

The meaning of semantics in broadcast : from words to ontologies

Bruno Bachimont
Université de technologie de Compiègne
Institut National de l'Audiovisuel

Semantics : why ?

- Issue:
 - Interoperability : exchanging data in complex processes;
 - Autonomy of pieces of data : even when they are isolated from their production context, we want to know who they are and what they mean;
- Problem:
 - Heterogeneity ;
- Approach:
 - Escaping from heterogeneity by abstracting data to a common level that captures data semantics, that is, what they really are and mean.

Semantics : what is it ?

- The meaning of a « signifier » ; several points of view:
 - Linguistic approach:
 - Meaning consists of words reformulating the signifier;
 - Psychological approach:
 - Meaning consists of a mental representation, the image one has of the signifier;
 - Ontological / metaphysical approach:
 - Meaning consists of a « real » entity to which the signifier corresponds;
 - Pragmatic approach:
 - Meaning consists of the usage one has of the signifier, or the actions it entails.
 - Operational approach:
 - Meaning consists of the actions and manipulation done with the signifier
 - Logical approach:
 - Meaning consists of the logical inferences one draws from the signifier
 - Etc.

In practice...

- Using semantics amounts to add some words, preferably expressed in english, uppercase, written in latin alphabet.



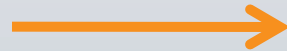
→ DOG

The situation

- The problem of using semantics is not solved but moved;
- The problem of characterizing the semantics of things considered in a given context is converted into the problem of making explicit semantics of the words used to expressed those things

In short...

Usual understanding of semantics
in AV communities



Meaning ?



DOG



Meaning ?



<biology:naturalcategory: dog>

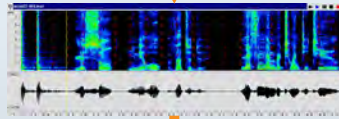
Meaning modelling
in knowledge engineering
and philosophy :
ontologies, formal
semantics, etc.

An opposition...

- In AV communities and Broadcast world:
 - A semantic approach is adding word to non textual objects
- In knowledge – concerned communities :
 - A semantic approach consists in formalising or modeling word / thing meaning.

Explanation: the unreasonable efficiency of language

Non-textual / verbal data



Semantically opaque :
data cannot describe itself

Textual data

dog

Semantically
transparent: a
word seems to be
autoexplicative

But...

- Unfriendly phenomena:
 - Homonymy: you may mean different things while telling the same thing;
 - Synonymy: different expressions may mean the same thing ;
 - Contradiction: opposite inferences may be drawn from the same expressions

Consequences

- Two main objectives :
 - Controlling what can be said or done while interpreting a word or label;
 - Having as many words and forms of expression at our disposal to express something ;
- Problem:
 - These objectives are contradictory :
 - Allowing many forms of expression forbids controlling ;
 - Controlling expression and their meaning forbids large expressivity

What is controlling meaning :

- Defining the authorised syntax and lexicon (what can be expressed) since labels (lexicon) and rules (syntax) have a predefined and precise meaning.
 - No ambiguity when interpreting expressions.
 - Approach adopted by semantic Web community : RDF, OWL, etc.
 - But need for formal languages and models :

$$\begin{aligned}\top^{\mathcal{I}} &= \Delta^{\mathcal{I}} \\ \perp^{\mathcal{I}} &= \emptyset \\ (\neg A)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus A^{\mathcal{I}} \\ (C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\ (\forall R.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \forall b. (a, b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\} \\ (\exists R.\top)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \exists b. (a, b) \in R^{\mathcal{I}}\}.\end{aligned}$$

Expressivity : what is it?

- Users are supposed to belong to a linguistic or specialised community :
 - They are allowed to use any terms since they are the best persons to know in context what is the required label, better than experts defining a priori some lists of authority.

