InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: Lexical Standards Report

Status:	Final (public)
Version:	3.0
Submission Date:	November 2003
Release Date:	December 2003
Author:	The InterPARES 2 Project
Writer(s):	Naomi Cull Deptartment of Linguistics, University of Toronto
Project Unit:	Terminology Cross-domain
URL:	http://www.interpares.org/display_file.cfm?doc= ip2_terminology_lexical_standards_report.pdf

1. Introduction

As a member of the Terminology Cross-domain Task Force, I have been investigating the standards used in constructing glossaries, thesauri, dictionaries and registers. The purpose of the research is to ensure that the methods used to create and maintain the *InterPARES 2 Terminology Database* are precise and consistent. To that end, Ian Lancashire and I have synthesized the established lexicographical definitions employed in the field and used by InterPARES for the various lexicographical documents. The sections that follow outline these definitions. Additionally, the standards used by InterPARES 2 and lexicographers for each type of lexicographical document are present below. We recommend that the following definitions be adopted:

(1) **COMPOSITE DEFINITIONS AND RECOMMENDATIONS (Lancashire and Cull)**

InterPARES 2 Terminology
definition: an unpublished Register and published hypertextually linked Glossary, Dictionary, and Thesaurus
purpose: to explain and map InterPARES 2 concepts for retrieval, and to relate them to the terminologies of archival science as employed outside InterPARES 2

Register
definition: an alphabetical word-list generated by mechanical computer concordance from a collection of InterPARES 2 documents minus stop words plus suggested terms created by researchers
model: output from *Concordance* software program
method: created by staff members from documents on Web site and online suggestion box submissions of co-investigators
published as: working document only, not part of final Terminology

Glossary
definition: an alphabetical conceptual vocabulary of technical words and phrases employed within InterPARES 2, in which each separate word-entry has a headword with qualifier information in brackets, a part of speech, a single definition or sense, and one illustrative quotation from an InterPARES 2 document, or a hyper link from the headword to the InterPARES 2 source document
model: Sue Ellen Wright's Web-Glossary FAQ
method: created by researchers from internal reports and studies with added entailed terms
published as: Web-based database

Dictionary
definition: a subject lexicon of English terms relevant to archival science employed within InterPARES 2 and the archival community at large, in which each (preferred and non-preferred) word-entry has a headword with qualifier information in brackets, a part of speech, syllabification, a usage label (e.g., *archaic, law, science,* etc.) when necessary, a list of known senses, and one or more sources for each sense (but without etymology, pronunciation, and illustrative quotations)
model: *Merriam-Webster Dictionary*
method: transfer Glossary intact, adding a source indication (i.e., that each sense is InterPARES 2); add headwords, senses, source attributions from a select list of subject lexicons relevant to current archival science, music, art, computer science as well as citations and direct quotations
published as: Web-based database

Thesaurus

definition: a master index of all headwords in the *InterPARES 2 Glossary* and *Dictionary*, classified by means of facets, codes and cross-references so as to indicate their semantic relationships, with part of speech but without definition or sense, illustrative quotation, source, or any other lexical information

model: The *Arts and Architecture Thesaurus* (with facets) and the *NASA Thesaurus* (without facets; modelled after ANSI/NISO Z39.19-1993. http://www.niso.org/standards/resources/Z39-19.html)

method: conflate all headwords in dictionary, add semantic encoding, and link them to Glossary (for InterPARES headwords) and Dictionary (for all other headwords)

published as: Web-based database

(2) **Further Recommendations for all Lexicographical Documents** (derived from the set of accepted standards used by lexicographers as described below, e.g., Wright, Pearce-Moses, the *NASA Thesaurus, Merriam-Webster Dictionary*)
(These recommendations do not include matters of display or format.)

1. Use infinitive verb forms without the infinitive particle 'to'. (Wright Biv; contra NISO)
2. Use the singular noun form (Wright Bv)
3. Enter complex terms in their natural order; and
4. Enter complex terms in their inverted word order with a cross-reference to the natural ordered entry (contra Pearce-Moses)
5. Provide part of speech information (Wright Bxi; but see page 6; contra Feb. 2003)
6. Do not limit the number of characters in the term entry for the glossary, thesaurus and dictionary. (Wright Cv; contra the *NASA Thesaurus*)
7. Use American English spelling for entries and list all spelling variations (cf. page 8)
8. List homographs as separate entries (Wright Av). The first entry in the dictionary, but not the glossary, should be the one that appears in IP2 documents; homographs from other sources follow.
9. Provide fully inflected forms, e.g., *oxen*, for irregularly spelled words in the dictionary and glossary.
10. Indicate which set of standards was employed in the creation of the lexicographical document, e.g., *ISO, NISO, OED, US DoD,* etc.[1]
11. Provide a bibliography of the source documents, preferably with hyperlinks to those documents (Wright Eii). The bibliography should also be standardized, e.g., using *ISO 690-1987 Documentation-Bibliographic references-Content, form and structure*. (Wright Div)
12. Provide full forms for acronyms that are not readily apparent to the non-specialist.
13. Use scope notes to give the date, author, URL, etc. of the reference indicated. Hyperlinks should be included to the actual source. (Wright Diii, Eii)
14. Use a spellchecker and grammar checker.
15. Implement electronically online, e.g., as a database that ColdFusion publishes to the Web.
16. Do not use the term entry in its own definition (Wright Ciii). In a phrasal headword, one can use a part of the term as long as it is elsewhere defined.
17. Use full sentences in the definition of an entry and not just a synonym or equivalent. (Wright Ci) The recommendations from the InterPARES 2 meeting in February 2003 allow for the use of equivalent expressions.
18. Include entries for all entailed terms, i.e., unfamiliar or technical terms used in the definition of a term entry. Include hyperlinks from entailed terms to the term entry in which the entailed term is used. (Wright Bxii)
19. Format all three lexical documents identically. (Wright Cvi)
20. Each entry in the glossary and thesaurus should represent a single concept. (Wright Aii)
21. Use the relationship indicator HN to indicate when a term was added to the thesaurus and the history of modifications to any part of the entry.
22. Include modifiers in the glossary and dictionary. For example, the modifier 'electronic' from 'electronic record' would have its own entry.

[1] ISO = International Organization for Standardization. NISO = National Information Standards Organization. OED = *Oxford English Dictionary*. US DoD = United States Department of Defense.

1.1 Definitions in the InterPARES 1 and 2 Documents

I have searched through the available documents on both the InterPARES 1 and 2 Web sites to find definitions for the four types of lexicographical documents (register, glossary, thesaurus and dictionary) and to see what standards are currently in use by the InterPARES Project. From my review of the online InterPARES documents, the list of documents I have reviewed is included in an appendix, the following definitions were found:[2]

(3) *Register*

Quoting from Luciana Duranti, in an email dated 2003/03/13:
"…the Terminology Cross-domain Team decided to generate a Register of working terms and phrases by means of concordance software that used the texts of the UBC Project Glossary, the InterPARES 1 book, and the InterPARES 2 proposal and overview of intellectual framework to create a corpus of words and phrases to be concorded." "The purpose of the Register is to identify words and phrases that need working definitions." A request for additions and deletions to the preliminary register was made, and respondents were asked to provide working definitions for additional terms and phrases.

(4) *Glossary*

a. *The InterPARES Glossary*[3]
'The InterPARES Glossary contains definitions for terms that are used in InterPARES research materials. The primary purpose of the InterPARES Glossary is to facilitate the communication of ideas and research findings by clearly defining key terminology that is used within the InterPARES Project (additional background information about the InterPARES Glossary is provided in the document's preface).'

b. In the paper *The InterPARES Glossary* by Ken Hannigan,[4] presented June 2002 at College Park, MD, the definition for the InterPARES Glossary is 'A controlled vocabulary of terms used in the InterPARES research' (slide 5).

c. From the Terminology Cross-domain Task Force Meeting[5]
'A terminology is not a dictionary but contains arbitrary definitions of words employed by researchers in a given project. These definitions need not reflect common usage. A glossary explains hard or difficult words in a text or texts. Its definitions are not arbitrary but are taken from common dictionaries. (I. Lancashire)'
'The IP1 Glossary defines the words in the IP1 text. The definitions were developed by the researcher who wrote those texts. These definitions were not constrained by the "outside" (L. Duranti)'

(5) *Thesaurus*[6]
'by thesaurus we mean a collection of different words that mean the same thing, that may offer preferred terms'

I have not found definitions for *dictionary* in the InterPARES documents as yet.

[2] I used the search tool available at.http://www.interpares.org/rws/index.cfm to look through the InterPARES documents for 'register,' 'glossary,' 'thesaurus' and 'dictionary.'
[3] http://www.interpares.org/reports.htm.
[4] http://www.interpares.org/documents/interpares_glossary_tf.pdf.
[5] February 13, 2003, p. 1, http://www.interpares.org/rws/display_file.cfm?doc=ip2_wk5_terminology_action_items.pdf.
[6] From the Terminology Cross-domain Task Force Meeting. February 13, 2003, p. 1, http://www.interpares.org/rws/display_file.cfm?doc=ip2_wk5_terminology_action_items.pdf.

1.2 Formal Definitions

In addition to the definitions for these four lexicographical documents as used in the InterPARES documents, there are formal definitions.[7]

(6) *Register* (from the online *Oxford English Dictionary 2*):

> 'A book or volume in which regular entry is made of particulars or details of any kind which are considered of sufficient importance to be exactly and formally recorded; a written record or collection of entries thus formed; +a list, catalogue; a record of attendance at a school.'

