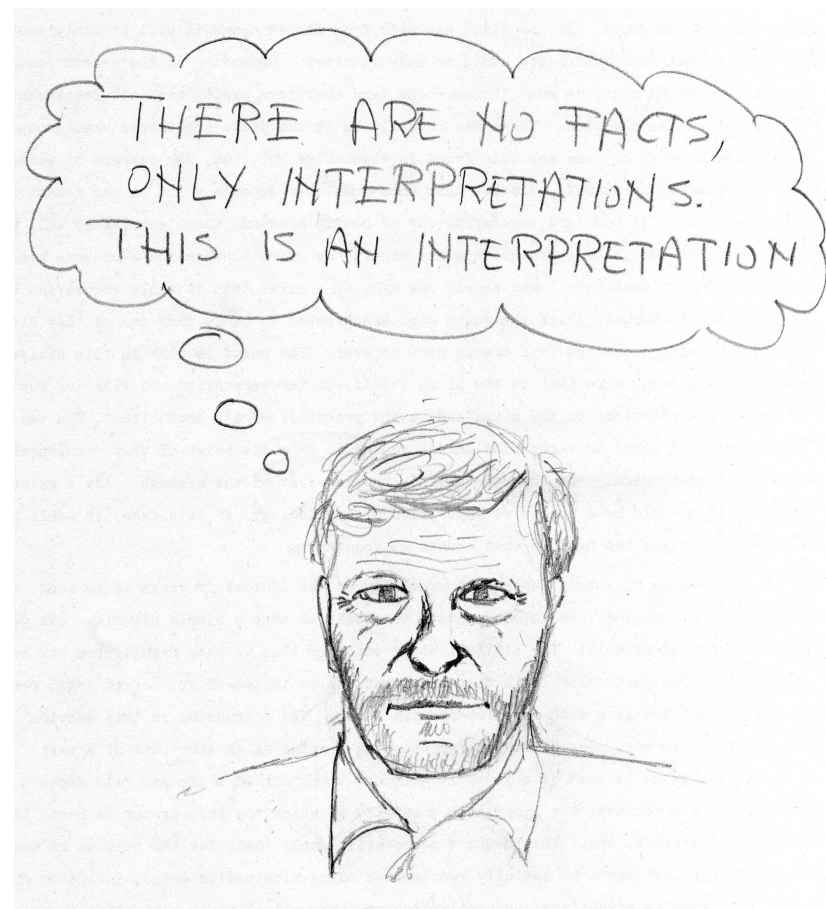
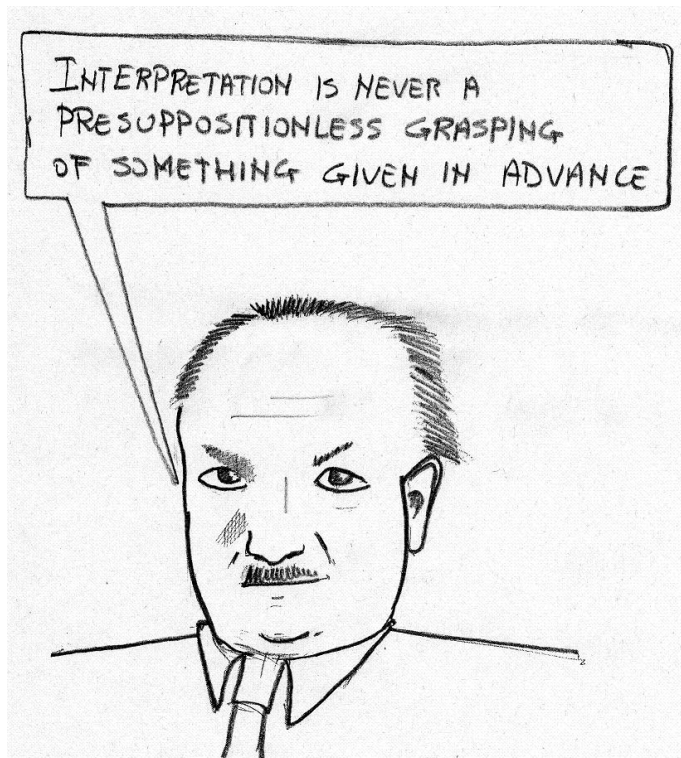
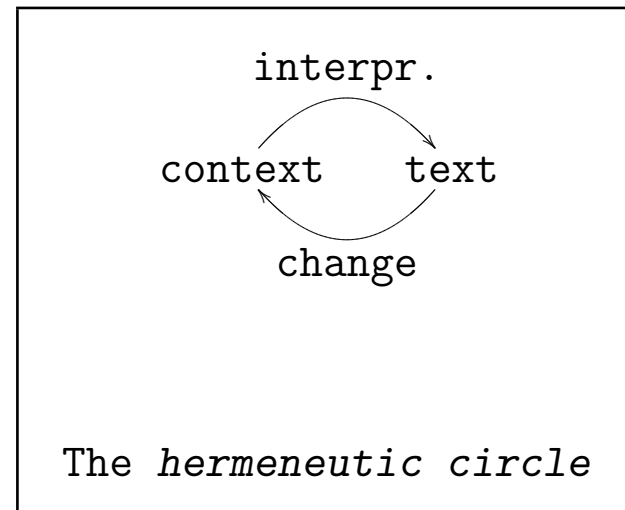


