LSP and Modern Technologies: Towards a Unified Theory of Knowledge

1. Introduction

The object of the Symposium was *Languages for Specific Purposes*. This initial sentence reveals and hides more than one meaning and begs a few questions. The first question that came to my mind was whether there is a language without a purpose, and I came up with the contradictory answer that no *language* has a purpose, and that no *instance of any language* does not have one. Purpose is part and parcel of our human condition (don’t we always act or speak with something at the back of our mind, or maybe at the front, if we are conscious at all of our words or actions?), and *language*, one of our fabricated concepts that still remains hard to define in spite of tons of linguistic literature, does not, as a concept, have a purpose. It can be argued that the purpose of language is communication, but it should rather be said that communication is the purpose of humans and that they use language amongst other means to achieve their goal. The term *language* is very ambiguous, and so is the term *object*. The latter could be synonymous with purpose. If I ask someone *What’s the object of your rude language?* I do not necessarily expect an answer as I know that the purpose of my interlocutor is to express antagonism, usually with the additional intention of destroying my views and ego. Rude language can take many forms, from a short sharp word to a long erudite argumentation sprinkled with disparaging terms, not to mention body language, but everybody can identify rude language by its purpose. So rude language is a type of LSP, although probably not the type targeted by the symposium organisers. An object can also be something *concrete* I can identify and manipulate with my sensory and motor systems, or something *abstract* that is in my mind, but we may well ask whether there is much difference between concrete and abstract language. Most terms, like *object* or *language*, are highly ambiguous, and most of the time we are not even aware of the ambiguity. Studies in neurolinguistics and neuropsychology have highlighted surprising ways in which we relate to our environment, our own bodies, and the many threads that underlie meaning. As a neurolinguist, I take the modest view that one’s knowledge can only be partial, and by this I do mean both incomplete and biased. Moreover, human knowledge is limited by our human nature, and we cannot expect to know the world as it is, but only as we see it through our mental representations. As a neurolinguist, my mental representations have developed in such a way that I believe that knowledge can only be relational, but that our limited human cognitive abilities should not be a source of mourning, as long as they allow us to improve our physical, mental, and social comfort. As one of the growing concerns of the 21st century is about knowledge, I wish to highlight what a theory of knowledge could be like from the point of view of neuroscience, how it can be in agreement with the semiotic points of view of Peirce and de Saussure, how it can encompass all types of knowledge, and how it points in the right direction to build prostheses that would help ‘the transition between an information to a knowledge society’. To become knowledge, information needs an interpreter. As a linguist, I propose in this paper to concentrate on how textual information becomes knowledge, and to refer to tools that already exist and are able to analyse differences in meaning in large text
corpora and to develop dictionaries of a new type. If they were developed further, they could also evaluate and compare knowledge in different domains and cultures, as well as assist text generation or translation. These tools are just the precursors of what I call artificial interpreters. In order to be consistent, they should be designed according to their specific overtly defined intentions. The intention – or purpose – of a particular model is realised by defining appropriate relation(s) and analysing their frequency. The intentional aspect is important because it is fundamental to human action, learning in particular. ‘The concept of artificial interpreters can have an infinite number of applications’.

2. Biological processes

2.1. The variability-universality paradox, knowledge and learning

Human language is at the same time universal and extremely diverse. This paradox is the result of basic principles of neuronal learning and human brain development. Humans benefit from a huge expansion of cortical areas that support the faculty to acquire language, sophisticated social behaviour and abstract knowledge. These areas develop over many years, and indeed throughout life. Universal processes allow for the great cultural variability that appears amongst human beings. ‘Variability expresses itself as different languages, different knowledge types and different LSPs.’

It is claimed that knowledge has been accumulating astronomically over the last decades. Yet it would be more appropriate to say that ‘information has accumulated’, and is ever growing. In order to transform information into knowledge, an interpreter is required. Individual humans, and groups of individuals, are aware of their limited capacity for grasping the totality of their own specialist domain – i.e. of their limited capacity for interpreting all the information in that domain. In order to cope with the vast amount of information available, stochastic models of language processing have been used with success, but most of them do not consider the way the brain actually processes language and knowledge. As it is argued here that human processing is relational, ‘a stochastic model should take this basic and ubiquitous property of learning and understanding into account’. Rather than taking language units as counting units for stochastic models, one should therefore count relations between these units. It is a fundamental change in the approach to knowledge and language processing, a change that gives hope for a move from an information society to a knowledge society. In order to lay the basis for artificial interpreters, let us first examine how humans “make sense”.