(7) *Glossary* (from *Making Dictionaries. A Guide to Lexicography and the Multi-Dictionary Formatter.* by David F. Coward and Charles E. Grimes. SIL International. Waxhaw, North Carolina. 2000. p. 67. http://www.sil.org/computing/shoebox/MDF_2000.pdf:

> 'A glossary is usually no more than a listing of the headword[8] (lexeme) and a simple gloss or two. Sometimes it also includes part of speech. It does not include example sentences, synonyms, multiple senses, etc. A glossary is sometimes a necessary minimum for archiving dying languages and cultures, but should not be the goal or final result of any significant fieldwork. Minimal entries in a dictionary, and typical entries in a glossary are about the same.'

(8) *Dictionary* (Coward and Grimes: 67):

> 'A restricted portion of the lexical database ($lexicon_2$) that is published for a primary purpose and a primary audience. A dictionary provides a systematic exploration of the vocabulary of a language, including, among other things, meaning, range and usage. A dictionary normally uses some convention of alphabetizing to organize the material. Dictionaries normally do not include housekeeping information, but extract information from the lexical database for formatting. The broadest kind of dictionary is a comprehensive general purpose monolingual or bilingual dictionary. More specialized dictionaries might focus on kin terms, body parts, plants, fish, or animals. A medium-sized dictionary for publication has around 5,000 entries. A significant dictionary has over 10,000 entries (counting headwords as an entry).'

(9) *Thesaurus* (Coward and Grimes: 67):

> 'A thesaurus is organized along different principles than a dictionary, generally around semantic domains. Very few general thesauruses for minority languages have been usable by the local communities. This is for a variety of reasons which are not yet well understood, but they include: the organizing categories chosen by the compiler do not fully match the categories recognized by the community themselves; and how to use a thesaurus is not immediately transparent, etc. An attempt at a published thesaurus for a language is not recommended until a full dictionary has been published first.'

2. Sources for Standards

In this section I outline the standards that are in current usage for the four lexicographical documents defined above, that is, the standards used by InterPARES 2 Project and those that are used and recommended by other organizations. As part of this survey, I will compile the information in an annotated bibliography. Some of the works to be included in the

[7] By 'formal' I mean official standards that are recognized and employed by lexicographers.
[8] According to the online OED, the headword is the subject of the dictionary entry or article.

annotated bibliography are listed in the bibliography at the end of this report. Here are the standards used to construct the InterPARES 1 and 2 glossaries as indicated in documents found at the InterPARES Web sites.[9]

2.1 Lexicographical Standards Adopted by the InterPARES Project

2.1.1 InterPARES 2 Project Standards for a Glossary (Terminology Database)

Within the InterPARES 2 Project, there are fifteen fields of inquiry and twenty countries pursuing research on electronic records,[10] thus it is crucial that strict standards be followed to ensure consistency within the glossary. According to Duranti 2001,[11] the same principles used to organize the InterPARES 1 Project have been extended to that of InterPARES 2. Thus I have been working under the assumption that the same standards used to construct the *InterPARES 1 Glossary* were used to construct the *InterPARES 2 Terminology Database* (July 17, 2003). According to page two of the *Glossary Committee Report*,[12] terms and their definitions entered into the *InterPARES 1 Glossary* 'complied with agreed rules of style governing grammatical form, use of the singular and plural, spelling, capitalization, and use of acronyms. The style format rules were based on *ISO 5964: Documentation-Guidelines for the Establishment and Development of Multilingual Thesauri* and on the *Oxford English Dictionary*.' Note that the style for the glossary is based on those used by the ISO for a thesaurus. Duranti (in an email to the ip2-terminology and Ian Lancashire, dated July 20, 2003) has since recommended that both the glossary and the thesaurus be monolingual given the time constraints. Using ISO standards for this and other lexicographical documents helps to achieve the objectives of the glossary that are: '1. internal consistency and 2. to be able to share definitions with outsiders, provides context.'[13]

The glossary was intended to be a conceptual vocabulary, not a linguistic analysis of words.[14] Thus information pertaining to the term entry's part of speech, etymology and other grammatical information is not meant to be included.

[9] A list of the InterPARES documents used in this report is included in the appendix.
[10] See *International Research on Permanent Authentic Records in Electronic Systems (InterPARES): Experiential, Interactive and Dynamic Records* written by Luciana Duranti 412-2001,
at http://www.interpares.org/rws/display_file.cfm?doc=InterPARES_2_detailed_proposal.pdf, p. 15.
[11] See *International Research on Permanent Authentic Records in Electronic Systems (InterPARES): Experiential, Interactive and Dynamic Records* written by Luciana Duranti 412-2001,
at http://www.interpares.org/rws/display_file.cfm?doc=InterPARES_2_detailed_proposal.pdf, p. 6.
[12] Available at http://www.interpares.org/book/interpares_book_h_part5.pdf, which is Part Five of *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*
at http://www.interpares.org/book/index.htm; no date is indicated on the document.
[13] Terminology Cross-domain Task Force Meeting. February 13, 2003, p.
1, http://www.interpares.org/rws/display_file.cfm?doc=ip2_wk5_terminology_action_items.pdf.
[14] Terminology Cross-domain Task Force Meeting. February 13, 2003, p.
2, http://www.interpares.org/rws/display_file.cfm?doc=ip2_wk5_terminology_action_items.pdf.

2.1.2 InterPARES 2 Project Standards for a Thesaurus

Duranti refers to the *ISO 2788-1986:Documentation-Guidelines for the Establishment and Development of Monolingual Thesauri* used for the InterPARES 2 thesaurus in an email to Jonathan Furner on Wednesday, July 09, 2003. As mentioned above, a more recent email to the InterPARES 2 list dated July 20, 2003 indicates that the thesaurus will now be monolingual.

2.1.3 InterPARES 2 Project Standards for a Register

I have as yet found no formal statement as to the style used for the Register of Terms and Phrases.

2.1.4 InterPARES 2 Project Standards for a Dictionary

I have as yet found no formal statement as to the style used for the dictionary; however, from the Terminology Cross-domain Task Force Meeting. February 13, 2003, p. 2, '[d]ictionary definitions should be written in a way that they could be substitute for the words in the text, so they could be used as surrogates for the word.'[15]

2.2 Lexicographical Standards

There are a number of standards used by lexicographical organizations, such as the ISO,[16] the *OED* (*Oxford English Dictionary*), and others. Some information regarding these standards is included here.

2.2.1 Standards for a Glossary

2.2.1.1 Wright 2002

Wright[17] gives a brief description of some of the standards used in a Web-based glossary. Some of her main points are listed below. Standards pertaining to a multilingual glossary have been omitted here.

(10)
A. Glossary Format
   i. The glossary should consist of a small number of domain-specific *terminology entries*, or *term entries*. These domains include but are not limited to art, history, law, the social sciences, etc.
   ii. The entry definitions should related to a single concept.

[15] http://www.interpares.org/rws/display_file.cfm?doc=ip2_wk5_terminology_action_items.pdf.
[16] International Organization for Standardization, http://www.iso.ch/iso/en/ISOOnline.frontpage.
[17] 2002; http://appling.kent.edu/ResourcePages/TSA-2003/TSAWeb/Final%20lineup/Terminology_Management_FAQ.htm.

iii. Only one entry should relate to a specific concept-oriented term.
iv. Synonyms are listed within a related entry and not listed separately.
v. Homographs, or polysemes (terms with the same spelling but different meanings) are listed in separate entries.
vi. A list of subfields should be created and used consistently with a term entry when necessary, e.g., **friction facing**, n, <*driven discs*>…

B. Term-Related Issues

The canonical form of an entry should be used; that is, the traditional form of a word as you might see it in a dictionary entry. For instance,

i. Correct spelling of term entries.
ii. Terms are in lowercase, unless uppercase is the norm, for acronyms for instance, e.g., NASA. Note that this is different from the *thesaurus* standard used by the ISO, where preferred terms are given in uppercase. If a language, such as German, writes a word with the first letter in uppercase, this tradition should be followed in the glossary so that users will know how to use the term.
iii. The singular form of a term is used, unless the term is normally used in the plural, e.g., pants, scissors, etc.
iv. Use the infinitive form of a verb, without the infinitive particle, e.g., walk, run, compute, not *to walk, *to run, *to compute., etc.
v. Multiterm words should be given in the spoken order of the word, e.g., digital component, not component, digital or digital record, not record, digital.
vi. Use boldface, a larger font, etc. to make the term entries stand out from the rest of the text.
vii. Indicate the difference in meaning between synonyms listed with the term entry.
viii. Indicate if the term is associated with an abbreviation, full form or variant.
ix. Indicate if variants are associated with a particular region.
x. Synonyms, abbreviations, full forms, and variants should be easy to find in an entry and not buried in a note.
xi. Provide grammatical information for the term entry and any synonyms listed within the entry. For example, in the separate listings for 'act' we would see **act** (n.) and **act** (v.).
xii. Terms that are used within in the definition or notes of a term entry are called *entailed terms* and should be listed separately in the glossary. 'Any possibly unknown term used in a definition should be clearly defined in its own term entry' (Wright 2002:3). Wright suggests creating hypertext links from words that occur in definitions to their entries in the glossary.

An important point to consider with respect to Bi, is that given the international nature of the project and the lexicographical documents, is an American or British spelling to be used for words such as 'favour,' 'behaviour,' 'colour,' etc., or will alternate spellings be provided in the entry? Since the most common spelling is usually the one used for the term entry, an American spelling should be adopted, with variants provided.