Strategies

Semantic labels

Control

- Thesaurus
- Ontologies
- Semantic web

NLP to make labels
more robust
regarding
variability of
expression

Expressivity

- Folksonomy
- Social tagging
- Social interaction

Statistical filtering
to control
dispersion

GRAAL :

**Robust and meaningful data,
intelligible par man and machines**

The different representations of semantics

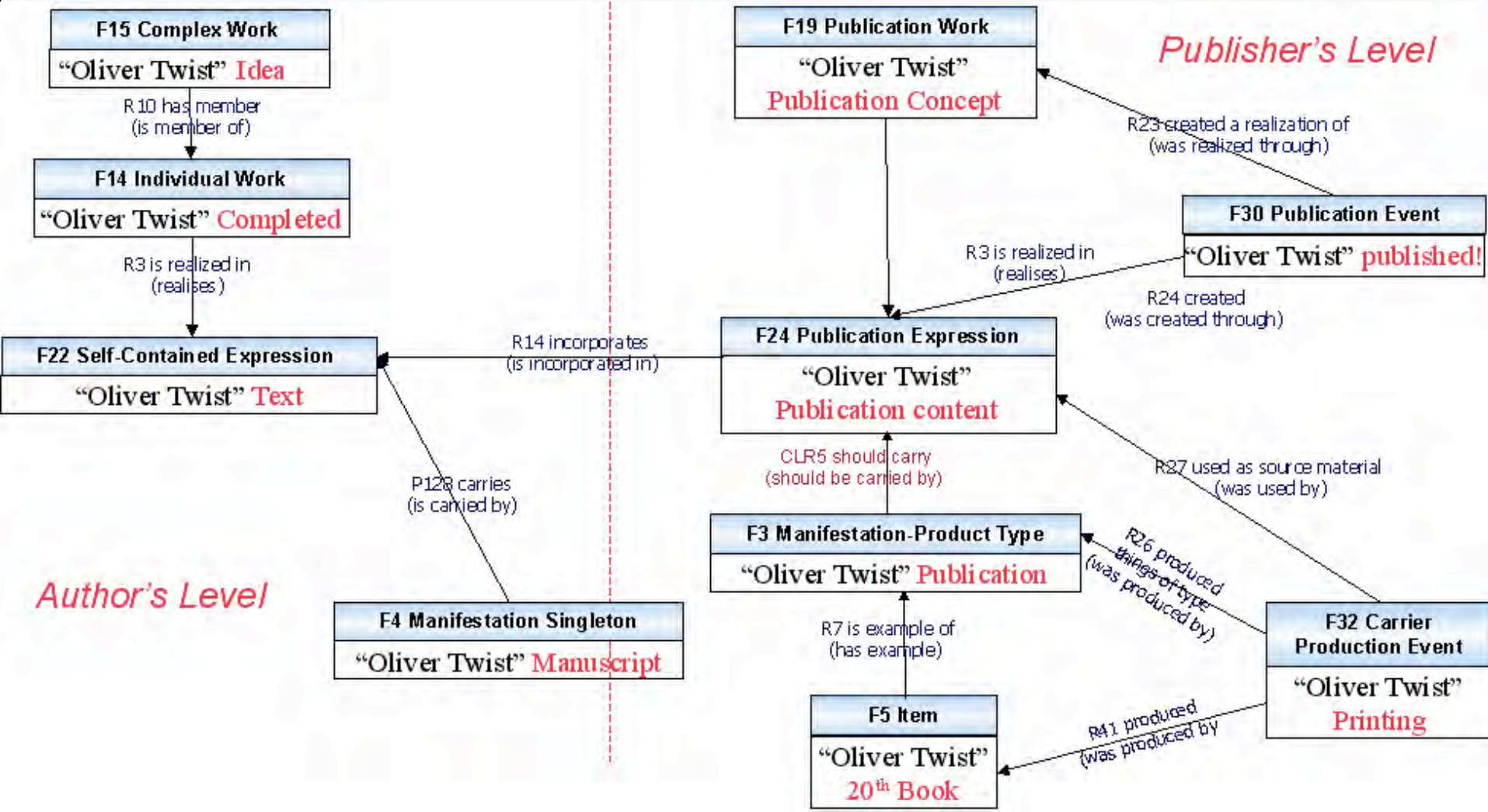
- Words
- Aggregated words:
 - Lexicon / dictionary
 - Thesaurus
- Modelised words :
 - Concepts;
 - Ontologies
- Elicited words
 - Social tagging
 - Folksonomies.