2.2. The human brain

We can only learn what our brains allow us to learn, but this only too often seems to be overlooked.

The cortex is not uniform: it resembles a quilted patchwork, and dedicated areas can be distinguished functionally as well as by looking at the particular cell architecture in those areas. The brain is a physical structure, and its activity can be measured by various tools and techniques.

The structure of the brain is determined only partly under instruction of the genetic code. Stimulation is crucial to develop structures that support particular activities, and highly skilled practice (such as playing the violin professionally) even results in visible growth of specific brain areas. Certain functions can also only be developed at a certain age, and we know how difficult it is to speak a foreign language without an accent if we learn it after a certain age. Yet it is always possible for a healthy brain to learn and this shows that the brain retains a degree of plasticity until late life. Brain structure determines brain activity, but brain activity also modifies structure.

Terms and concepts used in the neurosciences tend to err from brain to mind theory. Yet the brain is a physical object whose activity can be measured by various techniques, whereas the mind is an abstract construct structured with equally abstract constructs. Brain and mind theories
nonetheless merge because they both propose to be explanatory as far as functions are concerned. They also provide valuable hypotheses for one another.

Higher cognitive functions are not localized. They are supported by different brain areas that are connected to one another, often by very long nerve fibres. Yet we are not necessarily all connected in the same way: as Changeux (2003: 69) points out, functional invariance is compatible within connectional variability. After Geschwind (1965), brain processes have been envisaged as being supported by complex architectures of hierarchical and parallel neuronal networks. Geschwind suggested that taking either of the two previously rival approaches (atomistic or holistic) would inevitably lead to errors, and that animals and humans should not be considered as units but as ‘loosely joined wholes’. It was realized that language phenomena were so complex that neurolinguistics had to be supported by computer simulation – see Arbib/Caplan (1979); Perrin-Taillat (1984) and (1986). After a long debate about the modularity of the brain in the 1980’s and early 1990’s, and the premature death of Norman Geschwind, the ‘loosely joined wholes’ are now seen as a complex architecture of hierarchical and parallel networks (as briefly outlined section in 2.3. below). The notion of module is losing its vigour as it is becoming clear that different parts of particular networks are being used to subserve different functions.

2.3. Network generation: hierarchical and parallel processing

There is general evidence that stimulation plays an important role in all aspects of brain development after birth, including new cell generation, first discovered in primates by Gould et al. (1999). Synaptic generation (the development of connections between neurons) and degeneration (the death of connections between neurons) is the key to brain connectivity and has been known for a long time. The hypothesis was first proposed in the 1940’s by a Canadian psychologist, Ronald Hebb, and was confirmed in the domain of neuroscience by Markram et al. (1997).

The Hebb rule can be summed up by saying that neurons that fire together wire together. Frequency of co-firing is therefore essential to network connectivity. The more often co-firing happens, the stronger the connection. Conversely, if a neuron does not fire after receiving information from another neuron, the connection between the two becomes weaker and weaker and degenerates. If a neuron does not receive enough information from any other, it will itself degenerate.

Neurons that fire together eventually form a group that work together, and as a group they will connect to another group of neurons that fire just after receiving information from them. The Hebb rule therefore applies equally well to groups of neurons as it applies to individual neurons. As a result, ‘brain structure determines brain activity’, and ‘brain activity also modifies structure’.

The way information flows defines hierarchical networks, where groups of neurons of level \( n \) send information to groups of level \( n+1 \), and groups of level \( n+1 \) receive information from groups of level \( n \), as predicted by the Hebb rule. It follows that hierarchical networks work in parallel.

Hierarchical networks can share some of their constituent networks, or parts of them, as they work in parallel. Networks become interwoven into each other. As they share information – that sometimes travels in a loop – complex connectivity as well as functional and anatomical redundancy develop further in the central nervous system. Recent findings show that multimodal information is supported by multimodal networks that include multimodal neurons. Such is the case of bimodal visuo-motor neurons, trimodal somato-sensory, visual, and auditory neurons, and mirror neurons that have all been observed in recent years – see Rizzolatti/Sinigaglia (2008) for an account of studies in man and monkey.