C. Definition-related Issues

i. In the definition of an entry, it is not enough to use a synonym as the synonym itself may not be understood. Here is an example from Wright (p. 3): **diaphragm spring**, n, <*drive train technology*>, Belleville spring. If the reader doesn't know what a Belleville spring is, then he or she will have no understanding of what a diaphragm spring is either.
ii. In a hierarchically arranged conceptual glossary, the definition should contain superordinate rather than subordinate information; that is, the definition should be broad rather than narrow. For example, in a hypothetical entry for *apple*, the

definition should state the genus that the object belongs to, in this case 'fruit' and not the species, e.g., Macintosh.

iii. The definition should not contain the term being defined.
iv. Provide one good definition per term rather than several definitions in order to avoid confusion.
v. Definitions should be brief but complete. Additional information can be put into a note.
vi. Formatting should be consistent throughout. For example, the term entry should always begin with an uppercase letter, or always be in lowercase, or there should always be a period at the end of the entry or no period, etc.

Limiting the length of the term helps control the amount of information in the entry, which will be a means of avoiding repetition and confusion, thus the recommendation in Cv should be adopted.

D. Context-related issues

Contextual references are especially helpful for those unfamiliar with the way in which a term is used. For a list of different types of contexts, see the Glossary of Terms below.

i. Terms can be used in context by showing how the term is used in relation to other words that often occur with the term. This also serves to illustrate that the term is actually used in the way described in the definition.
ii. Be consistent in the documentation of contexts.
iii. An illustration of an object defined in a term entry is very useful. The illustration can be a bitmap in which the user can move the cursor to different parts of the illustration which are then hyperlinked to the term entries in the glossary.

E. Documentation

i. In order for users to trust what they read in a document, a hypertext glossary can have the terms linked to the proper documentation that shows the authenticity of the definition of the term.
ii. Ensure that all sources for the term information, such as definitions, contexts, etc. are identified.
iii. If information in a term entry comes from a Web site, the URL should be included in a note.
iv. A reference list that uses a bibliographical standard should be provided, especially in hardcopy material.

I strongly recommend including hyperlinks in the scope notes in the terminology database, as well as the other lexicographical documents, in order to provide the user with a path to the actual source as the term entry is used in context. One of the advantages of providing a hyperlink is that it saves on space since the entry itself would not need to contain a contextual reference.

Wright's suggestion that multiterm words should be given in the spoken order of the word, e.g., *digital component*, not *component, digital* or *digital record*, not *record, digital*, can be problematic. The advantage of using the natural order is that it provides the user a seemingly easier way to search for an entry as it is used in communication. But as Pearce-Moses (1989:3) points out in his paper on term entries for thesauri, it may be easier if compound terms that

begin with a generic term are inverted, e.g., *Art, School of* rather than *School of Art*. One reason for this is that the term entry would be listed in related areas and not lost amongst unrelated terms. Despite this possible drawback, I would still recommend listing the entries, simple or complex, in their natural order as this will make the search easier for users both familiar and unfamiliar with the related terms.

2.2.1.2 Glossary of Terms from ANSI/NISO Z39.19-1993

To illustrate the standards described by Wright (2002), some entries from the first part of the first page of the Glossary of Terms as it appears, verbatim, in ANSI/NISO Z39.19-1993, page 35, are presented below.

(11) **Glossary of Terms**

> The following definitions are derived from several related standards and specialized dictionaries, but all have been modified for the purposes of this standard. References to other standards appear in brackets. Underlined terms within definitions are defined elsewhere in this glossary.
>
> **assignment indexing**. An indexing method by which descriptors or subject headings from a controlled vocabulary are selected by a human or computer to represent the concepts or items in a document. The terms may or may not occur in thedocument. Cf. **derivative indexing**.
> **authority file**. A set of records of established headings and the cross-references to be made to and from each heading, often citing the authority for the preferred form or variants. Types of authority files include name authority files, subject authority files, and thesauri. autoposting. *See* **up-posting**.
> **bound term**. A descriptor consisting of a compound term. (The term was originated by Mortimer Taube in his *Studies in Coordinate Indexing,* vol. 1, 1953, p. 43.)
> **broader term.** A descriptor to which another descriptor or multiple descriptors are subordinate in a hierarchy. The relationship indicator for this type of term is BT.
> **candidate term.** A term considered for admission into a thesaurus because of its potential usefulness as a descriptor. Cf. **provisional term.**

2.2.2 Standards for a Thesaurus

2.2.2.1 Pearce-Moses 1989

Pearce-Moses discusses the choices involved in creating a thesaurus and talks about the benefits and drawbacks of a natural order for multiterm entries in a thesaurus with examples from his experience creating a photographic thesaurus. He emphasizes that clear and detailed scope notes provide the user with precise information on how the term entry is used.

Pearce-Moses (p. 1) describes a thesaurus as ‘a controlled vocabulary to facilitate information retrieval.’ This vocabulary is ‘a list of terms’ in which the ‘relationships between them’ is laid out for the user. Some of the relationships that hold between terms can be one of

synonyms, homonyms, broader terms, narrower terms, etc. Some of the problems that accompany the construction of a thesaurus is the fact that language and the meaning of words are ever changing, and one person's connotation may be different from another's, which highlights the importance of ensuring that a term entry's definition is consistent and reliable. One way to ensure this consistency is to draw the list of vocabulary items from the literature of the field (p. 2). However, this is not always sufficient if the field is new and evolving. In his work creating a photographic thesaurus, he found that 'historically important terms never achieved wide use, while other terms are notoriously corrupt for imprecise use'. One solution to imprecise terminology is to provide a single integrated list that's used for both simple and complex terms allowing the end user to join terms together according to how precise the terminology needs to be. This approach requires instructions in the thesaurus on how to assemble terms. One of the primary reasons for constructing a thesaurus in this way is it allows for a very simple entry search and at the same time, the user can build as much complexity as required into the term. As Pearce-Moses (p. 3) points out in his work with the photographic thesaurus, 'The vast majority of photographs can be described by approximately fifty terms. Most collections wouldn't worry about the variant of those fifty terms; they'll be happy as a clam in mud with "silver gelatin" without worrying that its [sic] more precisely "silver gelatin bromoiodide selenium toned photographic transparency."

As for the word order of term entries, Pearce-Moses (p. 3) notes that inverting forms helps keep related terms together, for example, *Art, School of* and *Art, Museum of* would both appear in close proximity to each other in a thesaurus while *School of Art* would appear with other types of schools, which may or may not be related to art. Therefore, he suggests that complex entries that begin with a generic term should be inverted. Pearce-Moses recommends using an inverted word order for complex terms that begin with a generic word. However, a problem that may arise is determining what constitutes a 'generic' word. For instance, should 'Art of Flying' be listed under 'Art' or 'Flying'? Given the recommendation for complex terms to be listed in their natural order by Wright (2002) above and ANSI/NISO Z39.19-1993. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri* (p. 13), I recommend entering complex terms in their natural (direct) order. Inverted order may be used in addition, with USE cross-references included directing the user to the term entry in its natural order.

Another important point that Pearce-Moses (p. 4) makes about constructing a thesaurus pertains to scope notes, which can be equally applied to a glossary and dictionary. He states 'A thesaurus without scope notes is merely a list of words". Information in the scope notes can help the user know how to use the term precisely. The scope note is vital in terms of increasing

'the accuracy of the terminology by defining it and restricting its precise use' and '[i]f this information is carefully researched, the thesaurus becomes more than a tool for authority, but acquires a secondary value on its own as an *authoritative* source.' As such, I strongly recommend providing hyperlinks in the scope notes that direct the user to the source document(s) where the term entry is found.

Pearce-Moses (p. 3) points out that imprecise terminology is a problem in the construction and use of a thesaurus. His suggested solution to provide a single integrated list that's used for both simple and complex terms depending on the needs of the user has advantages and disadvantages. One advantage is that the problem between entering multiterm entries in a natural order, e.g., *School of Art* versus inverted word order, e.g., *Art, School of*, is resolved since the user can select which terms to combine to refine the search. Second, users who require only a simple term, e.g., "silver gelatin" or complex terminology, e.g., precisely "silver gelatin bromoiodide selenium toned photographic transparency" should be satisfied. There are some disadvantages to having a single integrated list of terms as well. First, this approach requires instructions in the thesaurus on how to assemble terms, perhaps adding a level of difficulty at the user end. Second, asking the user to conjoin the necessary terms to find the definition implies that the user already knows the necessary terminology; however, this may not be the case, especially if the user is someone from outside the field. Because of this, I recommend providing as much information in the term entry as required by someone who needs more information, not less. Cross references may be used in the entry if there are alternate, less precise terms that can be used. A link from the entry to the document source will provide information on where and how the term is used, which will elucidate when the longer or shorter term is to be used.

2.2.2.2 ANSI/NISO Z39.19-1993: *Guidelines for the Construction, Format, and Management of Monolingual Thesauri.*[18]

To quote NISO:

> NISO's thesaurus standard (ANSI/NISO Z39.19) is the primary source for guidance on the construction of controlled vocabularies.
>
> The core strength of ANSI/NISO Z39.19 is that it offers a standardized way to organize many kinds of information, including conceptual and taxonomic information. Even if a specialized notation is preferred for relationships, the principles of design are still applicable. This core strength has made Z39.19 the leading national standard for thesaurus construction and the de facto international standard.[19]

[18] http://www.niso.org/standards/resources/Z39-19.html (under revision as of 2003).
[19] http://www.niso.org/committees/MT-info.html.