Skopos theory: even the functional relation is not an isomorphism; the translator decides the function that the translated text must have in the receiving culture.



Reading is an active, dialectic process.

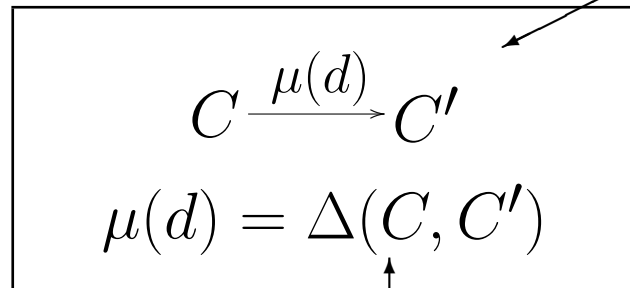
The context interacts with the text to refine interpretation.



CONTEXT BASED RETRIEVAL AND LANGUAGE GAMES

The meaning of a document is the change that it causes in the context of its interpreter:

Dialectic: the interpreter interacts with the text



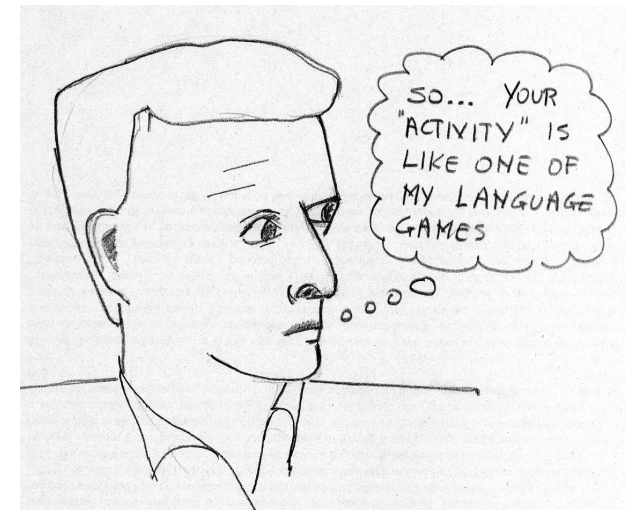
depends on the hermeneutic act

Ontology: $\perp \xrightarrow{\mu} C$
independent of the context of the interpreter.

Question: is it possible to formalize meaning? Only to the extent that it is possible to formalize context. In some cases it might be, in some cases it mightn't

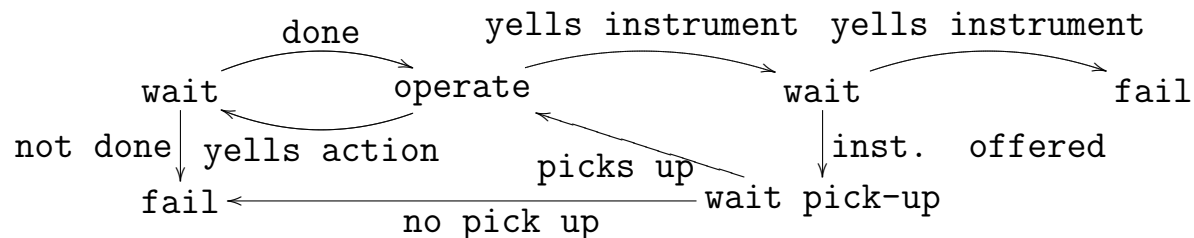
A query is a part of a linguistic activity, a **game**. In this game there are **moves** that change the context; the purpose of the game is to reach a certain context.

