

Towards a FrameNet Resource for the Legal Domain

Giulia Venturi¹, Alessandro Lenci¹,
Simonetta Montemagni¹, Eva Maria Vecchi¹,
Maria Teresa Sagri², Daniela Tiscornia², and Tommaso Agnoloni²

¹Institute of Computational Linguistics, CNR, Pisa, Italy

²Institute of Legal Information Theory and Techniques, CNR, Firenze, Italy

{giulia.venturi, alessandro.lenci,
simonetta.montemagni, evamaria.vecchi}@ilc.cnr.it
{mt.sagri, daniela.tiscornia, tommaso.agnoloni}@ittig.cnr.it

Abstract. In the AI&Law community, the importance of frame-based ontologies has been acknowledged since the early 90's with the Van Kralingen's proposal of a *frame language* for legal knowledge representation. This still appears to be a strongly felt need within the community. In this paper, we propose to face this need by developing a FrameNet resource for the legal domain based on Fillmore's *Frame Semantics*, whose final outcome will include a frame-based lexical ontology and a legal corpus annotated with *frame* information. In particular, the paper focuses on methodological and design issues, ranging from the customization and extension of the general FrameNet for the legal domain to the linking of the developed resource with already existing Legal Ontologies.

Key words: Frame Semantics, Legal Ontologies, Knowledge Representation, Corpus Annotation

1 Introduction

The last few years have seen a growing body of research and practice in the field of Artificial Intelligence and Law (AI&Law) for what concerns the construction of legal ontologies and their application to the law domain. The importance of this research area is testified by the different Workshops and Conferences which have been organized around this topic. However, as [1] points out, existing legal ontologies vary significantly, for what concerns their underlying structure and organization, the way they are constructed (either top-down or bottom-up) and how they are exploited in different applications. In this paper, we will focus on a particular type of ontology, the so-called *lightweight or lexical ontologies* [2], whose main feature consists in bridging the gap between the legal knowledge formalized in domain ontologies on the one hand and the legislative texts on the other hand; this follows from the fact that in this type of ontology legal concepts are paired with their lexical realizations. This feature makes this type of ontology particularly suitable for use in Information Extraction and Semantic

Tagging tasks. Note that these ontologies are typically bootstrapped from legal texts (either manually or through ontology learning techniques).

The most notable example of this type of ontology in the legal domain is represented by the JurWordNet ontology-driven semantic lexicon [3], together with its multilingual extension LOIS [4]. Both JurWordNet and LOIS have been developed following the WordNet (hereafter referred to as WN) design, where words expressing legal concepts such as ‘liability’, ‘sanction’, ‘violation’ are organized in *synsets* (i.e. sets of synonyms) in turn linked by hierarchical or taxonomical relations such as hyponymy and hyperonymy. Under this view, the meaning of a word is intended as a distinct, atomic semantic object, fully identified by its position in the general semantic network.

However, the taxonomical organization of legal concepts is not the only possible one. Legal experts claim that, despite their utility, WN-like resources are not completely adequate and satisfactory in order to represent events and situations typically expressed in legal documents: this is a consequence of the WN-model [5] they follow. Interestingly enough, this claim is in line with the Van Kralingen’s proposal of a *frame language* as a plausible method for the conceptual representation of legal knowledge [6]; in spite of the fact that this proposal dates back to the early ’90s, it still represents a need commonly felt in the AI&Law community.

In this paper we propose to face this need by developing a lexical resource based on Fillmore’s *Frame Semantics* [7] and on the organization principles underlying the FrameNet project [8] (hereafter referred to as FN)¹. In particular, we propose to build a FN-like resource specialized for the legal domain, by extending and refining the general purpose FN resource. By proceeding in this way, it will be possible to overtly represent the inner structure of complex situations in terms of their participants, e.g. “under which *Circumstances*, which *State of affairs* is sanctioned by which *Principle*”.

2 Starting points

In order to create a frame-based resource for the legal domain, our idea is to combine two different approaches from two different research communities, i.e. AI&Law and Computational Linguistics. In particular, we aim at revisiting Van Kralingen’s proposal of a *frame language* [6] for legal knowledge representation in the light of Fillmore’s *Frame Semantics* theory [7].

