245 VOCABULARY IN SELECTION SYSTEMS Buckland. March 9, 2003.

Multiple Vocabularies Every community (domain of discourse) develops its own vocabulary.
The repertoire of values in metadata systems (thesauri, classification systems) are “vocabularies.”
Problems arise from dissonance between two (or more) vocabularies: Searchers, authors, catalogers,...
For “automobile” look for term TL 205 in the Library of Congress Classification, for 180/280 in the U.S. Patent Classification, and for 3711 in the Standard Industrial Classification, but who knows that?
Using “natural” language as metadata does not eliminate the problem. Using “Automobiles” will find nothing in the U.S. import and export statistics. Data can be found using “Car,” but refers to railway and tramway rolling stock. Data relating to automobiles are under “Passenger Motor Vehicles, Spark Ignition Engine” but who would look there?
A search on “coastal pollution” in the Library of Congress Subject Headings and in Medical Subject Headings (MeSH). “Coastal” and "Pollution" as subject keywords yielded no results. Relevant records, found by searching for these two words in titles, revealed that the subject headings were, in ranked order:
LCSH: Marine pollution; Coastal zone management; Water -- Pollution; Petroleum industry and trade; Beach erosion; Coasts; Barrier islands; Coastal changes; etc.
MeSH: Seawater; Water pollution; Bacteria; Water microbiology; Air pollution; Environmental monitoring; Bathing beaches; Environmental pollution; etc.
All selection systems involve multiple vocabularies. Even in the most primitive case, where unedited natural language is searched with unedited queries, there are at least two: 1. The vocabulary of the author(s) of the document(s) searched; and 2. The vocabulary of the searcher. In operational systems the number of vocabularies is likely to be much larger, e.g. the vocabularies of the cataloger, modifications provided by the “see”, “see also, and searchers’ guesses at the “system vocabulary.”

Ambiguity in Multiple Stage Processes The intent of the transitions is, of course, to normalize term usage so that any discrepancies are rectified. If the searcher wants A and the author used B, then we might expect a cataloger to rectify the author's language into that of the searcher and the selection system. Ambiguity builds up cumulatively.
Inconsistency. Variation in an individual's usage or because many different persons are involved, each with his or her personal vocabulary (synchronous variation) -- and over time (diachronous variation). The likelihood of dissonance arising increases as the number of transformations, of persons involved, and the passage of time increase.
Mapping. “Vocabulary control” through USE references. Indexes to classification schemes.
Form and Significance. There is a duality between the form and significance of words: A same form of word can have different meanings in different vocabularies and the same meaning may be expressed by different forms of words. Minimally, four contingencies in the relationships between two words:

		SAME FORM	
		YES	NO
SAME MEANING	YES	Same	Synonym
	NO	Homograph	Not same

In a hypothetical closed system the same word would invariably have the same meaning defined a priori so homographs should be absent and synonyms, if not absent, should be recognized.
"Same" means, in practice, equally acceptable for the purpose at hand.
Recommended reading: Vocabulary as a Central Concept in Library and Information Science
http://www.sims.berkeley.edu/~buckland/colisvoc.htm