A query is a **move type** in a context linguistic game.

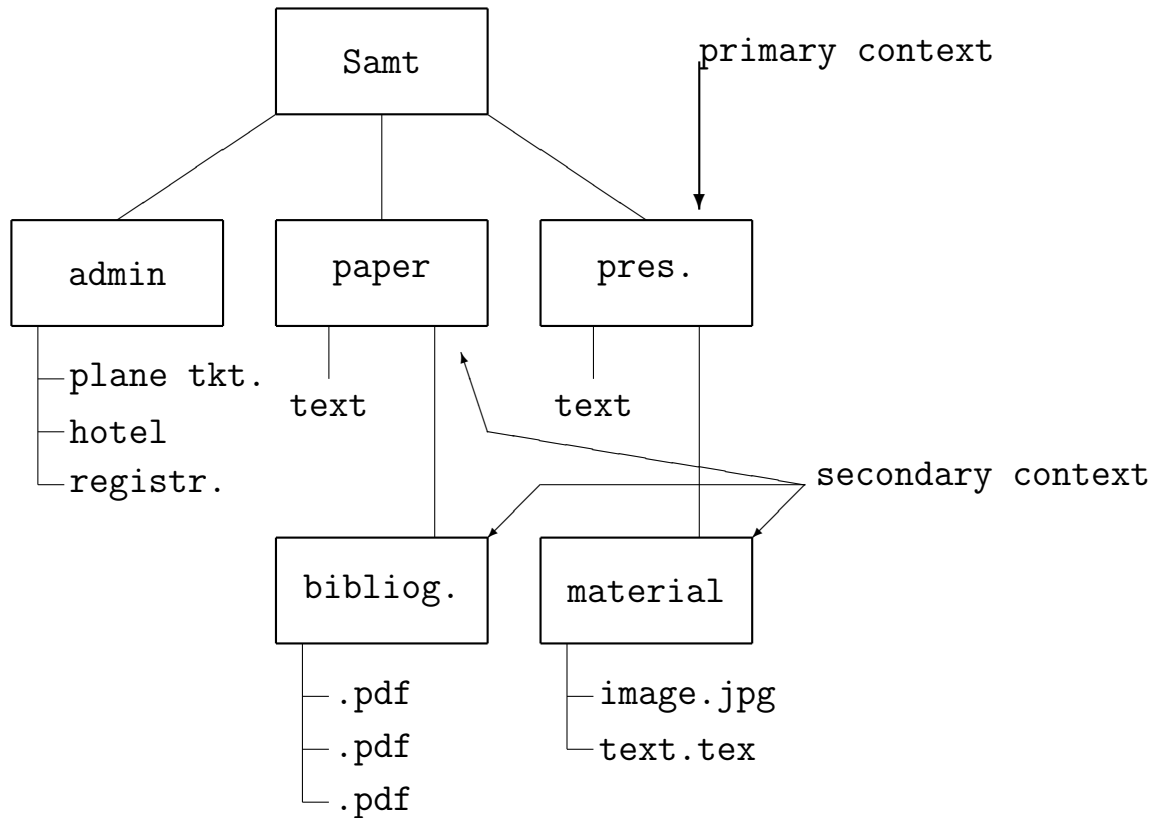


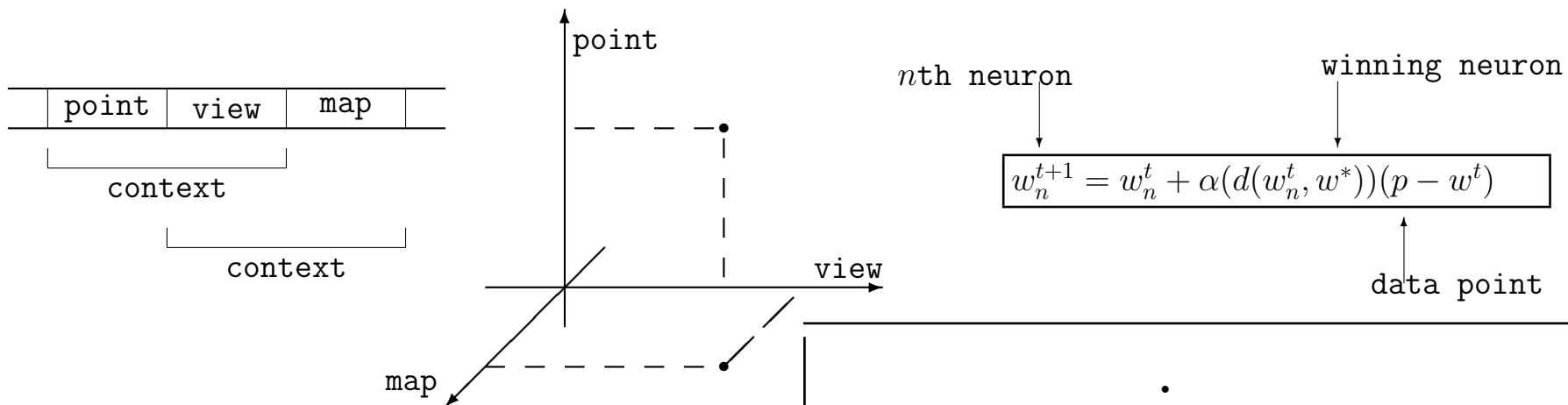
Many of these games are played as part of activities like writing a paper, preparing a presentation, looking for a place to go on a friday, etc. In some of these cases the context can be formalized and one can use it to do semantically meaningful query moves. In other cases this is not possible.

