Dissociation between regular and irregular in connectionist architectures: Two processes, but still no special linguistic rules

Marco Zorzi[a] and Gabriella Vigliocco[b]

[a]*Department of General Psychology, University of Padua, 35131 Padua, Italy.* [b]*Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706;* **mzorzi@psico.unipd.it** **gviglioc@facstaff.wisc.edu**
www.psychol.ucl.ac.uk/marco.zorzi/marco.html
psych.wisc.edu/faculty/pages/gvigliocco/gv.html

Abstract: Dual-mechanism models of language maintain a distinction between a lexicon and a computational system of linguistic rules. In his target article, Clahsen provides support for such a distinction, presenting evidence from German inflections. He argues for a structured lexicon, going beyond the strict lexicon versus rules dichotomy. We agree with the author in assuming a dual mechanism; however, we argue that a next step must be taken, going beyond the notion of the computational system as specific rules applying to a linguistic domain. By assuming a richer lexicon, the computational system can be conceived as a more general binding process that applies to different linguistic levels: syntax, morphology, reading, and spelling.

Two-process models represent a fairly general solution to the "quasi-regularity" problem in a number of cognitive domains. Quasi-regularity implies that, in a given problem domain, examples occur for which the solution is not fully consistent with the regularities represented in the problem set as a whole. Three paradigmatic quasi-regular domains within the English language are (1) the past tenses of verbs, (2) reading aloud, and (3) spelling.

For all these domains, various two-process models have been proposed. In the area of reading aloud, for instance, the idea that the regularities employed in reading nonwords (i.e., novel forms) may be represented separately from knowledge of individual word pronunciations has a long history (see, e.g., Baron & Strawson 1976; Coltheart 1978). A similar architecture has been postulated for spelling (see, e.g., Ellis 1982). The similarity between these two domains and inflectional morphology is striking. Neuropsychological studies of brain-damaged patients have shown double dissociations between the two forms of knowledge for all these domains.

In reading, surface dyslexic subjects show an impairment in reading irregular words but can read regular words and nonwords. Phonological dyslexics, on the other hand, cannot pronounce words they have not encountered before (e.g., nonwords). A corresponding dissociation has been observed in spelling (see Denes et al., 1999, for a review of acquired dyslexias and dysgraphias).

For the past tense domain, an equivalent dissociation has been recently shown by Ullman and colleagues (1997b). What is most striking, however, is that problems with the irregular inflections are associated with problems in reading and spelling irregular words, whereas problems in the production of regular inflections

are associated with problems in reading novel words (Ullman et al. 1997b). Furthermore, in the study by Ullman et al. (1997b) the patients performing poorly on regular past tenses were also those with "syntactic" problems (agrammatism and Parkinson's disease), whereas the patients performing poorly on the irregular past tenses were anomic (posterior aphasia and Alzheimer's disease). Although neuropsychological associations are not considered to be very informative (see Shallice, 1988, for discussion; also book reviews and Précis of Shallice's *From Neuropsychology to Mental Structure BBS* 14 (3) 1991), in the present case they may be.

Some neural network modellers have proposed that the distinction between word-specific representations and a computational system has no psychological validity and that a neural network with a single, homogenous route from input to output can handle both the regular and the irregular cases and still can generalize the regularities to novel cases (see, e.g., Plaut et al. 1996; Plunkett & Marchman 1993; Rumelhart & McClelland 1986; Seidenberg & McClelland 1989). However, the successful simulation of double dissociations has proved elusive for single-route models. Recently, Bullinaria and Chater (1995) have presented a very careful and insightful analysis of the properties of single-route neural network models, looking at how they manage to handle both productive regularities and exceptions in a single knowledge base and how these capacities dissociate under disruption (addition of noise, removal of hidden units, etc.). The authors demonstrate that, especially as the complexity (size) of the problem increases, double dissociations do not occur under disruption. They conclude that their results "set a challenge to modelling researchers to show that rule/exception double dissociations can occur in such networks" and they predict that "such a challenge cannot be met" (Bullinaria & Chater 1995, p. 260).

Neuropsychological double dissociations, however, can be easily handled by connectionist two-process models. Zorzi et al. (1998a) have shown that the distinction between word-specific information and (componential) knowledge about the regularities of the domain can be quite easily realized in standard neural networks if the input and output layers, as well as being linked via some kind of intermediate representation (hidden units), are allowed to make direct contact, that is, that part of the network is a *two-layer* net. A kind of "modular decomposition" (see Jacobs 1999) emerges in the system simply in response to the different computational demands posed by the problem of learning regular and irregular items. The use of two-pathway network architectures results in a decomposition of the problem in terms of regular versus irregular (or componential versus word-specific) for the different domains of reading (Zorzi et al. 1998a; 1998b), spelling (Houghton & Zorzi 1998), and past tenses (Westermann 1998).

If we consider inflectional morphology and reading, there is evidence that in both cases equivalent computational solutions can fit the observed data. Where does this lead us? To an interesting new and parsimonious hypothesis. Assuming a structured lexicon (see Levelt 1989; McDonald et al. 1994), we may not have to postulate that each domain draws from specific abstract linguistic rules (part of the language faculty) but instead might postulate some more general binding mechanisms that allow for the *componential* manipulation of stored codes (phonological codes in particular). These binding mechanisms may be shared between domains, and/or they may share resources. As has been shown by recent connectionist models, the ability of productive generalization in a given domain can be based on simple associative systems (rather than on abstract, algebra-like rules; see also McClelland & Plaut 1999). However, the issue of rules versus connections must not confound that of one versus two mechanisms: Connectionist two-process models (see, e.g., Zorzi et al. 1998a) dispense with explicit symbolic rules but predict double dissociations.