# An ontological approach to model and query multimodal concurrent linguistic annotations

**J. Seinturier, E. Murisasco, E. Bruno, P. Blache**

LSIS UMR CNRS 6168, LPL UMR CNRS 6057
Université du Sud Toulon Var, Université de Provence
F-83957 La Garde, France, 13604 Aix en Provence, France
{seinturier, murisasco, bruno}@univ-tln.fr, blache@lpl-aix.fr

**Abstract**

This paper focuses on the representation and querying of knowledge-based multimodal data. This work stands in the OTIM project which aims at processing multimodal annotation of a large conversational French speech corpus. Within OTIM, we aim at providing linguists with a unique framework to encode and manipulate numerous linguistic domains (from prosody to gesture). Linguists commonly use Typed Feature Structures (TFS) to provide an uniform view of multimodal annotations but such a representation cannot be used within an applicative framework. Moreover TFS expressibility is limited to hierarchical and constituency relations and does not suit to any linguistic domain that needs for example to represent temporal relations. To overcome these limits, we propose an ontological approach based on Description logics (DL) for the description of linguistic knowledge and we provide an applicative framework based on OWL DL (Ontology Web Language) and the query language SPARQL.

**Keywords:** Information systems,Knowledge engeeniring, OWL

## 1. Introduction

This work stands in the OTIM (Tools for Multimodal Annotation processing) project [1]. It aims at developing conventions and tools for multimodal annotation of a large conversational French speech corpus (Blache et al., 2010) (`http://aune.lpl.univ-aix.fr/~otim/`). OTIM can be summarized in two main steps.

The first step concerns *the multimodal annotation of a conversational speech between two persons*. It is under the responsibility of linguists; annotation is done according to different levels of linguistic analysis. Each expert has to annotate the same data flow according to its knowledge domain and the nature of the signal on which he annotates (signal transcription or signal). Experts generally use dedicated tools like PRAAT[2], ANVIL[3] or ELAN[4]. The qualifier multimodal is due to the nature of the studied corpus which is composed of text, sound, video. Within the project OTIM, linguists propose an encoding for annotating spoken language data, with the acoustic signal as well as its orthographic transcription. They have chosen to use Typed Feature Structures (Carpenter, 1992)(Copestake, 2003) (TFS) to represent in an unified view the knowledge and the information they need for annotation. Linguistic annotation tools rely on native and not often open formats which are not directly interoperable. TFS provides an abstract description using a high level formalism independent from coding languages and tools.

The second step concerns *the representation and manipulation of multimodal annotation*. We aim at providing linguists with a unique framework to encode and manipulate numerous linguistic domains (morpho-syntax, prosody, phonetics, disfluencies, discourse, gesture and posture (Blache et al., 2010)) in order to analyze and find correlations between annotated linguistic domains. For that, it has to be possible to bring together and align all the different annotations associated to a corpus. It implies the definition of a formal model for describing and manipulating them in a concurrent way. The main difficulty in defining a data model comes from the heterogeneity and the distribution of the resources. Concurrent manipulation consists in querying annotations belonging to two or more modalities and/or in querying the relationships between modalities. For instance, we need to be able to express queries over gestures and intonation contours (what kind of intonation contour does the speaker use when he looks at the listener?) and to query temporal relationships (in terms of anticipation, synchronization or delay) between both gesture strokes and lexical affiliates.

In this paper, we focus on this last step considering semantic web technologies for the development of a Knowledge-based Information System.

## 2. Context and Motivation

Linguistic knowledge is captured by means of three types of information : *properties* (the set of characteristics of an object); *relations* (the set of relations that an object has with other objects); *constituents* (complex objects composed of other objects). TFS proposes a formal presentation of each annotation in terms of feature structures and type hierarchies : properties are encoded by features, constituency is implemented with complex features, and relations make use feature structure indexing; each linguistic domain is represented as a hierarchical model. TFS enables linguists to represent in an unified view the knowledge and the information they need for annotation.