Example: the game that goes on between a surgeon and her assistants. Two moves: yelling the name of an instrument (scalpel!) or to name an action (vacuum).



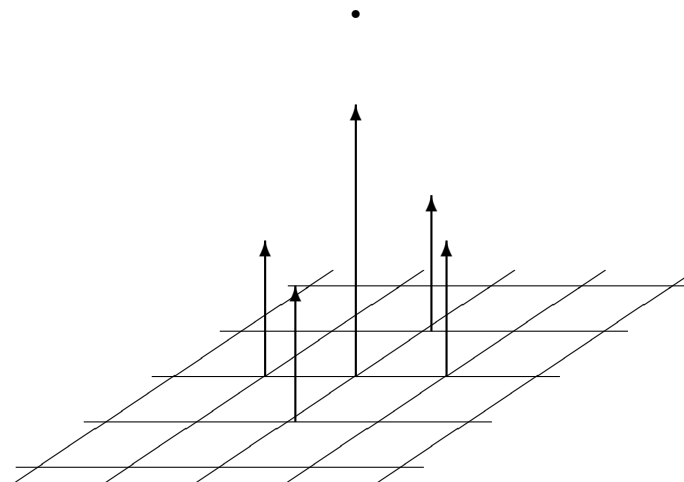
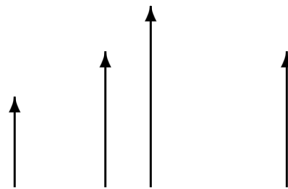
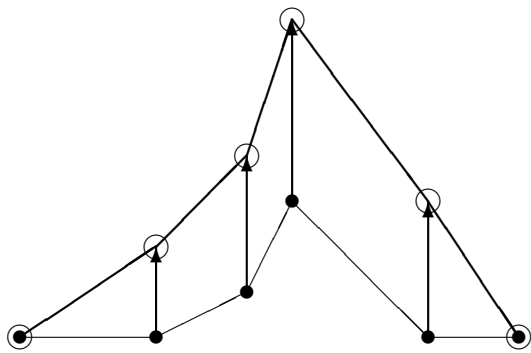
for activities whose context is in a computer, that is where we take the context! The context is given by all the documents in the folder in which one is working, the document in the folders around it, and the relations between these folders.