Standards for the construction of a monolingual thesaurus are set out in ANSI/NISO Z39.19-1993. Since this is a sixty-eight page document, all of the standards cannot be listed here; however, some of the more relevant criteria are discussed starting with the list of abbreviations, codes and conventions used in ANSI/NISO Z39.19-1993. Definitions for most of the unfamiliar terms can be found in the Glossary of Terms at the end of this report.

(12) **Abbreviations, Codes and Conventions used in ANSI/NISO Z39.19-1993**

Abbreviations (Thesaurus Codes) and Relationship Indicators (page xii)

BT	=	broader term
BTG	=	broader term (generic)
BTI	=	broader term (instance)
BTP	=	broader term (partitive)
GS	=	generic structure
HN	=	history note
NT	=	narrower term
NTG	=	narrower term (generic)
NTI	=	narrower term (instance)
NTP	=	narrower term (partitive)
RT	=	related term
SEE	=	equivalent to U (USE)
SN	=	scope note
TT	=	top term
U	=	use
UF	=	used for
UF+	=	used for … and …
USE+	=	use … and …
X	=	see from (equivalent to UF); reciprocal of *see*

Conventions:
Descriptors are in **boldface**.

Words discussed in the text are enclosed in ‘single quotation marks’.

One of the main features of a thesaurus is the display of semantic relationship that links terms. This linking is achieved through *relationship indicators*, such as the ones listed above. ANSI/NISO Z39.19-1993 (p. 13) describes three types of semantic relationships that all thesauri should have. These are: the *equivalence relationship*; the *hierarchical relationship*; and the *associative relationship*. Within each relationship there is reciprocity, that is, ‘every relationship indicated between term A and term B has a corresponding relationship from term B to term A’ (p. 13). The following is a table indicating the abbreviations for relationship indicators as shown on page 13:

(13) **Table of Abbreviations for Relationship Indicators** (ANSI/NISO Z39.19-1993:13)

Relationship	Relationship Indicator	Abbreviation
Equivalence	**USE**	none or U
(Synonymy)	**USED FOR**	UF
Hierarchy	**BROADER TERM**	BT
	NARROWER TERM	NT
Association	**RELATED TERM**	RT

The equivalence relationship is one that exists between preferred (the descriptor) and nonpreferred terms when each term refers to the same concept (p. 15). The descriptor is used as a substitute for the other terms that relate equivalent concepts. Synonyms contain cross-references to the descriptor. Equivalence relationships are expressed by using the following conventions:

(14) **Equivalence Relationship** (ANSI/NISO Z39.19-1993:15)

U or **USE**, which leads from a nonpreferred (entry) term to the descriptor,
and
UF or **USED FOR**, the reciprocal, which records entry terms leading to the descriptor.

Examples:

Aves USE **birds**	outline USE **shape**
birds UF Aves	**shape** UF outline

Equivalence relationships cover three types of terms: synonyms, quasi-synonyms and lexical variants.

The second type of relationship that exists in a thesaurus is the hierarchical relationship. According to ANSI/NISO Z39.19-1993 (p. 16) the hierarchical relationship is:

> 'the primary feature that distinguishes a systematic thesaurus from an unstructured list of terms, such as a glossary. It is based on degrees or levels of superordination and subordination, where the superordinate descriptor represents a class or a whole, and subordinate descriptors refer to its members or parts.'

The relationship indicators BT (Broader Term), which refers to the superordinate descriptor, and NT (Narrow Term), which refers to the subordinate descriptor, are used to express reciprocity in the hierarchical relationship.

(15) **Hierarchical Relationship** (ANSI/NISO Z39.19-1993:17)

mammals	**vertebrates**
BT **vertebrates**	NT **mammals**

Hierarchical relationships cover three situations: the generic relationship, the whole-part relationship and the instance relationship.

The generic relationship 'identifies the link between a class and its members or species' (p. 17). To test whether a generic relationship exists between two terms, one can form the statement "[narrower term] is a [broader term]" (p. 17). The relationship indicators used to show a generic relationship between terms is BT/NT, or if one wishes to be more specific, the indicators BTG (Broader Term (generic)) and NTG (Narrower Term (generic)) can be employed.

(16) **Generic Relationship** (ANSI/NISO Z39.19-1993:17)

rats	**rodents**
BTG **rodents**	NTG **rats**

As we can see, rats are rodents and rodent is the larger class to which rats belong.

The whole-part relationship pertains to 'situations in which one concept is inherently included in another regardless of context, so that the descriptors can be organized into logical hierarchies, with the whole treated as a broader term.' (p. 17). Again, specific relationship indicators can be utilized in order to represent the whole-part relationship between terms. These are: BTP (Broader Term (partitive)) and NTP (Narrower Term (partitive)).

(17) **Whole-Part Relationship** (ANSI/NISO Z39.19-1993:17)

central nervous system	**nervous system**
BTP **nervous system**	NTP **central nervous system**

The final situation covered by the hierarchical relationship is the instance relationship. This relationship shows 'the link between a general category of things or events, expressed by a common noun, and an individual instance of that category, often a proper name.' (p. 18). Again, specific indicator abbreviations can be used to show the precise nature of the relationship. These are: BTI for Broader Term (instance)) and NTI (Narrower Term (instance)).

(18) **Instance Relationship** (ANSI/NISO Z39.19-1993:18)

fairy tales
NTI **Cinderella**
Rumpelstiltskin

While specific relationship indicators such as those described above can be quite useful, my recommendation is to avoid the extra layer of complexity in the indicators and to simply use BT and NT when necessary.

The final semantic relationship to be discussed is the associative relationship.[20] This type of relationship exists between terms that are conceptually or semantically related but are not equivalent or hierarchical (p. 19). The use of one term generally implies another and 'one of the terms is often a necessary component in any explanation or definition of the other'.

(19) **Associative Relationship** (ANSI/NISO Z39.19-1993:19)

boats	**ships**
BT **vehicles**	BT **vehicles**
RT **ships**	RT **boats**

The next part of this section examines the standards that pertain to the scope, form and choice of descriptors as described in §3 of ANSI/NISO Z39.19-1993, unless otherwise stated.

(20) **A List of Standards Pertaining to the Scope, Form and Choice of Descriptors in a Monolingual Thesaurus (ANSI/NISO Z39.19-1993:2-10)**

1. Each descriptor represents a single concept.
2. Descriptors are restricted according to the meanings within the domain of the thesaurus.
3. Homographs should be avoided when possible.
4. The meaning of homographs can be clarified using standardized qualifiers in parentheses as part of the term entry/descriptor, e.g.,
   cranes (birds)
   cranes (lifting equipment)
5. Parenthetical qualifiers should be used with homographic terms even when the term entry 'is used in only one of its meanings within a thesaurus' (p. 3).
6. Use a compound term (if one exists) as a descriptor rather than a parenthetical record, e.g.,
   phonograph records *rather than* records (phonograph)
7. Scope notes are used to provide clarification as to the usage of the descriptor and tell the user the chosen meaning of the term. The scope note is not part of the term entry.
   illuminations
   SN Includes both the ornamental decoration and the illustrations in manuscripts as well as in some printing books, if done by hand.
8. Proper nouns are used for unique entities, e.g.,
   Artistotle
   World Health Organization
9. Descriptors should be nouns or noun phrases.
10. Infinitive verbs without the particle 'to' should not be used. Nominal forms or gerunds should be used instead, e.g.,
    catalysis *rather than* catalyze
    cookery *rather than* cook

[20] The relationship indicator RT (related term) can be used more extensively than as it is discussed here; however, the example shown here is a good representation of how the indicator is used to link terms in a thesaurus.

I disagree with the use of nominal forms in place of verbs since deriving the verb form from the noun form is not always easy, e.g., cook from cookery, distill from distillation. This is especially true if the search is not a native speaker of English. For this reason, I advocate using the infinitive form of the verb without the particle 'to.'

11. Avoid using articles before descriptors, e.g.
    arts *rather than* **the arts**
12. If the article is an integral part of the descriptor, use it in its natural order, otherwise invert the article, e.g.,
    Computer Place, The
    El Salvador
13. Count nouns should be expressed in the plural, e.g.,
    books
    singers
14. Noncount (mass) nouns and names of abstract concepts should be expressed in the singular, e.g.,

Mass nouns	Abstract concepts
copper	**Buddhism**
snow	**love; anger**

I recommend the use of singular noun forms for both all nouns unless the descriptor is normally used in the plural form, e.g., *scissors, pants*, etc. in order to maintain consistency among the thesaurus and the glossary and dictionary documents. However, it should be noted that the *NASA Thesaurus* (see below) also uses nouns in the plural form.