2.1 Frame-based Legal Ontologies

Amongst the bulk of Legal Ontologies built so far (see [1] for a state-of-the-art), the Van Kralingen and Visser studies are the only ones which envisage a frame-based ontology of law. In their collaborative project, Van Kralingen has defined a theoretical model (i.e. a *conceptual* ontology) and Visser has formalized it in

¹ <http://framenet.icsi.berkeley.edu>

an ontology [9]. The proposed *frame language* is based on the concept of a *norm* and of an *act* as legal conceptual primitives of the legal domain which can be conceived as *frames*, i.e. *data-structures for representing a stereotyped situation in which each element is represented*. Thus, the focus is on the inner structure of a *norm* and of a *legal act*, i.e. on what their building elements are. As shown in Table 1, a *norm frame* is defined as a template in which each element of a norm is represented as a slot of the norm frame. Since every legal action has many different aspects, a *legal act* has also been conceived as a frame. As shown in Table 2, each aspect of an action is represented as a slot of the *act frame* as well.

Element	Description
Norm identifier	The norm identifier (used as a point of reference for the norm).
Norm type	The norm type (norm of conduct or norm of competence).
Promulgation	The promulgation (the source of the norm).
Scope	The scope (the range of application of the norm).
Conditions of application	The conditions of application (the circumstances under which a norm is applicable).
Subject	The norm subject (the person or persons to whom the norm is addressed).
Legal modality	The legal modality (ought, ought not, may, or can).
Act identifier	The act identifier (used as a reference to a separate act description).

Table 1. A norm frame as defined in the Van Kralingen’s frame-based ontology [6]

2.2 FrameNet

The FN resource we started from is a lexical resource for English, based on *Frame Semantics* and supported by corpus-evidence. The goal of the FN project is to document the range of semantic and syntactic combinatory possibilities of each word in each of its senses. Typically, each sense of a word belongs to different Semantic Frame, conceived in [8] as “a script-like conceptual structure that describes a particular type of situation, object or event along with its participants and properties”. For example, the “Apply_heat” frame describes a common situation involving participants such as “Cook” and “Food”, etc. , called Frame Elements (FEs), and is evoked by Lexical Units (LUs) such *bake*, *blanch*, *boil*, *broil*, *brown*, *simmer*, etc. As shown by the following example, the frame-evoking LU can be a verb (bolded in the example) and its syntactic dependents (those written in subscript) are its FEs: [Matilde _{Cook} **fried** [the catfish _{Food}] [in a heavy iron skillet _{Heating_instrument}].

The type of representation produced by FN is a network of “situation-types” (frames) organized across inheritance relations between Frames, as opposed to

a network of meaning nodes, as in the case of WN. In FN, Frame Elements can be also specified with Semantic Types (i.e. ontological categories) employed to indicate the basic typing of fillers that are expected in the Frame Element. Most of these semantic types correspond directly to synset nodes of WN, and can be mapped onto already existing ontologies. FN currently contains more than 800 Frames, covering roughly 10,000 Lexical Units; these are supported by more than 135,000 FN-annotated example sentences.

Element	Description
Act identifier	The act identifier (used as a point of reference for the act).
Promulgation	The promulgation (the source of the act description).
Scope	The scope (the range of application of the act description).
Agent	The agent (an individual, a set of individuals, an aggregate or a conglomerate).
Act type	The act type. Both basic acts and acts specified elsewhere can be used.
Means	The modality of means (material objects used in the act or more specific descriptions of the act).
Manner	The modality of manner (the way in which the act has been performed).
Temporal aspects	The temporal aspects (an absolute time specification).
Spatial aspects	The spatial aspects (a specification of the location where the act takes place).
Circumstances	The circumstantial aspects (a description of the circumstances under which the act takes place).
Cause	The cause for the action (a specification of the reason(s) to perform an action).
Aim	The aim of an action (the goal visualized by the agent).
Intentionality	The intentionality of an action (the state of mind of the agent).
Final state	The final state (the results and consequences of an action).