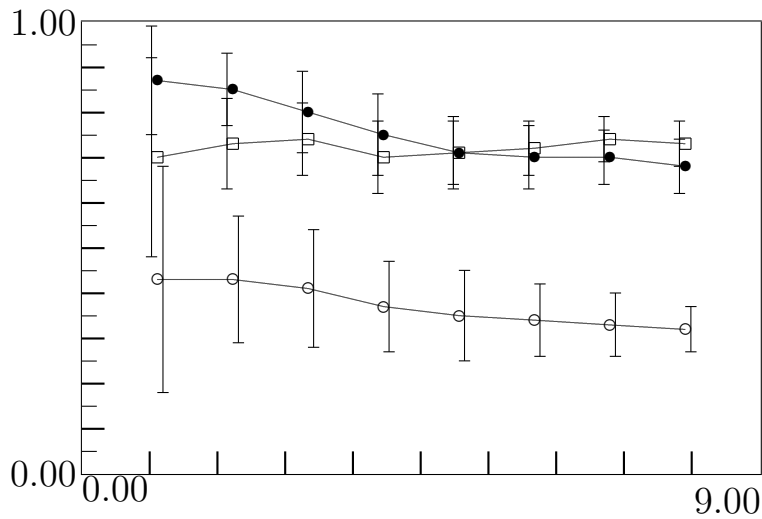


Q : query point

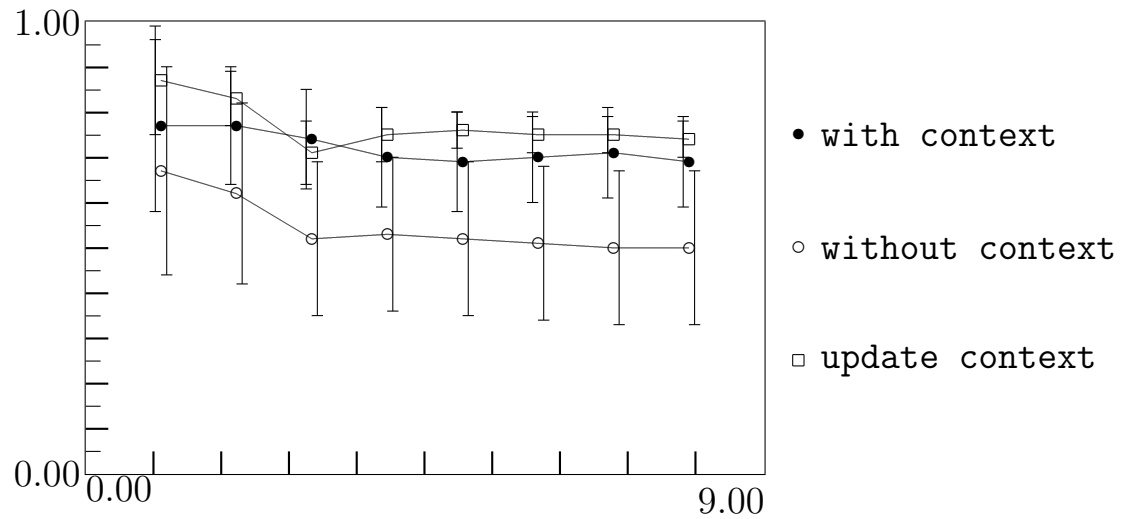
$$w_n^Q = w_n + \beta(d(w_n, w^*))(Q - w_n) \quad \beta \gg \alpha$$

$$\Delta w_n^Q = w_n^Q - w_n = \beta(Q - w_n)$$


Results obtained using the *google*(10^{100}) search engine and making brutal approximations: since google does not accept weights, replicate the words proportionally to their weight to approximate it.