Following Jacob/Jeannerod (2005), over-generalising recent findings on mirror neurons to mind-reading and social functions of animal and man is likely to lead to misconceptions. There are in our opinion many more networks involved in social behaviour than mirror neuron networks, and many parallel and hierarchical networks interfere with one another to serve language and knowledge processing. As these networks mature at a late stage of development, and are likely to change
over time, connectivity must vary from one individual to another, and looking for mirror neurons of social behaviour is likely to be a vain pursuit.

A basic principle underlies everything we learn: *a relation is always necessary*. This relation must be activated with sufficient impact to cause firing, and at a frequency high enough to allow the relation to materialise in brain matter. Once a relation is established – i.e. once neurons or groups of neurons are sufficiently connected – it can behave as a unit and connect to another group and so on *ad infinitum*. This is reminiscent of the semiotics of Peirce (1978) and de Saussure (1916–2005) – see section 3.2. below. *What is seen as a unit at a given level can be a relational structure at another.*

### 3. How do humans make sense?

By considering how neurons work one can only confirm that knowledge is relational, and can only be relational. Such a position was already taken by Spinoza (a philosopher who is popular in neuroscience – see Damasio (2003)), Peirce (a general scholar, semiotician and scientist), de Saussure (the founding father of linguistics), and Poincaré, a mathematician, scientist and philosopher, one of the last ‘generalists’ of knowledge, whose views are still widely accepted in science today.

**3.1. Universals of knowledge**

As mentioned in the introduction, the intentional aspect of knowledge is important. Eagerness to explore the unknown and seek explanations for everything, often at one’s own risk, seems to be universal. It is true of the infant who continuously explores her environment – and has to be continuously watched to avoid serious accidents – and it is also true of the scientist who works hard at her research, having grown a little older and presumably wiser between infancy and adulthood – or so it is hoped. This *apetitus noscendi* is, so to speak, the old saying re-enacted over and over again, *Curiosity killed the cat...but satisfaction brought it back*, the *sine qua non* condition of making life interesting, with the notion of *reward* attached to human undertakings.

If we consider the *intentions* of the Physical Sciences and Technology on the one hand, and those of the Human Sciences and Law on the other, they appear to be remarkably similar. Humans always have some *action* at the back of their mind, again with the notion of *reward* attached to it. The aim is to change one’s circumstances, which requires gaining power over the world – the physical world or the symbolic world as the case may be. This in turn requires a better understanding of the object of study, which is achieved by testing theories against the world, and by classifying the concepts that underlie theories. It can be claimed that testing has to do with the neuronal networks of action processing (the dorsal pathway) and classification with semantic processing (the ventral pathway), but both networks cooperate and as far as experts are concerned, one can assume that processing is a lot more complex and involves numerous networks distributed in the brain, including of course the frontal lobes.

We constantly make hypotheses and test them as we explore our physical world and our social world, and we gradually discover what works and what does not. The mini-models of the child are not very different from those of the adult, except that, as we grow older, models can be mediated by abstract language, mathematical language, and logic. The object of study is always envisaged dually: as a system – or structure – and as it happens – pragmatics: one has a theory, supported by concepts, and one tests it against the world. If experience tells you that your theory is wrong, you modify it, which changes your conceptualisation to some extent. If one fails to recognize that the conceptualization is wrong, a price will have to be paid: things will go wrong again. For instance, as long as the stock market is run on the basis of game theory, you may hire the best mathematicians in the world and yet still expose the world economy to the risks of big Casino traps. There is nothing wrong with game theory, except that it may not be the most adequate to run the economy. Mathematical language is good at making concepts unambiguous, or at tracing existing ambigu-
ity, but it must have adequate concepts to start with. The economic system, like any other system, is an abstract entity that cannot be conceived or refined without considering it as it appears. If it goes wrong, its underlying concepts should be modified.

Classification is a first-class tool to develop understanding. Poincaré expressed the view that classification is a useful commodity for the scientist, and not only for the scientist but for all men of past, present and future generations, and that this cannot be pure chance. This view is more generally taken up in science (Bouveresse 2003: 30-31). The most popular type of relation used to classify – *it is like/* it is not like – has to do with comparison and metaphor, the close companions of many classifications. Children, adults, all civilisations enjoy classifying things according to some criterion or other, and comparison and metaphor are omnipresent in language, even in the specialised languages of Physics and Law where one might least expect them. According to studies in anthropological linguistics, the taxonomies of plants in hunter-gatherer societies are very similar to the Linnaean taxonomy (Berlin et al. 1973; Atran 1990), even if they are not always in agreement with them. Classifications imply relations, whether the latter are implicit or explicit, and, according to Poincaré again, as quoted by Bouveresse (2003: 26), *only relations can be known* – which fits nicely with the way the brain works. We use all sorts of relations, similarity is just one of them, and as regards language, contextual relations are crucial.

