1. Automating some types of language processing holds great promise for helping us develop new ways of drawing insight from the world's linguistic legacy. But "promise" has many meanings, and this is a promise that has not yet been kept. [...] We find ourselves at the threshold of a new era. Behind us is an era of almost entirely manual markup and transcription; ahead we envision increasing reliance on automation for at least the more mundane parts of that work. [...] Why is the future perpetually just over the horizon? The reason, I argue, is simple: those who could build these marvels don't really understand what marvels we need, and we, who understand what we need all too well, don't really understand what can be built."

—Douglas W. Oard 2009, "A Whirlwind Tour of Automatic Language Processing for the Humanities and Social Sciences"

2. The history of information retrieval research has been strongly dominated by a focus on retrieval of what we might call "formal" content, content written with dissemination in mind. Such content potentially has high value, but constitutes only a tiny fraction of the words produced by our planet's 7 billion people. Recently, activities such as the TREC Blog and Microblog tracks have begun to explore how retrieval systems might be tailored to the unique characteristics of informal content. Perhaps unsurprisingly, it turns out that informal content poses unique challenges for Cross-Language Information Retrieval (CLIR) as well. In this paper, we begin to explore those challenges.

—Bagdouri et al 2014, "CLIR for Informal Content in Arabic Forum Posts"

3. It is important to recognize that "technological innovation" encompasses far more than mere technical innovation—equally important is our ability as a society to learn to productively use the technical capabilities that we can create. This chapter examines one such technology: helping users to find information in ways that "flatten" language barriers. In keeping with what is emerging as common usage, we refer to this challenge as "Multilingual Information Access (MLIA).

—Douglas W. Oard (2009) "Multilingual Information Access"

4. "Among all of the advances in CLIR, none has had anywhere near as large an effect as accurate translation probabilities. [....] The best reported results for systems that use translation probabilities well is closer to 100% of what would have been achieved using same-language queries. It is worth taking a moment to consider what that means—today we can build systems to search French documents that work (approximately) equally well regardless of whether the query is written in French or in English."

—Douglas W. Oard (2009) "Multilingual Information Access"

5. One commonly cited limiting factor for Web search engines has been the challenge of developing a suitable business model for monetizing MLIA. Regardless of the cause, it seems clear that developing a broader experience base with MLIA techniques will be an important next step. [...] These emerging capabilities are first steps in the direction of developing a richly multilingual information ecology that could support the next generation of research on information-seeking behavior in such settings.

—Douglas W. Oard (2009) "Multilingual Information Access"

6. A cross-language ontology is an ontology whose concepts are lexicalized in more than one language, That is, each core node in the ontology represents an abstract concept and has associated with it a "label" (a word or phrase) that identifies it in a natural language, given that concepts must be expressed in the vocabulary of some natural language, if users are to interpret the ontology. Each concept node in a RAVEN ontology is labeled by the following convention: the concept name for a node always begins with "C-" and ends with an English word or phrase that can define the concept to a native English speaker. The link between nodes specifies a relationship between the concepts. [...O]ntological relationships of three types are defined: hierarchical, equivalent, and associative relations. For example, in an ontology on transportation, the concept labeled "car" may have each type of link: one for an associative relationship to the concept labeled "gasoline", one for an equivalent relation to the concept labeled "automobile", and one other for a hierarchical link to the concept "vehicle".

—Valerie Cross and Clare Voss, 2000 "Fuzzy Queries and Cross-Linguistic Ontologies in Multilingual Data Exploitation."