**MEANING:** universal multisensorial meaning representation stimulating tailored information exchange

**About you**

This proposal is submitted by a team of researchers at the Centre for Computational Linguistics (CCL), part of the research group Formal and Computational Linguistics (ComForT) of the department of Linguistics. We are submitting this proposal on behalf of the Faculty of Arts of the University of Leuven (KU Leuven). Our research focuses on machine translation, language technology for social inclusion, digital humanities, and formal linguistics. As we are working in an interdisciplinary area, we closely collaborate with researchers in other domains (speech recognition, computer vision, artificial intelligence, information retrieval, translation studies, terminology, remedial education, cognitive psychology), and are in close contact with valorization partners, such as the translation industry, health care institutions and organizations promoting social inclusion.

We have participated in and coordinated several national, binational and European (EU) projects, and our team consists of people that have been external reviewers for FP7 projects about machine translation and ontologies, for NWO Netherlands and for NSF South-Africa. We have organized several conferences and workshops, about computational linguistics, machine translation, and natural language processing for social inclusion. We are a member of the European Association for Machine Translation.

**What is the challenge and the vision?**

In today's information society people and machines are confronted with lots of information, anywhere, all the time. The amount of information provided by the Internet of Things, Smart Cities and Social Media is still growing, leading to an information overflow. Most information is only relevant at certain moments (everyday life, travel) for certain users (customized). On the other hand, relevant information may be missed, because it is not tailored to the needs of the specific user. This can be due to a language barrier between the user and the information, cultural or social bias, profile mismatches between user and target group, mental and physical capacities requiring adapted information or medium of communication, ... Information is not restricted to spoken or written form, but can also be visual (icons, pictures, video), tangible (braille, 3D prints), or smellable. In order to deliver the requested, personalized kind of information (per individual user, per situation) the full meaning of the information (i.e. the denotation and the connotation) that can be communicated in any form should be automatically derived and encoded/decoded in a universal representation, as soon as it becomes available. This should be irrespective of the language and culture of origin, the topic of information, or the humanness of the agents. When deemed useful, it becomes available to a user (person, machine) in the appropriate, tailored format.

Figure 1 gives a schematic overview of this idea, showing some examples of message media and target groups. For each of these types of messages needs, we need a way to interface with the Meaning Representation layer, encoding or decoding the information from or into the desired medium, and new types of messages (new languages, new media, new type of agents) can be added in a later stage. The world is thus considered a single communication space, defeating the localization/personalization bottleneck.



Figure . Schematic overview of how different media interat with an abstract universal multisensorial meaning representation.

This proposal sheds a new light on internationalization, localization, personalization and customization of information. It covers all communication aspects (any sense or medium, any topic, language, or culture), and satisfies the needs of a particular user at a specific moment in time, for example depending on his location or information in his agenda.

Many different (digital) technologies and research areas are involved, such as knowledge representation, ontology building, artificial intelligence (natural language processing, speech technology, computer vision), cognitive psychology, semiotics, linguistics, remedial education, anthropology, law (privacy), gerontology, sociology, and marketing.

Industry, commerce, and societal actors, as well as human end-users should be involved at all time, in order to develop a universal meaning representation system and interaction mechanisms that suits needs in a satisfying way for all parties.

**Why is it good for Europe?**

In Europe we are confronted with a whole series of official European languages, and people should be able to communicate with each other, with authorities, with their environment (Internet of Things, Smart Cities, (care) robots), anywhere in Europe, whenever they want and in a designated way. Migrants who recently arrived do not yet master the language used in their new home country, and a fair number of settled migrants never properly learned their new home language well enough to get by in all circumstances (for example to explain physical symptoms to a doctor, or to understand his questions and explanations). Due to dementia or another illness, people can lose full mastering of their language. A large group of the European population is to some extent functionally illiterate in the language in which they want or need to communicate, amongst them people with an intellectual disability (autism, down syndrome, attention deficit disorder), a physical disability (hearing impaired or deaf, visually impaired or blind, olfactory impaired/anosmia). They need tailored information, in a natural language (including sign language, if necessary simplified or tactile – like braille), pictographic language (meaning expressed in pictographs) or any other means of communication.

Travellers, online shoppers, but also native speakers of the local (European) language can experience problems, confronted with instructions and manuals in a language they don’t master (a recipe in French, Japanese description on tea bags in an Asiatic shop).

In order to overcome this burden for once and for all, we need to develop a semantically and pragmatically inspired universal multisensorial meaning representation system (language, culture and medium-independent) interfacing with all European languages, the languages of the trading partners/countries, the mother tongues of the migrants and other vulnerable minorities, and the mother tongues of the major part of the tourists. It should be extendable to other languages and communication systems.

This meaning representation, together with a standardized set of metadata, should be able to express all information, including the information not explicitly expressed by the agents (for example, as they obey Grice’s maxims of quantity: 1) be as informative as necessary, and 2) be not more informative than necessary). What is necessary differs per information receiver and per situation.

With such a meaning representation, industry and commerce no longer need to make their information available in different languages, as only one will be enough in order to generate the appropriate meaning representation. By removing the language barrier, we remove one of the largest hurdles in truly creating a single European market. There is no more need for explicit translation, and more users will be reached, as the information can be generated in many languages and media simultaneously.

On the other hand, people are not bothered with information they are not interested in. By receiving only relevant information, they will be more receptive to the information they do receive when they want / need it.

By crawling the web and extracting meaning from it, through information extraction from the semantic web and from linked data we can build up a database of world knowledge, which can further improve and refine future interactions between the meaning representation layer and the communicators.

**What would it take to do it?**

Mankind has been looking for an appropriate way to represent meaning for more than 2000 years. When truly aspiring a universal meaning representation layer that can interact with many different media and languages, we can consider this as a very large scale project, perhaps similar in size to the Human Brain Project, or the Human Genome Project.

Concerning the architecture of the Meaning Representation layer, we need to combine existing views on meaning representation from different scientific areas (philosophy, cognitive psychology, artificial intelligence, linguistics) into a flexible and adaptable unifying *theory of everything*. Nevertheless, the project can be subdivided into managable areas focusing on specific domains and interactions, in which academia collaborates with industry and commerce on the one hand, and with end-users on the other hand, to make sure that there is real economic and societal added value. Depending on the number of subprojects, we estimate an effort of 8 to 10 years with a consortium of several tens of researchers. In order to ensure a least biased as possible aproach, research and user groups from all over Europe should be involved.

There are several European research initiatives related to our proposal, However. they do not have the broad perspective sketched above.