Table 2. An act frame as defined in the Van Kralingen’s frame-based ontology [6]

3 Our approach

This section outlines our approach to the construction of a FN resource for the legal domain. Our eventual goal is to instantiate the Van Kralingen’s frame-based approach to the representation of legal knowledge by exploiting the FN model. While the Van Kralingen’s methodology is mostly based on domain-theoretical assumptions, we are rather planning to develop a corpus-based lexical-semantic resource which permits accounting for how complex events and situations are expressed within legal documents. The linguistic-empirical evidence provided by such a corpus-based methodology results in a bottom-up organization of legal knowledge.

As opposed to a WN-like resource, we think that a FN-like approach can be particularly suitable for the legal domain for a number of reasons. While in WN words are organized as hierarchies or taxonomies of synsets, according to FN principles word senses are related to each other only by way of their links to common background Frames.

Moreover, as Fellbaum noted in [5], “WordNet reflects the structure of frame semantics to a degree, but suggested that its organization by part of speech would preclude a full frame semantic approach”. In FN, on the other hand, the lexical units that evoke a frame are not restricted to a single part of speech. For example, the Frame “Process_end” is evoked by both verbs such as *to conclude*, nouns such as *end* and adjectives such as *final*. This is a very important FN feature when dealing with corpora of legal language. According to [10], it is very common in legal texts that events are expressed through nominal rather than verbal constructions. It follows that, for example, the Frame “Prohibiting” can be evoked both by the verb ‘to prohibit’ and by the deverbal noun ‘prohibition’.

To our knowledge, the only effort within the AI&Law community devoted to the use of FN is reported in [11]. As part of a layered approach to a legal domain representation, the authors exploit nine Semantic Frames selected from FN. Different Frame Elements from different Frames have been occasionally combined to represent the legal-domain knowledge contained in six judicial judgments of the Supreme Court of Justice of Portugal. They overtly argue for “a corpus-based methodology for an ontology construction that seeks the rigorous linguistic analysis aiming at formalization”. Yet, differently from our approach, they do not explicitly aim at creating a domain-specific FrameNet resource.

During the initial design phase, we have considered what has been done in other specialized domains as well. For example, within the bio-medical domain a domain-specific FN extension has been proposed in [12], who successfully developed a BioFrameNet through creating new Semantic Frames relevant to the domain of molecular biology and linking them to domain-specific biomedical ontologies. However, in the construction of such a FN resource for the bio-medical domain the authors faced bio-medical language peculiarities which pose challenges rather different from ours. As laid out in Section 4.1, the specific relationship between the ordinary and legal language (i.e. their closed intertwining) raises more challenging issues.

Following the underlying organization of the FN model, we intend to produce:

1. a legal corpus annotated with frame information,
2. a lexical frame-based resource covering the legal and domain terms occurring in the annotated corpus.

4 Design issues

A number of issues worth discussing has been encountered during the design stage of a FN extension and specialization for the legal domain. They mainly concern the choice of i) whether and to what extent the general FN Frames

should be customized for legal text annotation purposes, and ii) how to ontologically type the lexical fillers of Frame Elements for domain-specific purposes.

4.1 FrameNet customization strategies for a *legal FrameNet*

Following the approach laid out in Section 3, we plan to build a legal domain extension of the general FN on the basis of the already existing set of Semantic Frames. An initial stage of corpus annotation has been foreseen as a first ‘investigation’ phase. In a later stage, in which a suitable amount of annotations will be done, there will be the choice of whether and which kind of customizations are needed according to the corpus evidence and domain requirements.

As pointed out in [12], a key issue encountered while dealing with domain-specific texts is whether or not the creation of a new Frame is warranted. Within the legal domain the situation is made more difficult since the technical language used in the legal domain is closely intertwined with common language. According to linguistic studies (see among others [13]), legal language, still differing from ordinary language, is in fact not dramatically independent from everyday speech. This implies that it is no longer simply an issue of keeping existing Frames or creating new ones from scratch to convey domain-specific semantics. Accordingly, the specialization phase is concerned with the following three customization strategies which differ in their increasing degree of modification to the general FN resource:

1. the exploitation of domain-specific Semantic Types which classify Frame Elements from the general FN repository,
2. the introduction of one or more new Frame Elements within an existing Frame,
3. the splitting with a new Frame.

An example of 1. is provided in the excerpt of annotation reported in Section 5 below, where the Semantic Type “LegalDescription” has been added to the Frame Element “Principle” in order to ontologically type the lexical filler of this participant to the Frame “Prohibiting”.