15. 'The most widely accepted spelling of words should be adopted. If variant spellings exist and are commonly recognized, each should be entered in the thesaurus, and a cross-reference should be made from the nonpreferred to the preferred form.' (ANSI/NISO Z39.19-1993:7)
    Romania *rather than* Roumania
    theater *rather than* theatre
16. Only well established abbreviations and acronyms should be used as descriptors, e.g.,
    DNA *rather than* deoxyribonucleic acid
    lasers *rather than* light amplification by stimulated emission of radiation
17. Full forms should be used rather than unfamiliar or less widely used abbreviations.
18. Descriptors should be written in lowercase characters except when the descriptor is a proper name, trade name and other terms that are usually capitalized.
19. 'Descriptors, nonpreferred terms, relationship indicators, and textual notes should be typographically distinguished. Suggested topographic specifications are: lightface or *italics* for nonpreferred terms, all capitals for relationship indicators such as USE, and **boldface** for descriptors.' (ANSI/NISO Z39.19-1993:24)

2.2.2.3 The *NASA Thesaurus*[21]

The *NASA Thesaurus* contains the subject terms from the documents in the NASA STI Databases. The original terminology was drawn from NASA's indexing vocabulary in the 1960s with additional terminology coming from the *DOD* (Department of Defense) *Thesaurus of Engineering and Scientific Terms*. There are two volumes and a supplement. *Volume 1*[22] is the *Hierarchical Listing With Definitions* and *Volume 2*[23] is the *Rotated Term Display*. The *Supplement*[24] contains updates, which take place every six months,, including all the new terms and associated hierarchies added since the cut off for the 1998 edition of the *NASA Thesaurus*. While various guidelines were used in the original construction of the *NASA Thesaurus*, currently the standard used is that of the National Information Standards Organization *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*, ANSI/NISO Z39.19-1993. Some of the terms used and the conventions that follow are from the Nomenclature and Conventions section of *Volume 1* of the thesaurus, pages vi-vii.

(21) **Nomenclature and Conventions used in the pdf version of the *NASA Thesaurus***

i. **Postable Terms**. 'Subject terms that have been approved for use in indexing, and thus can be 'posted.' In *Volume 1*, postable terms are shown in non-italic type.'

ii. **Nonpostable Terms**. 'Terms that are included for cross reference information and cannot be used for indexing. In *Volume 1*, nonpostable terms are set in italics.'

iii. **Term Selection**. Subject terms are drawn from aerospace literature.

iv. **Noun usage**. Term entries are given in noun forms.

v. **Singular vs. Plural**. In contrast to other lexicographical documents, the *NASA Thesaurus* uses plural forms for term entries. The singular form is used for non-count nouns and for terms that refer to unique entities, such as the *Mariner 10 Space Probe*.

vi. **Term Length**. 'No more than 42 characters, including spaces, are used for any postable subject term. Various words in longer terms are sometimes truncated. Full expanded forms of such truncated terms are generally included in the scope notes.'

vii. **Term Ambiguity**. 'When subject terms have more than one meaning in aerospace usage, or where distinction between terms must be made, clarification is provided in one of two ways:

a) Parenthetical qualifying expressions or glosses are added, becoming part of the subject term. For example:

sizing (shaping)
sizing (surface treatment)

[21] All the information regarding the *NASA Thesaurus* (its two volumes and supplement) come from the online pdf version available at http://www.sti.nasa.gov/thes1.htm.

[22] http://www.sti.nasa.gov/98Thesaurus/vol1.pdf.
[23] http://www.sti.nasa.gov/98Thesaurus/vol2.pdf.

b) Scope notes are also added for explanation or definition; they do not become part of the subject term. For example:

rotational states

SN (LIMITED TO MOLECULAR ENERGY LEVELS-EXCLUDES ROTATIONAL DYNAMICS OF VEHICLES OR OTHER BODIES)'

viii. **Word Order**. Multiterm or complex term entries are shown in their natural word order.

ix. **Abbreviations and Acronyms**. Some commonly used abbreviations and acronyms are used as postable terms. 'USE cross references are made from the unabbreviated forms.' e.g.,

Orbiting Solar Observatory
USE **OSO**

x. **Synonyms**. 'When candidate subject terms are true synonyms, one is chosen to be the valid, or postable term, and the other is provided with a USE cross reference.' UF = Used For (see below)

columbium
USE **niobium**

niobium
UF columbium

xi. **Array Terms**. 'Subject terms with meanings either too broad or ambiguous for effective indexing or retrieval of information, have been designated array terms and carry the following scope note (USE OF A MORE SPECIFIC TERM IS RECOMMENDED - CONSULT THE TERMS LISTED BELOW). Relationships with other postable terms are shown by the Related Term (RT) reference only' (see below). An infinity symbol (∞) precedes the array term.

∞ **beams**
SN (USE OF A MORE SPECIFIC TERM IS RECOMMENDED - CONSULT THE TERMS LISTED BELOW)
RT beams (radiation)
beams (supports)

xii. **Identifiers**. Identifiers are 'terms that designate unique entities' and are treated as regular terms with complete cross references. 'Most identifiers are proper nouns and many include a numeric or alphabetic designation for a particular model or item. As a general rule, identifiers are added to the Thesaurus only if they have an important relation to the aerospace sciences.'

F-111 aircraft
UF LASV
TFX aircraft
GS attack aircraft
.fighter aircraft
..**F-111 aircraft**
General Dynamics aircraft
.**F-111 aircraft**
Grumman aircraft
.**F-111 aircraft**
jet aircraft
.turbofan aircraft
..**F-111 aircraft**
supersonic aircraft
.**F-111 aircraft**
RT ∞ aircraft
mission adaptive wings
variable sweep wings

[24] http://www.sti.nasa.gov/98Thesaurus/thessup3.pdf.

Volume 1, the *Hierarchical Listing With Definitions*, contains all subject terms, approved USE cross references, and provides the hierarchical structure for each term. This hierarchical listing is meant to serve as an orthographic authority for upper/lowercase forms of all terms and cross references. To get an idea of what a listing looks like in the *Hierarchical Listing With Definition,* here examples from the Nomenclature and Conventions section of *Volume 1* are provided below. The cross reference relationships and their application as they are used in *Volume 1* are also provided.

(22) **The Cross-Reference Relationships used in the *Hierarchical Listing With Definition* of the *NASA Thesaurus,*** page viii; the cross references SN, DEF and ∞ are shown in the *Supplement*, page v..[25]

Cross Reference	**Notation**
Broader Term	GS[26]
Narrower Term	GS
Related Term	RT
Use	USE
Used For	UF
Scope Note	SN
Definition	DEF
Array Terms	∞

Broader Term - a term with a more inclusive concept. 'In the Generic Structure (GS)[27], broader terms appear above and to the left of the term referenced.'

reentry communication
GS telecommunication
. space communication
..spacecraft communication
…**reentry communications**

Telecommunications, space communication and *spacecraft communication* are broader terms than the term entry *reentry communication*.

Narrower Term - a term with a more specific concept. 'In the Generic Structure (GS), the narrower terms appear below and to the right (indented) of the term referenced.' There are an unlimited number of narrower terms.

GS **scanners**
.Coastal Zone Color Scanner
.horizon scanners
.infrared scanners
.ocean color scanner
.optical scanners
..flying spot scanners
..multispectral bandscanners
…Thematic mappers (Landsat)
.ultrasonic scanners

[25] The number of cross reference relationships used in the *NASA Thesaurus* is smaller than that used in ANSI/NISO Z39.19-1993.
[26] ANSI/NISO Z39.19-1993.(page 22, §6.2.2.2) states 'Multiple levels of hierarchy may be indicated without BT/NT notation, by using the abbreviation GS (generic structure) with indention and punctuation marks such as period and colon as cues to the levels of hierarchy.' They suggest looking at the *NASA Thesaurus* for an example of this.
[27] See the definition for *Generic Structure* in the Glossary of Terms below.

Coastal Zone Color Scanner, Horizon scanners, infrared scanners, ocean color scanners, *optical scanners*, and *ultrasonic scanners* are narrow terms than *scanners,* i.e., they are specific types of scanners. *Flying spot scanners*, *multispectral band scanners*, and *Thematic Mappers (Landsat)* are narrower than *optical scanners* and *scanners*, meaning that not only are they specific types of scanners they are also specific types of optical scanners.

Related Terms - 'the two terms are conceptually associated, but not equivalent or generically related. The RT relationship is reciprocal'.

radar equipment **radio equipment**
RT radio equipment RT radar equipment

Use (USE) - 'indicates that the term is not 'postable', i.e., not a valid term, and that the term or terms adjacent to the USE indicator should be used instead'. Nonpostable terms are in italics.

jet airstreams
USE **jet streams (meteorology)**

Used For (UF) - 'This relation is the reciprocal of the USE cross reference and indicates that the term listed above the UF indicator is a valid or 'postable' term, and term or terms adjacent to the UF indicator are nonpostable'.

jet streams (meteorology)
UF *jet airstreams*

The following is an example of a hierarchical listing with definition from the *NASA Thesaurus.* The numbered list provides information on how to interpret the listing.

(23) **Typical Hierarchical Listing with Definition** (*NASA Thesaurus*, *Volume 1,* p. xi)

1 **microbursts (meteorology)**
2 *(added January 1993)*
3 SN (EXCLUDES IONOSPHERIC RADIATION MICROBURSTS)
4 DEF A strong, localized downdraft that strikes the ground creating an outflow of severe winds near the ground that diverge radially from the impact point.
5 UF *bow echo microburst events*
6 GS meteorology
.micrometeorology
..**microbursts**
(meteorology)
7 RT aviation meteorology
flight hazards
thunderstorms
vertical air currents
wind shear

Key
1. Postable Term
2. Date Added
3. Scope Note
4. Definition
5. Use For Term
6. Generic Structure
7. Related Term

ANSI/NISO Z39.19-1993 uses the relationship indicator HN to indicate when a term was added to the document and the history of modifications to any part of the entry. I would recommend using the relationship indicator HN in the InterPARES 2 Project Thesaurus.

