

## Sense tagging: does it make sense?

Jean Véronis  
Université de Provence  
29, Avenue Robert Schuman, 13100 Aix-en-Provence (France)  
Jean.Veronis@up.univ-mrs.fr

Sense tagging is probably one of the challenges that corpus linguists have to face in the near future. So far, computerisation of this task has yielded very modest results despite numerous efforts, and sense tagging is turning out to be a touchy task. Difficulties stem from various sources, extracting disambiguating information from the context. However, one of the main problems that lies upstream of the disambiguating process is the sense inventory itself. Most tagging efforts rely on traditional dictionaries to supply the reference senses, or on computer-oriented resources such as *WordNet*, which do not differ significantly from traditional dictionaries in terms of sense division. The present paper shows that human taggers perform very poorly when given a traditional dictionary as the reference, and that machines should therefore not be expected to perform any better if the same kind of resource is used. A detailed analysis reveals the lack of distributional criteria in dictionary entries: traditional dictionaries are chiefly concerned with meaning definition, and not with the surface clues (syntactic, collocational, etc.) that are required to match a given sense with a given corpus occurrence. It is argued that no fundamental progress can be made until large-scale lexical resources have been built that incorporate extensive distributional information, and that, until that time, any massive sense tagging efforts based on traditional dictionaries or computer-oriented resources such as *WordNet* would not only be premature but also questionable in terms of resource management.

**Keywords:** sense tagging, polysemy judgements, interannotator agreement, dictionaries, distributional information, word sense disambiguation.