3.2. Peirce and de Saussure

As pointed out above, a system is an abstract entity that cannot be conceived or refined without considering its use and it is generally agreed that meaning is use: abstract meaning cannot be elicited in the absence of meaning as it appears. Yet it is difficult to elicit meaning as it appears without referring to abstract meaning. There is always a degree of indeterminacy of meaning. So how comes that we actually understand anything at all?

The linguistic sign of de Saussure (1916–2005) consists of the *signifiant* and *signifié* that are as inseparable as both sides of a sheet of paper. This implies a strong relation between form and meaning. The sign is therefore tripartite, consisting of the *signifiant*, the *signifié* and the relation between them. Besides, the linguistic sign derives its value from the relationships it entertains with other linguistic signs of the language. So the linguistic sign is never alone and can only be considered amidst its counterparts. Does this mean that anything goes, that relations are unconstrained, that any sign can be related to any other sign? And how should the relation between signs be defined? Is it contextual? Is it semantic? De Saussure’s theory accommodates both paradigmatic (semantic) and syntagmatic (discursive, contextual) relations of signs. Besides, one of its strong points is to emphasize that the object changes according to one’s point of view. This allows for the notion of *indeterminacy of meaning* developed in section 3.3. below.

The sign of Peirce (1978) is explicitly tripartite, consisting of the *object*, the *representamen* and the *interpretant*. It is also more general than de Saussure’s as its ambition is to be applicable to our whole knowledge system. The object is a thing being represented by another thing (the *representamen*) in a third thing (the interpretant). At first sight Peirce’s model seems more constrained than de Saussure’s model: only interpreters can play a role in the generation of signs. But this is an illusion because the third party (the interpretant) is in fact a relation that binds an object and a *representamen*. An interpretant is an object and a *representamen* bound together. Besides, the interpretant is ‘determined’ to be also a *representamen* of the object to another interpretant. This means that what is a sign at level \(i\) (interpretant \(I_i\), binding object \(O_i\) and *representamen* \(R_i\)) is part of a sign at level \(i+1\) (interpretant \(I_{i+1}\), binding object \(O_{i+1}\) and *representamen* \(R_{i+1}\), where \(R_{i+1} = \text{interpretant } I_{i+1}\)). Looking back at the view of complex networks expressed in section 2.3., one cannot help thinking that Peirce was a precursor in brain theory as well as in many other fields. The interpretant allows for a sign of any particular domain to participate as a building block in another domain (or level). It can again be concluded, as in section 2.3., that *what is seen as a unit at a given level can be a relational structure at another.*
3.3. The indeterminacy of meaning

There is evidence from studies in neuropsychology that both frequency and context play a vital part in lexical access – see for example Simpson et al. (1988), for a review. The issues of word ambiguity and the indeterminacy of meaning have been extensively studied. It has been shown that the more frequent a word is, the more readily available it is for various linguistic tasks. The notion of typicality can be interpreted in terms of frequency: the more frequent a word, the more typical it is for its category. An apple is a typical fruit because it is a very commonly encountered fruit in Western society. It has also been shown that a robin and a canary are considered as typical birds. If it may be assumed that robins are typical because they are frequently encountered, it is difficult to believe that the same applies to canaries – most people probably have never seen one in real life. But canaries do occur frequently in children’s picture and story books. As regards bananas, oranges and satsumas, one could assume that they should be as typical as apples are because we have been in contact with them just as much as we have been with apples for decades, and they are widely illustrated in children’s picture books as well. Yet this simply is not the case. The difference is that they do not have a strong footing in our language and mythology. The apple is omnipresent, from compounds (Adam’s apple), idioms (he is the apple of her eye), stories and fairy tales (Snow White, William Tell) and mythology (the apple of discord; the fruit of knowledge). The apple appears in all manner of discourse, as well as in painting, which clearly is not the case of the more mundane oranges and satsumas. The concept of apple is twofold: it refers to objects in the real world, and it refers to a symbolic object. Whereas the former emerges in our mind from our contact and interaction with physical objects, the latter is generated from language. Our physical environmental context plays the vital part in the development of concrete language, and the linguistic context plays the vital part in the development of abstract language.