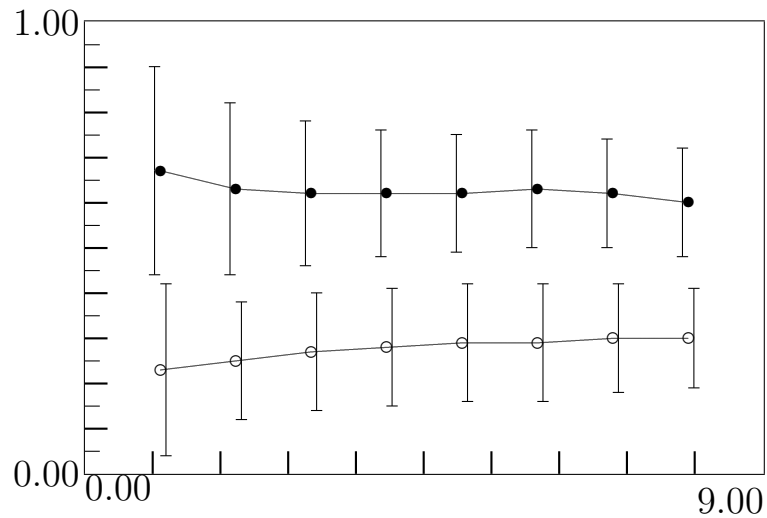
Context: preparation of a technical article on data integration.



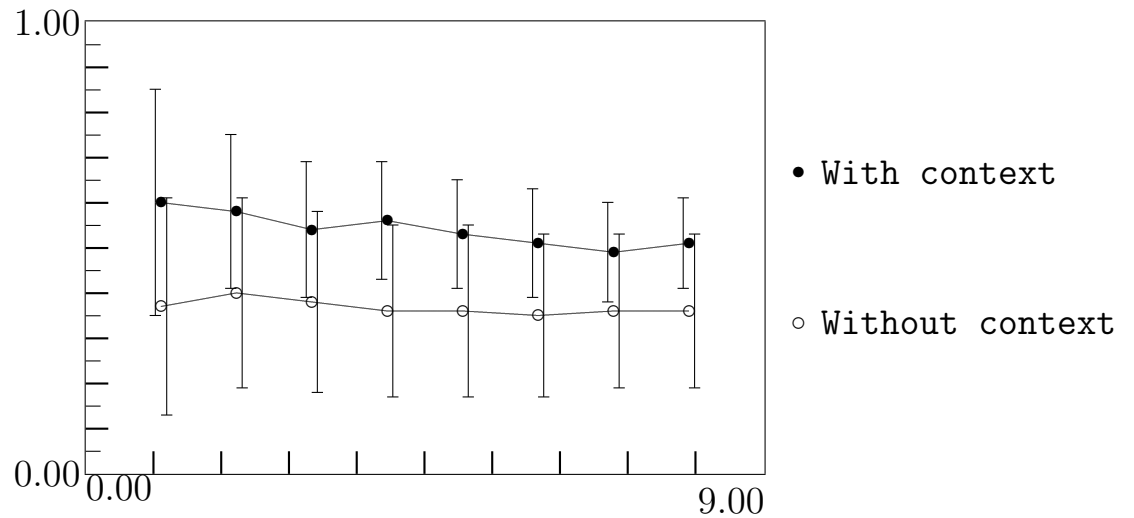
Context: preparation of a collection of neurophysiology papers.

Abscissa: number n of results considered; ordinate: precision of the first n results. Average and variance over 30 queries.

Image retrieval results obtained using the *google images* search engine and making the same brutal approximations as before. Search and context were text-based.



Context: preparation of a technical article on data integration.



Context: preparation of a collection of neurophysiology papers.

Abscissa: number n of results considered; ordinate: precision of the first n results. Average and variance over 30 queries.