Volume 2, the *Rotated Term Display* is a reference tool to help locate terms in *Volume 1*. It contains the postable (preferred) terms and nonpostable (nonpreferred) terms found in the *Hierarchical Listing* arranged in a KWIC (key-word-in-context) index. The postable terms 'microbursts' and 'maps' appear as follows:

(24) **Rotated Term Display Listing** (the *NASA Thesaurus*, *Volume 2,* p. 210)

microbursts (meteorology)

(25) **Rotated Term Display Listing** (the *NASA Thesaurus*, *Volume 2,* p. 200)

astronomical **maps**
lunar **maps**
radar **maps**
radar clutter **maps**
relief **maps**
weather **maps**
____meteorological charts

Notice that there is no definitional information here. The user is expected to consult *Volume 1* of the thesaurus in order to locate the definition. As with *Volume 1*, cross references from nonpostable to postable terms are used in *Volume 2*. Words with little semantic or searchable value, termed *stopwords*, are omitted in the rotated term display. Similarly, numeric strings and non-alphanumeric characters are not included in the sort. To disambiguate homographs, words that are spelled alike but having different meanings and origins, *glosses* are used. In the *NASA Thesaurus*, a gloss is 'a word or words enclosed in parentheses at the end of a term' (*Volume 2*, p. v). In the rotated term display, parentheses have been ignored so that glosses appear with related terms:

(26) **Glosses in the Rotated Term Display Listing** (the *NASA Thesaurus*, *Volume 2,* p. v)

geometrical theory of diffraction
analytic **geometry**
angles **(geometry)**
Bose **geometry**
chords **(geometry)**
variable **geometry** structures

2.2.3 Standards for a Register

ISO 10241-1992 International Terminology Standards (Preparation and Layout) might be a suitable set of standards for the InterPARES 2 Register since it is likely related to creating a word list. However, this is difficult to say since I do not currently have the document. The following quote is from the abstract:

> Establishes rules for the preparation: preliminary work (needs analysis, target group, subject delimitation, sources, number of concepts, choice of languages, schedule) and working procedure (collecting and recording terminological data, establishing the term list, concept fields and concept systems, formulating definitions, coining and selecting terms), as well as for the terminography (structuring of entries, order of entries, indexes, graphic representation, bibliography). Does not deal with changes that may be necessary when an International Standard is adopted as a national standard.[28]

2.2.4 Standards for a Dictionary

2.2.4.1 The *Oxford English Dictionary*

The standard style in the online *Oxford English Dictionary* shows headwords. For each headword, there is a page that lists the number of entries found for the headword. Another page lists the various senses,[29] that is, concepts or meanings, associated with the chosen entry of the headword. For example, there are seven entries for the word *slate*. Here is an example of the partial listing of the term *slate* in its first sense.[30]

(27) *OED 2* entry for 'slate'

> **slate**
>
> **slate** sl*e*it, sb.1 Also (chiefly north. and Sc.) 4- sclate (5 sclathe), sklate (9 sklet); 5-9 sclait (6 sclayt), 5 sklaytt, 6 sklaitt, 6-9 sklait. ad. OFr. *esclate* fem., in the same sense as *esclat* masc., whence slat sb.1 After *c* 1630 the forms with *scl-, skl-* are exclusively northern and Scottish. The earliest example of the form occurs in sense 3, but the development of the senses must have been the same as in slat sb.1
>
> 1.
>
> a. A thin, usually rectangular, piece of certain varieties of stone which split readily into laminæ (see 4), used especially for the purpose of covering the roofs of buildings.

[28] Available at http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=18278&ICS1=1&ICS2=20&ICS3=.
[29] The *OED 2* online definition of 'sense' is 'The meaning or signification of a word or phrase; also, any one of the different meanings of a word, or that which it bears in a particular collocation or context.' (http://www.chass.utoronto.ca/patbin/new/oed-idx?fmt=entry&type=entry&byte=412331313)
[30] Taken from http://www.chass.utoronto.ca/patbin/new/oed-idx?fmt=entry&type=entry&byte=431155163.

Also freq. called a *roofing-slate*, and with distinguishing terms, as *blue, green, grey, white slate(s)*. For the older Sc. use of the word see skaillie.

<alpha>

•1455 *Anc. Cal. Rec. Dublin* (1889) 284 Sclatys, bordes, gottorys, schall ly upon the key be the spase of xx. dayes.

•1456 Sir G. Haye Law Arms (S.T.S.) 228 As a sclate fell of a hous and slewe a man.

•C. 1540 Boorde The Boke for to Lerne B ij, Many tyles or sklates.

•1584 Reg. Privy Council Scot. III. 678 Becaus thair is sklaittis, lyme, sand and tymmer to be transportit..to his said palice.

•1832 Carrick in Whistle-Binkie (1890) I. 213 Some o' them gaed ower the sklates As weel's your dainty dow.

<beta>

•1530 Palsgr. 706/1, I sclate a house with stone slates.

•1570 Levins Manip. 39/12 A Slate, tyle, tegula, later.

•1600 J. Pory tr. Leo's Africa iii. 202 The roofe is couered with certaine blacke stones or slates.

•1662 Gerbier Principles 36 The Roof..should be covered either with Lead or blew Slates.

•1745 Season. Advice Protestants 17 The Houses, that were formerly in good Repair, and cover'd with Slates, decay.

•1758 J. S. Le Dran's Obs. Surg. (1771) 65 A Slate fell upon her Head from the Top of an House.

•1811 Farey Derbyshire I. 428 At Sheffield these white and grey Slates are exclusively used.

•1841 James Brigand xix, The house was built of cold grey stone, with a roof of slates.

•1889 H. C. Seddon Builder's Work (ed. 2) 231 Ordinary roofing slates are sold by the number... Some of the largest sized slates are.sold by the ton, and hence are called ton slates or weight slates.

b. A slab of slate, +or other stony substance; a laminated rock.

•1601 Holland Pliny xvii. viii. I. 506 It [the Columbine marl] will resolve and cleave into most thin slates or flakes.

•1601 R. Johnson Kingd. & Commw; 27 The Irish Ocean, a sea so shallow, and so full of rocks and slates [etc.].

•1876 Encycl. Brit. IV. 500 He [the slater] supplies sawn slates for shelving in larders and dairies.

The electronic dictionary entry has the headword in boldface, the pronunciation, part of speech, homonym number, usage labels, variations on the spelling, followed by a list of the senses and

some etymological information for the various subheadings within each of the given senses. Quotations are included to illustrate how each sense of the word is used. Where appropriate, a citation contains the date, author, work (i.e., title) and the quotation. Importantly, all entries use a consistent structure like the one mentioned and shown here. As expected, this is clearly and necessarily much different from the glossary style used in the InterPARES 1 and 2 glossaries, which have the headword in bold and italics and have much less information on the usage of the word.

2.2.4.2 Coward and Grimes (2000)

In this section, I outline some of the details Coward and Grimes (2000) discuss when one creates a dictionary.

1. Identify the user audience (academic, local, general, government, mixed).
2. Coward and Grimes (page 70) recommend building a dictionary for a local audience so that it serves a variety of needs for a wide audience and is still useful to a narrower audience. In addition, one can supplement this with another document that provides additional information for a secondary audience, e.g., for an academic audience.
3. Trilingual dictionaries are not recommended. Instead, bilingual dictionaries with various pairings, e.g., French-English, German-English, German-French, etc., are preferred. One of the reasons is that semantic categories between languages are not always one-to-one (see page 71). That is, what may be a noun in one language may be a noun in another but a whole phrase in third. Another reason not to have a trilingual (or more) dictionary is that words from languages of radically differing cultures may not be easily intertranslatable.
4. A text-based lexicography, i.e., compiling a list of lexemes from a set of natural texts, such as the InterPARES documents, for instance, is a good source of headwords. In addition to this, words of a single semantic domain should be compared and contrasted with a subset of words within them in order to determine the precise differences in meaning, e.g., a list of emotion words such as *hate, loathe, despise* or *adore, love, cherish*, etc., have different connotations that may need to be addressed, especially in a bilingual dictionary, in order to ensure that the correct word is used in the correct context.
5. A minimal entry, the word and a gloss, is sufficient for a glossary, though it may also contain additional information such as when the entry was made, altered, the part of speech of the entry, etc.
6. The dictionary can be organized in terms of being root morphemes (structure-centric units) or surface-form lexemes (meaning-centric units (page 77). In determining which organization to use, it's important to keep the audience in mind. For instance, a linguist might be interested in a root-based dictionary for research purposes, but this would make it more difficult for someone without specific morphological knowledge to use. What is preferable is a lexeme-based dictionary that has the basic entries and includes information on roots and affixes (see page 78). In the case of verbs, the use of citation forms is preferred, e.g., *run*, not *runs, ran*, etc. See page 83 for list of advantages

2.2.4.3 *Merriam-Webster Dictionary*[31]

Information regarding the structure of the online *Merriam-Webster Dictionary* can be found at http://www.m-w.com/entries.htm. Some of the more relevant standards used in this dictionary follow. I will use the entry 'accuracy' to illustrate the standards.

(28) One entry found for **accuracy**.

