Saussure meets the brain

Jan Koster

University of Groningen

1 The problem

It would be exaggerated to say that Ferdinand de Saussure (1857-1913) is an almost forgotten linguist today. But it is certainly the case that Saussure, considered the most important linguist of the century in Europe until the 1950s, hardly plays a role in current theoretical thinking about language. As a result of the Chomskyan revolution, linguistics has gone through a number of conceptual transformations, which have led to all kinds of technical pre-occupations that are far beyond linguistic practice of the days of Saussure. For the most, it seems, Saussure has rightly sunk into near oblivion. Nevertheless, there is one famous maxim of Saussure's that is highly relevant for current thinking about cognition, in particular about the relation between the mind and the physical sciences. According to Saussure, the relation between a sign (*signifiant*) and what it stands for (the conceptual *signifié*) is accidental: *le signe linguistique est arbitraire*. What Saussure says about the individual sign can also be said about complete representations: there is no intrinsic relation between a representation and what it represents.

Such ideas about the arbitrary relation between form and meaning were not at all original with Saussure. Aristotle recognized the conventional nature of signs in his *De Interpretatione* and the Saussurean idea can also be found in Stoic philosophy of language. No matter the history of the idea, it is still higly relevant today. Computers, for instance, process information that is representational in nature and therefore, according to the tradition from Aristotle to Saussure, without intrinsic significance. Like all other

representational information, everything found in computers is only something in relation to external interpreters, such as human beings.

John Searle's famous Chinese room argument, which purports to show that semantics is not intrinsic to syntax, is just a variant of the Saussurean insight. The same can be said about Searle's more recent statement that syntax is not intrinsic to physics. These much advertized "discoveries", in other words, just echo the insights of a very long tradition.

It should be noted that the Saussurean maxim not only applies to digital computers but also to analogous machines with parallel architecture or to neural networks. The Saussurean thesis can be taken to be completely general: all forms of representation only represent under an interpretation that is extrinsic to the representation itself. A clear statement of similar ideas can be found in Wittgenstein's Blue and Brown Books, in which the author argues that the interpretation of a representation cannot be some kind of shadow representation. From a slightly different angle, the issue is discussed in Polanyi (1962, 81), where it is made clear that the application of a formalism to reality involves some degree of indeterminicy. The manner of application of a formalism cannot be part of the formalism itself.

The implications of such simple and obvious traditional ideas are profound. Since representations are arbitrary and in need of an extrinsic interpretation, there cannot be something like an autonomous world three in the sense of Popper (1972). Neither scientific formalisms nor religious or legal texts can stand on their own feet. In all cases, we have to deal with the indeterminicy of extrinsic interpretation, which is amply clear from the histories of science and religion and also from our judicial practice.

The question which I would like to address is whether the traditional insights in question have consequences for the brain sciences or for our theories of cognition.

2 The dual nature of the brain

One sense in which the term "interpretation" is used is in the translation from a representation into another. Let us call this form of interpretation t-interpretation. The ideas discussed in the previous section imply that human understanding cannot be a form of t-interpretation. Under any

form of t-interpretation, we remain in the realm of representation, which is arbitrary and in need of further external interpretation. So, to see human understanding as a form of t-interpretation would lead to infinite regress: each next representation would stand in need of another interpretation, and so on.

In order to escape infinite regress, there must be another form of interpretation, human understanding, which I will refer to as u-interpretation. U-interpretation seems to involve the conscious mind of a living human being and cannot be found at the level of dead matter. Computers only involve t-interpretation and also the actions of robots can be seen as t-interpretations of program states. If we abstract away from meaning, t-translation can be taken also to include the processes that lead from DNA (a representation) to protein synthesis.

Everything we know about the brain or about language involves representations and t-translations, while u-interpretation remains completely in the dark (perhaps necessarily so). John Searle's recent writings can be seen as an attempt to dismiss a cognitive science based on representation and t-translation in favor of a cognitive science that tackles u-interpretation head-on. It seems to me that this plea for a shift in our efforts is ill-adviced and will run into the unsurmountable problems which are ultimately implied by the Saussurean rift between form and meaning,

Also Edelman's 1992 harsh attack on "functionalism" (the somewhat misleading cognitive science name for Saussurean ideas about mental representation) seems completely beside the point. Edelman denies that the working of the mind is based on symbolic representation and bases his view on the incompatibility of representationism with the insights of modern biology. It is hard to determine what Edelman is talking about because it is literally inconceivable that somebody can memorize a poem without acquiring a representation of that poem. Similarly, a huge segment of human language is explained by theories of the representational sort, namely generative grammars. There are not only no alternatives of comparable scope to such theories, it is also naive to point out that a biological interpretation is missing. As is implied by the tradition from Aristotle to Saussure, it is in the very nature of representations that they cannot be explained in terms of biology (Edelman) or physics (Penrose). Human representations are coded somehow in the brain and therefore physical and biological objects, but their representational nature is just as much beyond physics and biology as the

working of a computer.

It is no doubt true, as Searle (and also Edelman 1992) seem to imply, that there cannot be ultimate understanding of the brain and human understanding in terms of representation and t-interpretation. As said before, we can only avoid infinite regress by somehow discharging the world of representation. In this sense, there is a homuncular residue in our theories of cognition that nobody knows how to get rid of, neither Dennett (1991), nor Edelman or Searle.