In the above example, it should be noticed that apple comes to be considered as typical through three different types of frequent encounters:

- encounters of objects in the physical world: in terms of brain processes, the specific associative areas of the five sensory modalities (seeing, hearing, feeling, smelling, tasting) interact with one another, with the motor modality, and with the speech modalities (articulating words and hearing them) to yield concrete language;
- encounters of visual symbolic representations of the objects (pictures in children’s books, paintings): the visual stimulus is itself a symbol and to work as such a different type of relation is necessary. In order to be established, it uses relations established above and reinforces them with new ones;
- encounters of linguistic representations of the objects (stories / discourse): yet another type of relation needs to be established. It uses relations established above as well as intra-discursive relations and inter-textual relations.

*For abstract language, intra-discursive and inter-textual relations dominate.* This view does not deny the fact that abstract language anchors itself in concrete language as our cognitive system develops in childhood (via metaphor, metonymy, action and purpose), but it highlights the idea that abstract meaning detaches itself from concrete notions in the course of development to acquire an independence of its own, even if the more concrete aspects remain in the background. This independence asserts itself in our late teens and adulthood and to a large extent through exposure to written language.

And now we come back to the importance of the point of view, first put forward by de Saussure. Our beloved *apple* will appear very different and appeal to different domains of our cognitive system depending on how we consider it. If we consider eating it our tasting and smelling modalities will be the most active, with our sensory-motor system playing a part too since we have to handle it as well as bite and chew it, and our visual modality will assist in the process; even our hearing modality will record the crunching noise. If we are peeling it, our tasting and smelling
modalities will be in the background as they may be twittering in anticipation, but the lead role will be taken by the sensory-motor and visual systems. The visual system will be primarily active if we are painting or drawing it. This illustrates the fact that polysemy is inherent to even the more concrete words: apple has different meanings depending on context. Now if we are reading Snow White or William Tell, our emotional system will come into play as it is being activated by the building up of the story as a whole. If we consider apple in a religious context, it will elicit feelings of guilt towards sex (according to popular belief spread by Church discourse) or towards the acquisition of knowledge (if we read the Bible). Here it is interesting to point out that the original fruit of knowledge (a substitute for the fire stolen from the Gods in Greek and Roman mythology) has become apple through a long pictorial and verbal tradition that dates back to the Middle Ages. This exemplifies the fact that words, and more generally signs, pertain to more than one level of our knowledge system.

If we consider how our knowledge is supported by our brain structure, we come to the view that meaning is inherently ambiguous. The belief that concrete meaning is less ambiguous than abstract meaning may come from the fact that a concrete (i.e. extra-linguistic) context is more likely to elicit the appropriate meaning than a linguistic abstract context.

4. Knowledge types, language types, and variability issues

4.1. Differences between the Physical Sciences and the Human Sciences

The Humanities and Science widely differ in their object of study. Whereas the Physical Sciences and Technology concern themselves with the ‘world out there’ (the natural world), the intentions and actions of the Human Sciences are directed towards the social world (the symbolic world). The question of the distinction between natural meaning – independent of the observer – and non-natural meaning (in the sense of Grice) is relevant here. If one accepts this distinction, it follows that the Physical Sciences do give us at least a partial view of the physical world as it is. If one rejects it, it follows that humans have no way of knowing what the world out there is like, and we may question whether it is at all real, and whether our own body and most familiar physical objects are not an illusion. If we opt for the latter option, we may either commit collective suicide, or decide that, as long as we feel reasonably happy, we might as well carry on as if we had opted for the first option. If our purpose is not to encourage mankind to commit collective suicide but to improve its physical, mental, and social comfort, the distinction can be considered of no interest to achieve that goal.

I still believe in a natural approach, and consider that what we learn in the neurosciences is not totally imaginary. We have reasons to rely on scientific methodology: theories and models are explicitly stated; they are constrained by mathematical language that either eliminates the ambiguity of its terms or controls it with a probabilistic approach. For science and its technological applications, departing from thoroughly tested scientific methodology is out of the question. Should rules and procedures not be applied with the utmost precision, most scientific experiments or technological applications would end up in disaster. The very fact that they do not implies that the success of the mathematical approach cannot be pure chance.