Special attention is paid to the introduction of a new Frame Elements within an existing Frame. It is such the case of the following sentence *Il venditore deve consegnare al consumatore beni conformi al contratto di vendita* ‘The seller must deliver goods to the consumer which are in conformity with the contract of sale’, which instantiates the Frame “Being_obligated”, evoked by the Lexical Unit *deve* ‘must’. A new Frame Element “Beneficiary” should need to be added to the list of the semantic roles already existing to the Frame at hand, in order to describe the *addressee of the duty* (i.e. ‘to the consumer’). The original Frame only includes a Frame Element “Duty” (in this case ‘deliver goods’) and “Responsible_party”, i.e. ‘the person who must perform the Duty’ (in this case ‘the seller’).

In a sentence such as *uno Stato membro può vietare, per motivi di interesse generale, la commercializzazione sul suo territorio, tramite contratti negoziati a distanza, di taluni prodotti e servizi* ‘a Member State can prohibit, for reasons

of general interest, commercialization on its territory, through contracts negotiated at a distance, of certain products and services’, the splitting with a new Frame “Authority_prohibiting” is needed. The syntactic realization of the sentence above shows that it is an *enacting authority* (i.e. ‘a Member State’) which enacts a normative principle, i.e. a prohibition, rather than a “Principle” which prohibits a “State-of-affairs”.

4.2 Towards an ontological typing of Frame Element

In designing a FN extension for the legal domain, we considered the ontological typing of Frame Elements as a fundamental stage. According to [8], the general use of Semantic Types in FN is “to record information that is not representable in our frame and frame elements hierarchies”. It is done through the categorization of the sort of lexical fillers that is expected in a Frame Element. We intend to exploit this FN usage in order to domain-specifically categorize Frame Elements involved in a situation expressed by legal texts, on the basis of Legal Ontologies. As pointed out in [14], the real benefit of integrating a lexical and an ontological resource follows from distinguishing lexicalized and not-lexicalized concepts through keeping them as *distinct layers of semantic information* but even linking them.

The domain ontology we intend to use is the Core Legal Ontology (CLO)² [2], that specializes the DOLCE foundational ontology library³ [15]. CLO was chosen since it provides lexicalizations of ontological classes (i.e. juridical concepts), both in Italian and in English. Moreover, it has been exploited as an ontological resource reference in LOIS and in the DALOS project [16].

The possibility of mapping this FN-like resource onto a so-called *lexical ontology*, such as JurWordNet, is still under discussion.

5 An example of legal texts annotation

In this section, we report an example of annotation carried out on the Directive 1999/44/EC of the European Parliament and of the Council of 25 May 1999 on certain aspects of the sale of consumer goods and associated guarantees. For the annotation we used the Salsa Tool [17], freely available for research purposes. It offers a graphical representation of a text, already annotated at the syntactic level, and allows the user to annotate Frames and Frame Elements. Figure 1 shows the annotation of the following sentence: *La decisione 90/200 ha vietato l’esportazione dal Regno Unito di taluni tessuti e organi bovini solo il 9 aprile 1990* ‘The decision 90/200 prohibited the exportation from the United Kingdom of certain bovine tissues and organs only the 9th April 1990’.

Two Frames have been annotated: i) a Frame “Prohibiting”, evoked by the Lexical Unit *ha vietato* ‘prohibited’, together with three Frame Elements, i.e