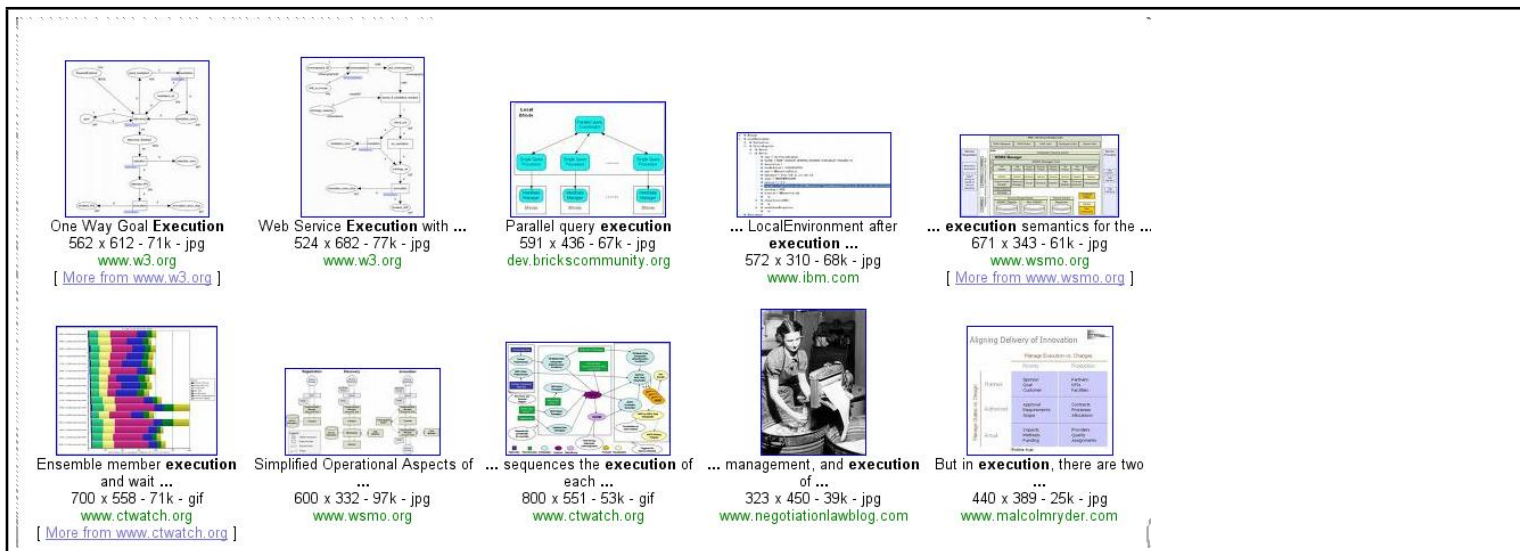
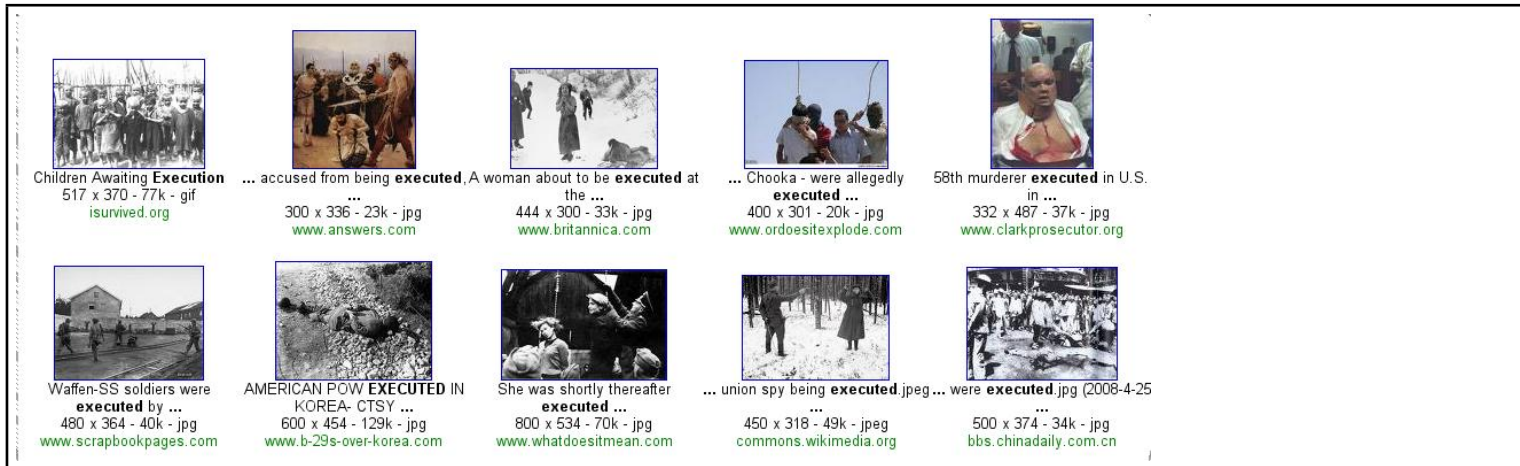


Image Results for the query *executed* (!) in the *computing* context.

CONCLUSIONS