Although the Human Sciences and Law have adopted methods from modern Science, they are heavily language dependent. Because the concepts manipulated by philosophy, psychology, sociology, economics, law, and – last but not least – grammar and linguistics can neither be perceived through our senses nor measured by instruments, the only available measurement tools involve the use of probability. Yet it is not uncommon for studies to persistently yield contradictory results. This may result from many different causes: inadequate sampling, liberal use of confidence intervals, unmotivated rejection of sub-samples, and sometimes deliberate misuse of probabilistic studies to deceive for the advantage of some interest group. But the major cause is probably conceptual variability. In the course of speech or even written discourse, we shift from one meaning to another without realizing that we do. Experts, of course, tend to avoid this, but there are shifts
of meaning between individuals and cultures, that become greater as non-experts constantly interfere.

Unlike experts in the Physical Sciences who are left reasonably undisturbed to carry on with their activities, experts in the Human Sciences and Law have to take into account the views of other members of the social world. The terms they create are not only used by specialists who do not necessarily share the same view and may attribute them different meanings, but are frequently used by lay persons. They are also often imported into different cultural backgrounds that will interpret them according to their own traditions. The net result is that the concepts of the Human Sciences and Law are ambiguous despite the specialists’ efforts to constrain them. Variability is the rule rather than the exception, synchronically as well as diachronically. We can create artificial interpreters that assess shifts of meaning by defining relations explicitly and controlling variability within a mathematical framework.

Variability in the domains of the Human Sciences and Law may not be easy to appreciate but it is impossible to do without it in a democratic society. Specialists use theories, models, and classifications that create images of members of society, and as a result lay members of society have a right to debate in order to avoid being caught up in the images created by the experts. This constitutes the inalienable right to express one’s opinion. Should concepts of the Human Sciences and Law become rigid, there would be no scope left for debate, and totalitarianism would be on our doorstep.

4.2. Expert language

LSPs differ between themselves. Instances of language of the Physical Sciences and Technology tend to be interspersed with mathematical language and iconic representations: pictures, schemas, diagrams, etc. One approach to support translation is to make reliable glossaries and databases that offer reliable equivalents to the linguist, and to promote linguist-expert cooperation. For technical glossaries, relating terms to pictures, schemas and videos may be a valuable option. The Physical Sciences are also discursive, and the concepts they use can be elicited from text. All symbolic objects (large samples of pictures, diagrams, videos or texts) could be analysed by artificial interpreters, whose interplay can be envisaged as an alternative to traditional glossaries. Yet only texts are being considered here. Because mathematical language gives Science rigour and precision, the mapping of meaning by an artificial interpreter is expected to reflect this specificity.

Instances of the language of the Human Sciences and Law are discursive. As pointed out above, the object of study – a particular aspect of humanity – cannot be easily delimited, identified or measured, as is the case with aspects of the physical world; the actors involved – both specialists and non-specialists – give the same words different meanings, so that ambiguity constantly pervades discourse. As a consequence, it is desirable to measure differences in meaning, both synchronically and diachronically, in order to assess cultural variability, and existing prototypes of artificial interpreters can already do that (see section 4.4. below).

Both reading expertise and domain expertise ought to be taken into account to devise expert language interpreters. Experts are expert readers in at least one language, and they are experts in a particular domain. Unlike speech, reading and writing do not come naturally to infants or older children: expert – that is fluent and efficient – reading and writing are the outcome of a long learning process involving many years of training. Written language adds new dimensions to our neuronal networks and revolutionises both our competence and our strategies. Expertise requires practice of reading and writing as well as learning grammatical rules as set by those who codified the language. These grammatical rules are normative, and it is important to be aware that they are, in that respect, unlike the descriptive syntactic rules devised by linguists to account for a natural (oral) language as it occurs. Yet both grammatical and syntactic rules can be described in terms of relations. LSPs have their specific normative rules: if one wants to be part of a community, learning them is a
must to prove that one can express oneself as it is customary to do within that community. These rules can also be expressed in terms of relations.

If one was interested in child language development, it would be valuable to consider sound and articulatory movement and their relationship to phonetic and semantic units. If one was interested in the acquisition of reading by children, it would be interesting to see how patterns of strokes on paper or screen relate to phonemes, syllables, words, and meaning. But as far as the expert reader is concerned, these lower level relations can be overlooked because one can assume that they function as units. This is not unreasonable because, as we have seen in sections 2.3. and 3.2., what is seen as a unit at a given level can be a relational structure at another. An expert reader can recognize the various flexional forms of a word as representing the same word automatically. He can also recognize idiomatic expressions for what they are. An expert in a particular field can equally well identify technical collocations. Considering the constituent words and the syntax of idiomatic expressions and technical collocations is therefore undesirable if one proposes to create an artificial interpreter for assessing meaning variability for LSPs. Various forms of the same term, idiomatic expressions and technical collocations can therefore be labelled as single units before such an interpreter sets to work on the corpus.