² <http://www.loa-cnr.it/>

³ <http://dolce.semanticweb.org>

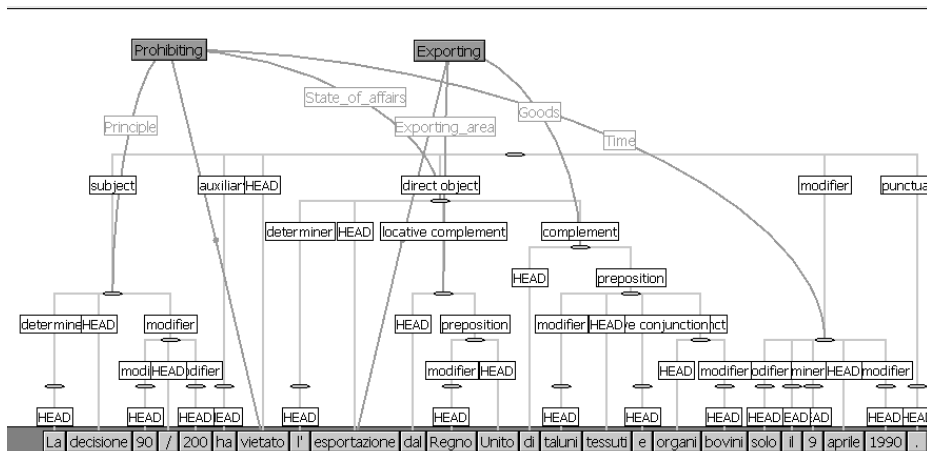


Fig. 1. An annotation example

“Principle”, “State-of-affairs” and “Time”, and ii) a Frame “Exporting”, evoked by *l’esportazione* ‘the exportation’, together with “Exporting_area” and “Goods” as participants. It should be noted that the annotated Frames refer respectively to the *legal domain* properly and to the *commerce domain* which is regulated by the Directive at hand. Interestingly, the two Frames are closely intertwined, in the sense that the textual span of the Frame Element “State-of-affairs”, part of the Frame “Prohibiting”, (i.e. *l’esportazione dal Regno Unito di taluni tessuti e organi bovini* ‘the exportation from the United Kingdom of certain bovine tissues and organs’) instantiates in turn the Frame “Exporting”.

The annotation of the textual span of Frame Elements was carried out on the top of the syntactic dependency relations automatically detected by the DeSR syntactic parser [18]⁴. The use of the Semantic Type “LegalDescription”, node of CLO, has been envisaged in order to ontologically type the lexical fillers (i.e. *la decisione 90/200* ‘the decision 90/200’) of the Frame Element “Principle”.

6 Conclusion

In this paper we introduced our approach to the construction of a FN resource for the legal domain. Through a customization phase of the general FN, we intend to produce i) an annotated corpus of legal texts and ii) a frame-based lexical-semantic resource. A strategy devoted to ontologically type the lexical fillers of Frame Elements annotated is foreseen as well in order to domain-specifically categorize participants involved in a situation expressed by legal texts. Through this, the developed FN resource will be linked to already existing

⁴ The parser used for this example was trained on a corpus of Italian newspapers; we are currently considering whether to develop a domain-specific version.

legal ontologies, thus resulting in a combined resource giving access to both the lexical and ontological aspects of legal texts.

Even though we present a work which is at an early stage of development, we foresee a number of possible applications and future extensions. Firstly, a frame-based annotated corpus of legal texts can be used to train test tools for semantic processing of legal texts, such as Semantic Role Labeling (SRL) tools. Namely, these SRL tools will be developed using language-independent unsupervised or semi-supervised machine learning algorithms, trained on the annotated corpus. The encouraging results achieved so far by SRL systems in the general language domain [19], are seen as an interesting opportunity to advance the state-of-the-art of Textual Case-based Reasoning (CBR) in the legal domain (see [20] for a frame-based approach).

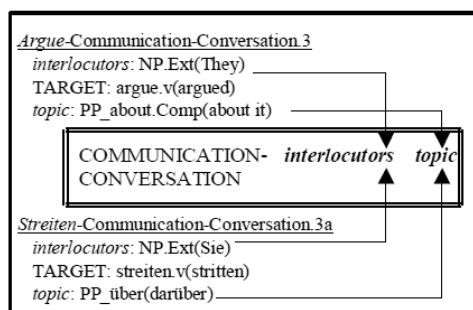


Fig. 2. Semantic frame as an interlingual representation [21]

Secondly, a multilingual FrameNet-like lexical resource can support semantic searching of legal texts in different languages. As reported in [21], where a bilingual German-English dictionary has been built on *Frame Semantics* principles, Semantic Frames are used as structuring devices to link multilingual lexicon fragments. Figure 2, extracted from [21], shows how a given combination of semantic and syntactic combinatorial properties of a given lexical unit in the source language has a correspondence link to its counterpart in the target language.

