We suggest that the standard fine-grained division of senses and (larger) homographs by a lexicographer for use by a human reader may not be an appropriate goal for the computational WSD task. We argue that the level of sense-discrimination that natural language processing (NLP) needs corresponds roughly to homographs, though we discuss psycholinguistic evidence that there are broad sense divisions with some etymological derivation (i.e., non-homographic) that are as distinct for humans as homographic ones and they may be part of the broad class of sense-divisions we seek to identify here. We link this discussion to the observation that major NLP tasks like machine translation (MT) and information retrieval (IR) seem not to need independent WSD modules of the sort produced in the Research field, even though they are undoubtedly doing WSD by other means. Our conclusion is that WSD should continue to focus on these broad discriminations, at which it can do very well, thereby possibly offering the close-to-100% success that most NLP seemingly requires, with the possible exception of very fine questions of target word choice in MT. This proposal can be seen as reorienting WSD to what it can actually perform at the standard success levels, but we argue that this, rather than some more idealized vision of sense inherited from lexicography, is what humans and machines can reliably discriminate.

3.1 Introduction

In Chapter 2, Kilgarriff identifies the source of the WSD “problem” as the attempt to assign one of several possible senses to a particular occurrence.
of a word in text – in particular, pre-defined sense lists provided in dictionaries and similar lexical resources. He goes on to suggest that the proper assignment of word senses requires a vast amount of lexical, syntactic, and pragmatic knowledge, together with generative procedures that can be exploited for every occurrence – a position reminiscent of the artificial intelligence (AI) community’s objections to statistical natural language processing (NLP) two decades ago. At the same time, Kilgarriff gives a nod to “the important role” of pre-established lists of word senses for WSD, by which we assume he means that the identification of some limited number of broadly defined senses is useful in language processing applications. He seems to be suggesting, at least obliquely, that while lexicographers and linguists seek to represent word meaning in all its depth and complexity, NLP can provide some useful results by relying on far less. This is exactly right, but it begs the question of how much – or, more to the point, how little – information about word meaning is actually required to do something useful in NLP, given our current capabilities.

Interestingly, although this question should be pivotal for those engaged in the WSD activity, within the NLP community very little progress has been made toward answering it directly. Perhaps this results from aiming too high: for example, the organizers of Senseval-2 state that “[Senseval’s] underlying mission is to develop our understanding of the lexicon and language in general” (Edmonds and Kilgarriff 2002:289). It is difficult to resist the temptation to answer the hard questions that have been debated by philosophers and linguists for millennia, rather than continue hard practical work within the considerable constraints on our current understanding of lexical semantics. But as Robert Amsler recently pointed out,

I fear the state of our understanding of theoretical lexical semantics is about where astronomy was 2000 years ago. The theory or even the logical arguments as to what stars in the heavens (or the semantics of words) must be will be debated for years to come without affecting the work of those of us empirically measuring what is observable and predictable (Senseval discussion list, 1 http://listserv.hum.gu.se/mailman/listinfo/senseval-discuss 27 August 2004).

Here we take a practical view of WSD, beginning with a reconsideration of the role of lexicographers in word-sense disambiguation as a computational task, as providers of both legacy material (dictionaries) and special test material for competitions like Senseval. We suggest that the standard fine-grained division of senses and (larger) homographs by a lexicographer for use by a human reader may not be an appropriate goal for the