4.3. Texts as physical objects and artificial interpreters

All domains of knowledge have produced masses of physical objects: ‘texts’, either couched on paper or recorded digitally. Texts exist independently of those who have produced them – specialists in the case of LSPs – and of those who are going to read them – specialists or non-specialists, native speakers or not. The interpretation of a text by individual human readers may differ greatly, but large textual corpora are big physical objects that are no figment of anybody’s imagination, or interpretation.

The body of text in a particular domain may be considered to represent common knowledge in that domain. But saying that it represents common knowledge is not the same as saying that it is common knowledge. Texts are not knowledge, but information; in order to become knowledge, they need an interpreter (an individual, a group of individuals, or an artificial interpreter, in any case a processor that defines relations between bits on information in the text). Knowledge is relational and can only be relational, and the frequency of each relation determines the emergence of meaning. In order to devise an artificial interpreter, both aspects need to be implemented.

As we have seen in section 3., abstract word meaning comes from a syntagmatic relation, that is to say a contextual relation. Human interpretation depends on the knowledge network of the interpreter, and this network is unstable from “womb to tomb” because the activity of interpreting itself changes the network. Moreover, a human interpreter does not explicitly choose what relation(s) he uses and the relations probably change in the course of interpretation, leaving the interplay of relations at least to some extent – maybe to a large extent – unconscious and uncontrollable.

As regards artificial interpreters, the system has to start from somewhere (this approach is legitimate because what is a unit at a given level can be a relational structure at another). Accordingly, initial relations must be explicitly defined to fit the purpose of the interpreter. In order to avoid any arbitrary form of meaning, the defined relation can only be contextual, of the type $x$ belongs to the same context $c$ as $y$ or $x$ is $n$ units before $y$ within context $c$. Such relations are determined by the definition of context $c$, the definition of the units denoted as $x$ and $y$ (the type of elements considered within $c$), and, in the second relation, the value of the integer $n$. The relation will be different depending on whether the purpose of the model is to elicit grapho-phonology, morphology, syntax or lexical semantics, and it should ideally be supported by experimental studies in neurolinguistics and psychology. In order to establish a parallel with neuronal network architecture (and the Hebb rule), the system must count, for any pair $(x, y)$, how frequently the relation occurs in the corpus, and define a frequency threshold. Relations can only be established
if they occur frequently. As a consequence relations that occur below a certain frequency should be eliminated. This will also automatically eliminate low-frequency units. Frequency threshold is different from the threshold activity value necessary for a neuron to fire. The latter applies to the activity of established networks in use, the former is a frequency value, not an activity value. It is the frequency of occurrence of a relation necessary to establish that relation, and therefore to structure networks. Frequency threshold has to be defined, and it can be expected that there is an optimum value. As the purpose of an artificial interpreter, as well as the units it relates, are clearly defined, it constitutes an explicitly defined model, constrained by a mathematical probabilistic approach. Moreover, relations do not vary during the course of interpretation, as is the case with a human interpreter. An artificial interpreter therefore eliminates subjectivity and allows humans to keep track of particular knowledge aspects objectively. By coupling various interpreters together, it should be possible to reflect the multifaceted aspects of human knowledge, while eliminating subjectivity.

4.4. Existing models: new types of ‘Dictionaries’

Semantic Atlases have been devised by Sabine Ploux and her collaborators (see Jacquemin / Ploux 2008 for an overview; Ploux Semantic Atlases [online] for more information). The approach is mathematical and statistical and the models fulfill some of the criteria mentioned above. They concentrate on lexical aspects of language and their great advantage over ordinary dictionaries and semi-automated models is that they avoid any arbitrary delimitation of meaning. They extract meaning completely automatically by defining a relation between words, and applying graph theory and correspondence factor analysis. In doing so, they define units of meaning much finer than the word. These units make it possible to find equivalents in different languages (provided the model has been applied to equivalent large corpora in those languages) and provide a tool to appreciate differences in meaning (either within or between languages).