Main Entry: **ac·cu·ra·cy**
Pronunciation: 'a-ky&-r&-sE, 'a-k(&-)r&-
Function: *noun*
Inflected Form(s): *plural* **-cies**
Date: 1662
1 : freedom from mistake or error : **CORRECTNESS**
2 a : conformity to truth or to a standard or model : **EXACTNESS** b : degree of conformity of a measure to a standard or a true value -- compare PRECISION 2a

(29) **Standards used in the online *Merriam-Webster Dictionary*[32]**

1. The number of entries found for the word is given. If there is more than one entry, a box below the number of entries is visible which allows the user to select an entry to view.
2. The headword is shown in boldface and lowercase, as is the case with the *OED 2* online. If a word is normally used with the first letter in uppercase, then it is listed as such in the *M-W Dictionary*.
3. The headwords are listed in alphabetical order letter by letter, ignoring intervening spaces or hyphens, e.g., *battlement* precedes *battle royal*
4. Headwords containing an Arabic numeral are alphabetized as if the numeral were spelled out, e.g., *three-color, 3-D, three-decker*
5. The abbreviations for the parts of speech are given for the each entry of the headword in italics.
6. Homographs are listed separately and are distinguished by superscript numeral before the entry:
   [1]pine . . . *n*
   [2]pine *vi*

7. Homographs are listed historically with the one first used in English entered first.
8. The pronunciation of the headword is provided, as is the syllabification.
9. Spelling Variations. When there is more than one acceptable spelling of a headword, all variants are provided. Here, both variants for 'ochre' appear in boldface, indicating that either variant is acceptable, neither is preferred, e.g.,
   Main Entry: **ocher**
   Variant(s): or **ochre**

'If two variants joined by or are out of alphabetical order, they remain equal variants. The one printed first is, however, slightly more common than the second:
plow *or* **plough**

When another spelling is joined to the main entry by the word also, the spelling after also is a secondary variant and occurs less frequently than the first:
can-cel-la-tion *also* **can-cel-ation**

[31] Online at http://www.m-w.com/cgi-bin/dictionary.
[32] http://www.m-w.com/entries.htm.

> Secondary variants belong to standard usage and may be used according to personal inclination. If there are two secondary variants, the second is joined to the first by *or*. Once the word *also* is used to signal a secondary variant, all following variants are joined by *or*:
>
> [1]**Shake-spear-ean** *or* **Shake-spear-ian** *also* **Shak-sper-ean** *or* **Shak-sper-ian**
>
> 10. The inflected form of the headword is provided, e.g., **-cies**. Note that an non-native English speaker may not know how to use this inflected form, thus the entire inflected for should be different when there is a change in the spelling of the word.
> 11. Etymology. The etymology of a headword is normally provided except where no etymology can be found or where the etymology is deemed unnecessary. The etymology, when provided, is given in boldface square brackets **[ ]** .
> 12. Date refers to the earliest recorded use in English. The date refers to the meaning of the sense as it's given in the dictionary, which may not be the same as the earliest meaning of the word.
> 13. Each sense of the headword is provided, along with synonyms for that particular sense.
> 14. Synonyms (from http://www.m-w.com/cross.htm)
> 'A cross-reference immediately following a boldface colon is a synonymous cross-reference. It may stand alone as the only definitional matter for an entry or for a sense or subsense of an entry; it may follow an analytical definition; it may be one of two synonymous cross-references separated by a comma:
>
> **gar-ban-zo** . . . *n* . . . : CHICKPEA
> **[1]ne-glect** . . . *vt* . . . **1** : to give little attention or respect to : DISREGARD
> **[2]main** *adj* . . . **1** : CHIEF, PRINCIPAL
>
> A synonymous cross-reference indicates that a definition at the entry cross-referred to can be substituted as a definition for the entry or the sense or subsense in which the cross-reference appears.'
>
> The directional cross-reference *compare* directs the user to a word that provides further information.

With regards to spelling variations in cases where there is no preferred spelling, I would recommend providing all variations as to eliminate the problem of choosing between an American or British spelling for some words. However, if there is a preferred spelling, both spellings may be listed with a cross reference from the nonpreferred to the preferred spelling, e.g.,

(30) **behavior**
UF *behaviour* (Brit.)

2.2.4.4 *The Canadian Oxford Paperback Dictionary*

While it is useful to refer to the standards used in online dictionaries such as the *OED 2* and *Merriam-Webster* since the InterPARES lexicographical documents will be in electronic form, it is also useful to have standards for a printed dictionary. In this section I present the standards used in the *Canadian Oxford Paperback Dictionary* (2002). I use the homographs for 'batch' to illustrate these standards.

(31) *Canadian Oxford Paperback Dictionary* (2002) entry for 'batch'

batch[1] •*n.* **1** a number of things or person s forming a group or dealt with together. **2** an instalment (*have sent off the latest batch of proofs*). **3** a quantity produced by one operation, or the amount of material necessary for this (*a batch of doughnuts*). **4** (*attrib.*) using or dealt with in batches, not as a continuous flow (*batch production*). **5** *Computing* a group of records processed as a single unit. • *v.tr.* arrange or deal with in batches.
batch[2] •*v.intr. & tr. N Amer, Austral., & NZ* (esp. in **batch it**) live alone and keep house for oneself, esp. temporarily.

(32) **Some standards used in the *Canadian Oxford Paperback Dictionary* (2002)**[33]

1. The headword is shown in boldface and lowercase, in roman type, unless the word is of foreign origin and not naturalized in English, in which case it is shown in large, bold, italic type, e.g., ***Gemeinschaft***
2. Headwords are listed in alphabetical order letter by letter, ignoring intervening spaces or hyphens, e.g., *battlement* precedes *battle royal*
3. Headwords containing an Arabic numeral are alphabetized as if the numeral were spelled out, e.g., *three-color, 3-D, three-dimensional*
4. The parts of speech are given for the each entry and its derivative in italics.
5. Homographs are listed separately and are distinguished by a superscript numeral after the entry:
   pine[1] . . . *n.*
   pine[2] *v.intr.*

6. 'Definitions are listed in a numbered sequence in order of comparative familiarity and importance, with the most current and important sense first. They are subdivided into lettered sense (**a, b,** etc.) when these are closely related or call for collective treatment' (p. vi).
7. The pronunciation (but not the syllabification) of the headword is given in phonetic transcription using the International Phonetic Alphabet but only for rare or difficult to pronounce words.
8. Spelling Variations. When there is more than one acceptable spelling of a headword, variants are provided. in boldface in brackets before the definition, 'or are given their own headword entry and cross-reference when these are more than seven entries away from the main headword. The main headword represents the most common form in Canadian usage' (p. iii).

a. **phony** *informal* (also **phoney**) •*adj.* (**phonier, phoniest**) **1** sham; counterfeit; fake. **2** insincere. •*n.* (*pl.* **-ies** or **-eys**) a phony person or thing. **phoniness** *n.*

b. **artefact** *var. of* ARTIFACT.

c. **artifact** *n.* a product of human art and workmanship. **artifactual** *adj.*

If only one of the senses has a variant spelling, the variant will appear at that sense.
Regional spelling variants are indicated in brackets after the entry, e.g.,

fibreglass *n,* (also esp. *US* **fiberglass**)

10. The inflected form of the headword is given after the headword only when the form is irregular or difficult, e.g.,
char[1] *v.tr. & intr.* (**charred, charring**)

11. No etymological information is provided.
12. Each sense of the headword is provided without synonyms.
13. Illustrative, contextual examples are shown in italics in brackets.
14. Cross-references and symbols (¶ indicates usage; indicates defined phrases or idioms;

• introduced a part of speech

2.2.5 Summary and Recommendation

What is readily apparent if one compares a headword entry in the *Oxford English Dictionary,* the *Merriam-Webster Dictionary* and the *Canadian Oxford Paperback Dictionary*, is that the amount of information decreases from the first to the last. I recommend adopting most of the standards used in the *Merriam-Webster Dictionary* as it provides a sufficient amount of information. However, I believe that the formatting style used in the *Canadian Oxford Paperback Dictionary* is easier to follow than Webster's. Since etymological information is not necessary in the *InterPARES 2 Dictionary*, though the date a term begins to be used is, this type of linguistic information is best omitted. Additionally, linguistic information such as syllabification seems equally unnecessary. I would recommend giving inflected forms in full for the ease of native and non-native English speakers alike.

3. InterPARES' Application of the Standards

In this section, I comment on how the InterPARES 2 Project follows or veers from the lexicographical document styles.

3.1 Register

I cannot comment on how the Register of Terms and Phrases accords with the practices set out by the project since I have not found a set of standards for a register within the online InterPARES documents.

3.2 Glossary

The criteria for a Web-based glossary set out in Wright (2002) have been followed in the *InterPARES 2 Terminology Database* with the following exceptions:

(33) a. The terminology entries in the InterPARES 2 Terminology Database begin with an uppercase letter rather than a lowercase letter, e.g., ***Access.***

According to Wright (2002), the entire headword in a glossary is in boldface and lower case, so that 'accuracy' should be displayed as **accuracy** rather than *Accuracy*.

b. Grammatical information is not included with each entry but only for entries where the word can belong to more than one syntactic category, e.g., ***Act*** (n.) and ***Act*** (v.). My

[33] http://www.m-w.com/entries.htm.

advice would be to follow Wright's recommendation to include syntactic information for each entry.

c. Terms that are used within definitions or notes (entailed terms) are not always listed separately in the glossary. For example, in the *InterPARES 2 Terminology Database* entry for the term Dynamic Record (August 25, 2003), the entailed term *spreadsheet* is part of the definition of the entry *dynamic record* but is not itself listed in the glossary:

Term:

- **Dynamic Record**

Definitions:

- Dependent upon data that might have variable instantiations and be held in databases and spreadsheets. [Ross, 2000] (Duranti-MCRI 412-2001 InterPARES 2 detailed proposal forum- line 140)

My recommendation is to list all the entailed terms separately in the glossary with hyperlinks to the term entry in which the entailed term is used.

d. In some instances, the term entries in the *InterPARES 2 Terminology Database* use the terms in the definition of the terms. For instance, the entry for ***Act*** (v.) is 'To give origin to an act (n.) (UBC Project Book Glossary)' (August 25, 2003).