Figure **??** graphically describes TFS representation of the prosodic domain; a formal definition can be found in (Car-

[1] supported by the French ANR agency
[2] http://www.fon.hum.uva.nl/praat/
[3] http://www.anvil-software.de/
[4] http://www.lat-mpi.eu/tools/elan/

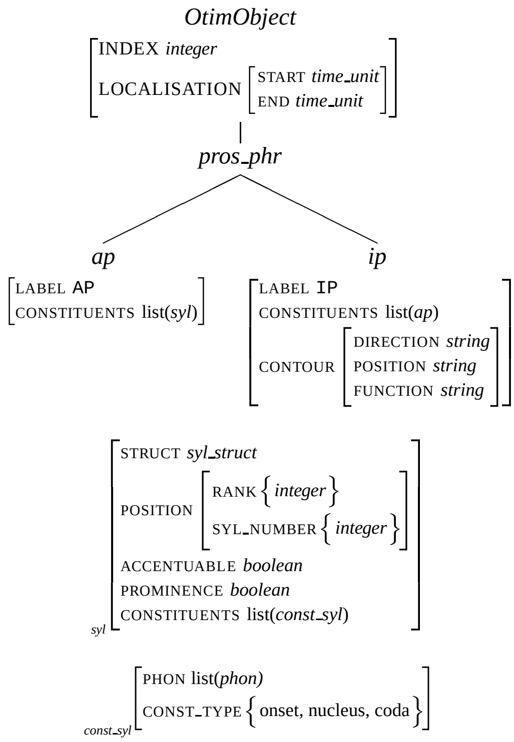


Figure 1: TFS representation of the prosodic domain

penter, 1992) (Copestake, 2003). For sake of simplicity, we do not detail the meaning of every feature used in the example.
Due to its theoretical nature, TFS representation cannot be used within an applicative framework and has to be implemented into other formalisms. Besides, TFS expressivity is limited, for example for temporal relations bacause object anchoring is absolute. Moreover, when linguists need to annotate coreferences or disfluences which are organized around objects, it would be useful to have an object anchoring which is conflicting with the underlying model of TFS which is a Directed Acyclic Graph (DAG).

## 3. Contributions

We propose an knowledge representation formalism which be an alternative to TFS : an ontological approach based on Description Logics (Baader et al., 2003) (DL) and on semantic web technologies for the development of a linguistic Knowledge-based Information System.
Ontologies enable experts to share and annotate information in their respective knowledge domain both represent semantic descriptions of linguistic domains and data. We have proved in (Seinturier et al., 2011) the theoretical correspondence of TFS and DL (both enable to represent DAG). From this result, our objetives are:

- the definition of a linguistic ontology from the TFS provided by linguists
- the definition of an applicative framework by means of semantic web proposals such as OWL-DL[5]) for the representation of this ontology and SPARQL[6] the querying language of semantic web for its manipulation.
- the demonstration of our first promising results for the exploitation of multimodal linguistic annotations

Some linguistic projects have a similar objective than OTIM, for instance NITE[7], AGTK[8], PAULA[9], XStandoff (Sthrenberg and Jettka, 2009). Our approach differs from them because we focus on an ontological contribution. These proposals generally propose toolkits for multilevel annotation by means of libraries of data and annotation management. Moreover, linguistic annotation tools rely on native and not often open formats which are not directly interoperable. The multiplication of annotation schemes and coding formats is a severe limitation for interoperability. One solution consists in developing higher level approaches (Ide and Suderman, 2007)(Sthrenberg and Jettka, 2009). However, these experiments still remain very programmatic.

### 3.1. Creating OWL ontology

Creation of the OWL ontology follows two steps. First of all, the terminological knowledge from the TFS is implemented into OWL using the Protege[10] ontology editor. The Protege framework was initially designed for biologists and biochemists. This characteristic is quite interesting because this is not a computer scientist tool and so there is no need of a specific knowledge in computer science to use it.
The user interface relies on a graphical and textual description of the concepts, relations and individuals. Within the OTIM project, the ontology has been hand maded using Protege instead of processing TFS. This choice comes from the fact that we use the OWL-DL expressiveness to integrate descriptions that was impossible to represent (for example time relations or cyclic references). At this time, a complete ontology including prosody, phonetics and lexical domains terminology is available. Figure 2 shows the ontology of the prosodic domain. This ontology is linked with two other domains: the phonetics domain, which is a part of the OTIM knowledge representation framework, and the time domain given by a standard ontology of the W3C.