In the Audiovisual World

- Many controlled vocabularies:
 - SMPTE dictionary
 - Proprietary thesaurus : INA's one
 - P-Meta
- No ontologies specific to audiovisual, but:
 - CIDOC – CRM
 - FRBR, FRBRoo

Which are close to audiovisual preoccupations.
- Social tagging becomes more popular among TV or Archive actors.

FRBRoo



Publishing with users

Les sites de l'Ina pour : ▶ les étudiants, les universitaires et les professionnels | English

ina.fr **meotic** Et si vous lui donniez rendez-vous demain soir ?

je suis je recherche Inscription GRATUITE

Se connecter | S'inscrire | Accéder à la BOUTIQUE | 0 article

CHAÎNES | FOCUS | PLAYLIST | JEUX & COMMUNAUTÉ | **ina.PUB**

Nouveautés | Sport | Politique | Art et Culture | Publicité | Divertissement | Fictions et animations | Sciences et techniques | Histoire et conflits | Economie et société

PLAYLIST > SCIENCES ET TECHNIQUES

Trier par : date | vues

05/09/2009  **INTERNET**
Eagle1

05/09/2009  **INTERNET, L'ANTICIPATION**
Eagle1

04/09/2009  **VE-TCHERNOBYL**
slesage

Rechercher sur Ina.fr Vidéo OK

VIDÉOS LES PLUS vues



Indexing with users



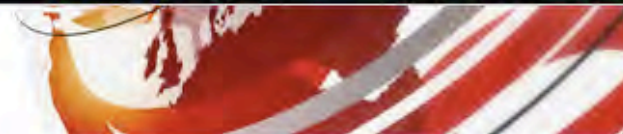
Low graphics Help

Search

Explore the BBC

NEWS

▶ Watch ONE-MINUTE WORLD NEWS



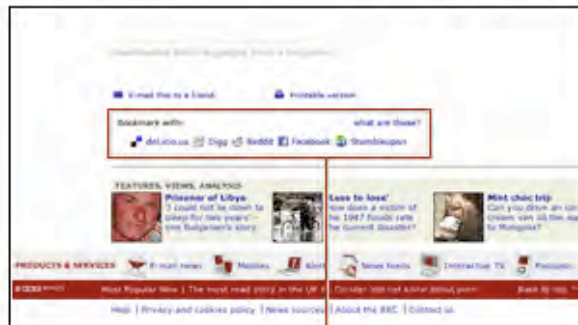
News Front Page



- Africa
- Americas
- Asia-Pacific
- Europe
- Middle East
- South Asia
- UK
- Business
- Health
- Science & Environment
- Technology
- Entertainment
- Also in the news
- Video and Audio
- Programmes
- Have Your Say
- In Pictures
- Country Profiles
- Special Reports

Social bookmarking links

Every story page on the News and Sport websites now includes links to a range of social bookmarking / sharing websites.



Bookmark with: [what are these?](#)

[del.icio.us](#) [Digg](#) [Reddit](#) [Facebook](#) [Stumbleupon](#)

Social bookmarking links can be found at the bottom of all story pages

These sites allow you to store, tag and share links across the internet. You can share these links both with friends and people with similar interests. You can also access your links from any computer you happen to be using.

PRODUCTS AND SERVICES



Podcasts

Full programmes and highlights to download



E-mail news

Sign up to a range of newsletters

▶ E-mail FAQs



Alerts

Download the news ticker and alerts service

▶ Alerts help



Mobiles

Get the top stories, wherever you are



News feeds

Let the news come to you

Conclusion

- Be careful with arrows, boxes and letters;
- Writing a word in natural language does not amount to semantic modelling;
- Issues:
 - defining who/what/when should handle semantics:
 - The system: apriori controlling of data semantics;
 - People : aposteriori filtering of meaningful and adated data produced by user;
 - What kind of semantics we need :
 - Compliance with a reference (even if nobody understands it, systems can)
 - Intepretation by people
 - Inference or calculus by machines
 - ...