The first models, based on synonym dictionaries, use a synonymy relation (Ploux 1997). Later models, of greater interest for LSP analysis, are based on large corpora. They use a contextual relation, the context being a sentence, or a window of \( x \) words (Ji/Ploux 2003). Originally, the contextual model generated a lot of noise, and this was remedied by various ad hoc devices. All forms of the same word were also counted as different words, which interfered with statistical analysis: relations that should have been significant appeared as different relations, and were not accounted for. Current models of contextual Semantic Atlases have introduced lemmatizers to remedy this problem, and they have also introduced some form of basic syntax that reduces the arbitrariness of contexts. Although these measures are a great improvement, one might regret that they still suffer from a theoretical deficit. From a linguistic point of view, closed-class words are still being rejected by the current models, and therefore still remain unrelated to word morphology, phrase structure and discourse structure (three levels of linguistic analysis articulated by closed-class words). From a neurolinguistic point of view, the definition of context could be improved by considering reading strategies of expert readers, in particular with reference to neurolinguistic studies on eye-movement. From an LSP point of view, they do not address the issue of expert knowledge – domain expertise or reading expertise.

Despite these criticisms, current models perform relatively well. They have shown their superiority to dictionaries in at least two ways: they have eliminated any arbitrary delimitation of meaning and have provided a unit of meaning much finer than the word. This makes it possible to map bilingual dictionaries automatically by using two large corpora, one in each language. These finer grain units of meaning also make it possible to measure differences of word meaning within or across languages. Besides, since a model can only elicit meaning present in the analysed corpus, it automatically produces specialised dictionaries from specialised corpora, and analysing texts from a particular culture automatically elicits the meaning of the terms within that culture. By analysing another corpus – in the same language or in another language – and mapping terms as it has been done for bilingual dictionaries, it could be possible to highlight differences in meaning.
across cultural systems. And as Semantic Atlases draw terminology from context, they present a substantial additional advantage: it is possible to trace back the contexts in which any particular term appears, which paves the way towards automatic text generation and translation.

5. Conclusion

This paper started with a few casual remarks highlighting the ambiguity of words. In the course of argumentation, it was shown that all words are ambiguous, even the more concrete ones, because of the way we process them. Ambiguity is omnipresent, particularly in the Human Sciences. The Physical Sciences fare better in that respect because they use mathematical language. Unlike natural language, mathematical language is precise, and provides a logical framework to eliminate or at least control the ambiguity of concepts that underlie a particular theory. The idea is not to get rid of ambiguity in the Human Sciences because, as it was pointed out, it is necessary in a democratic society, but to assess it in order to map out differences. The idea is to treat symbolic objects as physical objects in order to create artificial interpreters. In the case of language and LSPs in particular, symbolic objects are texts, or any discourse that could be materialized as text, and the concepts that underlie the theory are concepts borrowed from linguistic theory. Because it has been argued that knowledge is relational, and can only be relational, it has been proposed that mathematical language should control the relations between linguistic units rather than linguistic units themselves. There are infinite possibilities of relations, and the purpose of a particular interpreter is implemented by defining a particular relation explicitly. Different relations have different properties – reflexive, symmetric, transitive – and the choice of a mathematical framework depends on these properties. Because the only relations that must survive are those that occur frequently, the approach must be probabilistic, and artificial interpreters must operate on large corpora to achieve statistical significance. The implementation of artificial interpreters requires a multidisciplinary team of linguists, neurolinguists, mathematicians, and programmers. Once implemented and tested, it must be validated, by specialists of the domain under consideration in the case of LSPs. As regards meaning, one can expect greater variability in specialised texts from the Human Sciences than in texts from the Physical Sciences because the concepts that underlie the latter have already been constrained by mathematical language.

Artificial interpreters should also be conceived from the start with a view to cooperating with one another, each dealing with a particular level of representation. Yet a single interpreter, as exemplified by the Semantic Atlases of Sabine Ploux and her collaborators, can already perform rather well. Artificial interpreters exist as prototypes, and developing them further is a matter of political will. Talented linguists, mathematicians, programmers, and domain specialists tend to be employed in sectors that offer them good salaries, and a substantial research budget is needed to develop artificial interpreters. The ‘society of knowledge’ has a price!

References


Markram, Henry/Lubke, J/Frotscher, M/Sackmann, B 1997: Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. In Science 275, 213-215.