### 3.2. Managing data and querying with SPARQL

Management and querying of OWL data relies on the standard SPARQL (Prud'hommeaux and Seaborne, 2007) querying language. SPARQL enables to match graph pattern against the graph of RDF/OWL triple ($WHERE$ clause) and identifies values to be returned ($SELECT$ clause). The $FROM$ clause enables to identify the data sources to query. The $FILTER$ clause add constraints to the matching pattern and give more filtering capabilities. By convention, variables are marked with a $'?'$. By default the graph pattern is a conjunction of triple. Each triple (*subject*, *predicate*, *object*) represents a piece of knowledge and means the *subject* has a *predicate* with *object* as value.

[5]http://www.w3.org/TR/2004/REC-owl-features-20040210/
[6]http://www.w3.org/TR/rdf-sparql-query/

[7]http://groups.inf.ed.ac.uk/nxt/
[8]http://weblex.ens-lsh.fr/projects/xitools/logiciels/AGTK/agtk.htm
[9]http://www.sfb632.uni-potsdam.de/ d1/paula/doc/
[10]http://protege.stanford.edu/

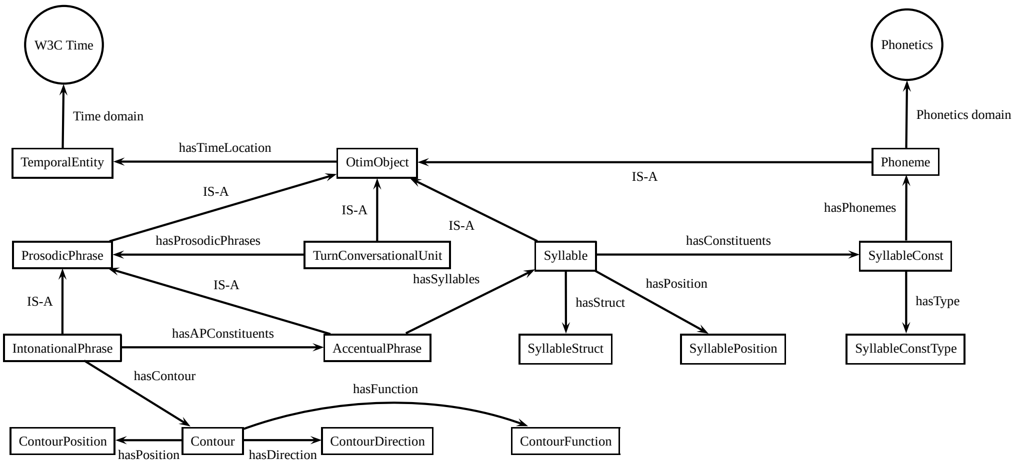


Figure 2: Ontological representation of the prosodic domain

We express within the OTIM project the linguistic inter domain queries designed on TFS by SPARQL queries on the OWL representation. A sample query expressed in natural language is:
*"We need the list of phonemes that are associated with the accentual phrases stated between the second 35 and the second 55 of the speech."*
This query takes into account the prosodic domain (accentual phrase), the phonetic domain (phoneme) and the time. Such a query is represented in SPARQL by:

```
1.  SELECT    ?phoneme
2.  FROM      otim − prosody.owl, otim − phonetics.owl
3.  WHERE {   ?const rdf:type prosody:SyllableConst
4.          . ?const hasPhonemes ?phoneme
5.          . ?syl rdf:type prosody:Syllable
6.          . ?sc hasConstituents ?const
7.          . ?ap rdf:type prosody:AccentualPhrase
8.          . ?ap hasSyllables ?syl
9.          . ?t rdf:type time:TemporalEntity
10.         . ?ap hasTimeLocation ?t
11.         . ?tref time:contains ?t }
```