The problem is that everything that is known about the brain in terms of neurology lies in the area of representation and t-interpretation. Sensory impressions, broken up in features and analyzed by numerous modules, are processed and "translated" into electrico-chemical signals, which can in turn be transported to the relevant central areas of the brain, where sometimes very specific cell groups respond to very specific information. At no point during such processes do we "break out" the world of representation and t-translation. The same can be said about the differential establishment of neuronal connections in networks of whatever size and about the selective reinforcement of neural activity at synaptic clefts under the control of neurotransmitters. It is clear that the brain is not a digital computer with serial Von Neumann design. But it is also clear that our current neuro-sciences can only describe the brain as any other information processing system in the sense that it exclusively involves representations and t-translations that are arbitrary from the point of view of meaning and u-interpretation. In this respect, the neuro-sciences have nothing to offer that goes beyond the usually dismissed computer metaphor.

According to the the tradition epitomized by Saussure's maxim this is exactly what one would expect: anything material, including the neurophysiological (or Penrose's quantum physical processing), is arbitrary with respect to meaning and u-interpretation. The impossibility to bridge the gap between form and meaning is not just an imperfection of our current scientific understanding but a matter of logic: representations cannot represent their own interpretation without infinite regress.

This is not to say that u-interpretation falls outside physical reality or forces some form of ontological dualism upon us. If we wish, we may stick with some form of monistic materialism and assume that the brain not only contains massive representational systems but also areas that are responsible for u-interpretation. It is not unreasonable to assume that certain purely

material processes manifest themselves as understanding (or as consciousness, knowledge, experience, qualia). The problem is not necessarily ontological but epistemological. Even if we could describe exactly which material processes are responsible for certain forms of understanding -a remote but perhaps not impossible goal- we would not rid ourselves of Saussurean arbitrariness. For us as "external" observers, the constitutive processes that discharge representations into understanding, would be indistinguishable from t-interpretation. We would never understand why other representations and other material processes on these representations could not lead to the same forms of understanding.

3 Summary and conclusion

If anything material is arbitrary with respect to u-interpretation and if u-interpretation is the core of consciousness, perception, knowledge, experience and other such qualia-invested phenomena, we seem, in practice, to be doomed to a form of epistemological (but not necessarily ontological) dualism. On the one hand, we have neural representations and processes and on the other hand, we have their u-interpretation. As for the neural representations, the situation is entirely the same as for computers, contrary to what is often assumed in this area. Since u-interpretation is extrinsic to material representational systems, there is little reason to expect that biology or physics will contribute much to our understanding of the mind. There is no reason to assume that neurophysiology is more crucial to the workings of the mind than silicon is to the workings of computers. Like computation, the scientifically accessible part of the mind is ultimately a realm comparable to mathematics, something that can be implemented in a certain type of hardware but which has no necessary relationship with it.

If these conclusions are correct, the cognitive sciences have a closer family resemblance to mathematics than to biology or physics. Much of the brain can only be understood in non-physical terms, as a representational system that functions with respect to an extrinsic u-interpreter, a homuncular residue that cannot be discharged for the reasons given above.

This conclusion is more or less the opposite of the one arrived at by Searle (1992) and Edelman (1992). Instead of emphasizing the representational

nature of much of the brain, they tend to dismiss it by denying the relevance of computational theories of cognition. If I am right, however, computational and other representational theories are the only ones attainable for cognitive science. Searle and Edelman are quite right in their insistance that computation is extrinsic to biology or physics, but nothing whatsoever follows from this observation. It certainly does not follow that biological systems cannot be used for computation or that human cognition must not be characterized in computational terms. At best, it follows that cognition is more than just computation, that there must be mechanisms responsible for u-interpretation, which give life to our computational/representational mind. However, like everybody else, Searle and Edelman have failed so far to demonstrate that our homuncular residue is more a matter of biology or physics than the powers of representation and computation that it brings to life.

It is perhaps important to remember that our deepest and in some sense most successful form of rational inquiry is not biology or even physics but mathematics. Mathematics is a human enterprise and falls as such within the limits of our biologically-given cognitive capacity. But this does not mean that it can be explained in biological terms and even less that it should give rise to Edelman-style complaints about a lack of significance in the shining light of modern biology.

The existence of mathematics simply shows that there is successful rational life outside physics and biology. So far, the emergence of more or less successful cognitive sciences like theoretical linguistics points in the same direction.

References

Dennett, D. (1991), *Consciousness Explained.* Little, Brown. Boston.

Edelman, G. (1992), *Bright Air, Brilliant Fire. On the Matter of the Mind.* The Penguin Press, London.

Penrose, R. (1994), *Shadows of the Mind. A Search for the Missing Science of Consciousness.* Oxford University Press, Oxford.

Polanyi, M. (1962), *Personal Knowledge.* Routledge and Kegan Paul, London

Popper, K. (1972), *Objective Knowledge.* Oxford University Press, Oxford.

Searle, J. (1992), *The Rediscovery of the Mind.* MIT Press, Cambridge, Mass.

Wittgenstein, L. (1958), *The Blue and Brown Books*. Harper and Row, New York.