References

1. Valente A., Types and Roles of Legal Ontologies, in Law and the Semantic Web, LNCS, Volume 3369/2005, Springer Berlin / Heidelberg, 2005, pp. 65-76.
2. Gangemi, A., Sagri, M.T and Tiscornia, D. (2005), A constructive framework for legal ontologies, in Law and the Semantic Web, Benjamins, R., Casanovas, P., Breuker, J., and Gangemi, A., (eds), Springer Verlag.
3. Sagri M-T. (2004), Tiscornia D., Bertagna F., Jur-WordNet, in Proceedings of the Second Global WordNet Conference, pp. 305-310, Brno, Czech Republic, January 20-23.

4. Peters, W. , Sagri, M. T. , Tiscornia, D., The structuring of legal knowledge in lois, *Artificial Intelligence and Law*, vol. 15, pp. 117-135, 2007.
5. Fellbaum, C. (ed) (1998), *WordNet: An electronic lexical database*. MIT Press.
6. Kralingen, R., Oskamp, E., and Reurings, E. (1993), Norm frames in the representation of laws, in *Proceedings of Jurix*.
7. C. J. Fillmore, Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, (280):20–32, 1976.
8. Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C.R., and Scheffczyk, J. (2006), *FrameNet II: Extended Theory and Practice*, available online at <http://framenet.icsi.berkeley.edu/>
9. Visser, P.R.S., *Knowledge Specification for Multiple Legal Tasks; A Case Study of the Interaction Problem in the Legal Domain*, Computer/Law Series, No. 17, Kluwer Law International, The Hague, The Netherlands, 1995.
10. Venturi, G., *Parsing Legal Texts. A Contrastive Study with a View to Knowledge Management Applications*, in *Proceedings of Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Workshop *Semantic Processing of Legal Texts*, Marrakech, Morocco, May 26–1 June 2008, CD-ROM.
11. Isa Mara da Rosa Alves, Rove Luiza de Oliveira Chishman, Paulo Miguel Torres Duarte Quaresma, *The Construction of a Juridical Ontology*, in *Proceedings of ICAIL 07 June 4-8, Palo Alto, CA USA*.
12. Dolbey A, Ellsworth M, Scheffczyk J (2006), *BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies*, O. Bodenreider, ed., *Proceedings of KR-MED*, 87-94.
13. Mortara Garavelli, B., *Le parole e la giustizia. Divagazioni grammaticali e retoriche su testi giuridici italiani*, Torino, Einaudi (2001).
14. Prévot, L., Borgo, S., and Oltramari, A. (2005), *Interfacing Ontologies and Lexical Resources*, In *Proceedings of OntoLex 2005 - Ontologies and Lexical Resources*, Jeju Island, Republic of Korea, 15 October.
15. Masolo, C., Gangemi, A., Guarino, N., Oltramari, A., and Schneider L. (2004), *Wonderweb deliverable d18: The wonderweb library of foundational ontologies*, tech.rep.
16. Agnoloni, T., Bacci, L., Francesconi, E., Peters, W., Montemagni, S., Venturi, G. (2009), *A two-level knowledge approach to support multilingual legislative drafting* in: J. Breuker, P. Casanovas, E. Francesconi, M. Klein (eds.), *Law, Ontologies and the Semantic Web* Amsterdam, IOS Press.
17. Erk, K., Kowalski, A., and Pado, S. (2003), *The salsa annotation tool-demo description*, In *Proceedings of the 6th Lorraine-Saarland Workshop*, 111-113.
18. Attardi, G., and Dell’Orletta, F., (2009), *Reverse Revision and Linear Tree Combination for Dependency Parsing*, in *In Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies short papers (NAACL HLT)*, Boulder, Colorado. A detailed description of DeSR parser is available at <http://sites.google.com/site/desrparser/>
19. Gildea, D., and Jurafsky, D. (2002), *Automatic labeling of semantic roles*, In *Computational Linguistics*, volume 23, 245-288.
20. Mustafaraj, E., Hoof, M., and Freisleben, B. (2006), *LARC: Learning to assign knowledge roles to textual cases*. In *Proceedings of FLAIRS*, 370-375.
21. Boas, Hans C. (2002), *Bilingual FrameNet Dictionaries for Machine Translation*. In M. Gonzalez Rodriguez and C. Paz Suarez Araujo (eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation*. Las Palmas, Spain. Vol. IV: 1364-1371.