We assume ithat the time bounds given are represented as a $TemporalEntity$ named $tref$. The $SELECT$ clause specifies that the result to build is made of phonemes. The clause $FROM$ contains the two data sources on which the query is processed (the two target domains prosody and phonetics). The $WHERE$ clause describes the patterns for a phoneme to match. The $WHERE$ clause is a logical conjunction (symbolized by .) of $9$ triples. The first $6$ triples (lines $3$ to $8$) describe the structure of the data and how to get a phoneme list from an accentual phrase. The last $3$ triples (line $9$ to $11$) describe what are the selected accentual phrases regarding the time criterion. The relation $contains$ applied to the variables $t$ and $tref$ represents the $contains$ relation of the Allen Algebra (Allen, 1991) which is implemented within the W3C time ontology.

## 4. Implementation and results

The OTIM framework for linguistic multimodal annotations management has been implanted within a Java/OWL framework. The OWL standard used is OWL-DL as this is the specification that gives all the expressiveness we need and guarantees some calculability results that are critical for querying data. The Java framework is based on two packages:

- A specific OTIM package that enables to deal with linguistic tools and data.
- The Jena[11] package that provides robust OWL capabilities as SPARQL querying and logic reasoning.

The OTIM package has been developed for interfacing with widely used linguistic tools and data repository (the tools that are the most used within the project are PRAAT and ANVIL). The Jena package is developed by the Open Jena project and provides advanced OWL processing methods that can be embedded within a Java application. Jena also provides relational mapping of OWL data that makes optimal SPARQL queries by translating them into relational queries. These characteristics guarantee that the use of the developed Java/OWL is efficient.

## 5. References

J. Allen. 1991. Time and time again : The many ways to represent time. *International Journal of Intelligent Systems*, 6(4):341–355, july.

Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. 2003. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

Philippe Blache, Roxane Bertrand, Brigitte Bigi, Emmanuel Bruno, E Cela, Robert Esperrer, Gaelle Ferre, Marion Guardiola, Daniel Hirst, Ep Magro, Martin JC, Meunier C, Morel Ma, Elisabeth Murisasco, Nesterenko I, Nocera P, Pallaud B, Prevot Laurent, Priego-Valverde J, Julien Seinturier, Tan N, Tellier Marion, and Rauzy Stephane. 2010. Multimodal annotation of conversational data. In *Proceedings of the fourth linguistic annotation workwhop (LAW)*, pages 186–191. Assocation for computational Lunguistics (ACL), 15 july.

Robert L. Carpenter. 1992. *The Logic of Typed Feature Structures*, volume 32 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, The

[11]http://openjena.org/

Edinburgh Building, Shaftesbury Road, Cambridge CB2 8RU, United Kingdom, c. j. van rijsbergen edition.

Ann Copestake, 2003. *Collaborative Language Engineering: A Case Study in Efficient Grammar-based Processing*, chapter Definitions of Typed Feature Structures. CSLI Publications, Ventura Hall, Stanford University, Stanford, CA 94305-4115.

N. Ide and K. Suderman. 2007. Graf: A graph-based format for linguistic annotations. In *Linguistic annotation workshop (LAW)*. Association for computational Linguistics (ACL).

Eric Prud'hommeaux and Andy Seaborne. 2007. Sparql query language for rdf (working draft). Technical report, W3C, March.

J. Seinturier, H. Rouine, E. Murisasco, and E. Bruno andP. Blache. 2011. Knowledge-based multimodal data represnetation and querying. In *Int. conf. on Knowledge Engineering and Ontology Developmement (KEOD 2011)*, Paris, France, October.

M. Sthrenberg and D. Jettka. 2009. A toolkit for multi-dimensional markup - the development of sgf to xstandoff. In *Proceedings of Balisage: The Markup Conference 2009*. Assocation for computational Lunguistics (ACL), Balisage Series on Markup Technologies, vol. 3.