Whenever possible, the term entry should not be used in its definition.

e. Earlier inconsistencies in the formatting of the glossary have been remedied

f. The most recent version of the InterPARES 2 Project Terminology Database (20030908) contains very few scope notes. This solves the problem of some of the inconsistency present in previous versions of the glossary; however, it is now not possible to determine the source of the entries.

I recommend including scope notes in the glossary with hyperlinks to the source documents.

3.3 Dictionary

Since the recommendations and standards regarding the construction of a dictionary were made after the creation of the *InterPARES 2Project Dictionary*, and no set of guidelines for a dictionary existed within the online InterPARES documents at the time of writing this report, a comparison between it and the recommendations included here is difficult. For this reason, I repeat the list of recommendations for creating a dictionary from page one and compare an entry from the *InterPARES 2 Project Dictionary* by way of showing a similar entry from the online *Merriam-Webster Dictionary*.

(34) **Recommendations for the InterPARES 2 Project Dictionary**

Dictionary

definition: a subject lexicon of English archival terms employed within InterPARES 2 and the archival community at large, in which each word-entry has a headword, a

part of speech, a list of known senses, and one or more sources for each sense (but without etymology, pronunciation, and illustrative quotations)
model: *Merriam-Webster Dictionary*
method: transfer Glossary intact, adding a source indication (i.e., that each sense is InterPARES 2); add headwords, senses, and source attributions from a select list of current archival subject lexicons
published as: Web-based database

The following dictionary headword 'accessibility' as it appears in the *InterPARES 2 Project Dictionary* is shown below. Following this is the headword 'accessible' as it appears in the online *Merriam-Webster Dictionary* since there was no separate headword 'accessibility.' In the *Webster* entry, I have omitted the pronunciation, the date the term is first used in its current sense and I put the grammatical category next to the headword rather than on a separate line below the headword which is more in keeping with a paper version of a dictionary.

(35) **InterPARES 2 Project Dictionary entries for 'accessibility' and 'accessible'**

Accessibility
The availability of archival material for consultation. *(SAA Glossary)*

The availability of archival material for consultation as a result of legal authorization and the existence of finding aids. *(SLAIS Glossary)*

Within computing, may also be used in a more to describe the physical process of retrieving information from storage media.*(Richard Pierce-Moses)*

Accessible
See accessibility

(36) **The word 'accessible' contained within the headword 'accessible' in the online *Merriam-Webster Dictionary*** (http://www.m-w.com/cgi-bin/dictionary). (Underlined words are hyperlinked to separate entries in the dictionary.)

One entry found for **accessible**.

Main Entry: **ac·ces·si·ble** (*adj.*)
Date: date of entry; or date of first use in the InterPARES documents, for example
1 : providing access
2 a : capable of being reached <*accessible* by rail>; *also* : being within reach <fashions at *accessible* prices> **b** : easy to speak or deal with <*accessible* people>
3 : capable of being influenced : **OPEN**
4 : capable of being used or seen : **AVAILABLE**
5 : capable of being understood or appreciated <the author's most *accessible* stories> <an *accessible* film>
- **ac·ces·si·bil·i·ty** *noun*
- **ac·ces·si·ble·ness** *noun*
- **ac·ces·si·bly** *adverb*

There are a few observations worth making regarding the entries for 'accessibility' and 'accessible' in the *InterPARES 2 Project Dictionary*. First, cross-references such as 'see' are not used in dictionary entries. Second, the headword should begin with a lowercase letter and

not be italicized. Third, each sense of the headword should be numbers as it is in the *Webster Dictionary*. Finally, parts of speech should be included next to the headword.

3.4 Thesaurus

As yet, there is no InterPARES 2 thesaurus on which to comment at the time of writing this report.

4. Conclusion

The set of recommendations provided at the beginning of this report reflects the standards generally accepted and in use by lexicographers. As such, they will be instrumental in helping to achieve the reliability, accuracy and authenticity necessary in the InterPARES 2 Project's lexicographical documents.

APPENDIX

1. *The InterPARES Glossary* (from http://www.interpares.org/reports.htm). The link http://www.interpares.org/documents/InterPARES Glossary 2002-1.pdf where this document is located was broken as of July 17, 2003. However, this document is also available at http://www.interpares.org/_private/glossary.cfm. No date is given for this document.

2. *The InterPARES 2 Project Glossary*. July 14, 2003. http://www.interpares.org/rws/display_file.cfm?doc=ip2_glossary.pdf&CFID=2033&CFTOKEN=37935291.

3. *The InterPARES 2 Project Glossary*. September 18, 2003. http://www.interpares.org/rws/display_file.cfm?doc=ip2_glossary(20030908).pdf.

4. *The InterPARES Glossary* by Ken Hannigan, presented June 2002 at College Park, MD. http://www.interpares.org/documents/interpares_glossary_tf.pdf (link located at http://www.interpares.org/papers.htm).

5. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. http://www.interpares.org/book/index.htm. There is no year specified for this document, though the glossary contained within says December 2001, therefore, I assume this is the date for the whole document.

6. *International Research on Permanent Authentic Records in Electronic Systems (InterPARES): Experiential, Interactive and Dynamic Records,* by Luciana Duranti 412-2001. http://www.interpares.org/rws/display_file.cfm?doc=InterPARES_2_detailed_proposal.pdf.

7. The *Glossary Committee Report* http://www.interpares.org/book/interpares_book_h_part5.pdf, Part Five of *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* at http://www.interpares.org/book/index.htm December 2001.

8. Email correspondence between Luciana Duranti and Jonathan Furner. July 09, 2003.

9. *The InterPARES 2 Project Register of Terms and Phrases*. July 17, 2003. http://www.interpares.org/rws/display_file.cfm?doc=ip2_register.pdf&CFID=2033&CFTOKEN=37935291.

10. InterPARES Glossary System: *System Integration Strategy*. version 1.0. http://www.interpares.org/rws/display_file.cfm?doc=IGS_SystemIntegrationStrategy_Draft_v1.pdf. April, 20, 2000.

11. *InterPARES 2 Project. Organizational Policy. Draft 9*. http://www.interpares.org/rws/display_file.cfm?doc=ip2_organizational_policy(10-2002).pdf. October, 2002.

12. InterPARES Glossary System. *System Overview*. version 1.0. http://www.interpares.org/rws/display_file.cfm?doc=IGS_SystemOverview_Draft_v1.pdf. April 20, 2000.

13. *Terminology Cross-domain Concordance Report*. http://www.interpares.org/rws/display_file.cfm?doc=ip2-terminology_glossary_concordance_report.pdf. No date is provided for this document.

14. Email correspondence from Luciana Duranti to ip2-terminology@interchange.ubc.ca. July 20, 2003.

15. *InterPARES 2 Project Dictionary*. http://www.interpares.org/rws/display_file.cfm?doc=ip2_dictionary(20030908).pdf. September 19, 2003.

GLOSSARY OF TERMS

Associative Context
A context that is used in order to show that the term and the concept are connected, but provides little detail about the concept. (Wright 2002:4)

Broader Terms
'Broader terms are used for index entries that are "higher" or less specific in the conceptual hierarchy associated with the main term.'
http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=thesaurus&#HierarchyDisplay

Count Nouns
'Count nouns are names of objects or concepts that are subject to the question 'How many?' but not 'How much?'. (ANSI/NISO Z39.19-1993:6)

Defining Context
Explains the context but is not a complete definition. (Wright 2002:4)

Descriptor
'A type of heading that is a term chosen as the preferred expression of a concept in a thesaurus.' (ANSI/NISO Z39.19-1993, p. 35)

Explanatory Context
'Includes explanatory material, but is not as complete as a definition or defining context.' (Wright 2002:4)

Generic Structure
'A thesaurus format that indicates all hierarchical levels of descriptors within an alphabetic display by means of codes, indention, and/or punctuation marks.' (ANSI/NISO Z39.19-1993. p. 36)

Gloss
'A word or words enclosed in parentheses at the end of a term.' (The *NASA Thesaurus, Volume 2*, p. v)

Headword
The subject of the dictionary entry or article. (*OED 2* online)

History Note
'A note in a term record in a thesaurus that provides the date of entry of a descriptor as well as the history of modifications to its scope, relationships, etc.' (ANSI/NISO Z39.19-1993:36)

Indexing
'Indexing is the process by which subject terms or classification symbols are assigned to concepts dealt with in documents. It includes any system in which the selection and organization of indexing terms call for human intellectual decisions, regardless of whether computer assistance is also used to store and manipulate these terms or to identify documents to which certain terms or combinations of terms have been assigned'. (ANSI/NISO Z39.19-1993. p. 1)

Lexeme
A word-like grammatical form intermediate between morpheme and utterance, often identical with a word occurrence; a word in the most abstract sense, as a meaningful form without an assigned grammatical role; an item of vocabulary. (*OED 2* online)

Lexical Variant
Lexical variants are not synonyms but rather 'different word forms for the same expression'. Types of lexical variants include direct vs. inverted word order, spelling variants, full names vs. abbreviations, etc. (ANSI/NISO Z39.19-1993:16)

Narrower Terms
'Narrower terms are used for index entries that are "lower" or more specific in the conceptual hierarchy associated with the main term.'
http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=thesaurus&#HierarchyDisplay

Noncount (Mass) Nouns
'Noncount nouns are names of materials or substances that are subject to the question 'How much?' but not 'How many?'. (ANSI/NISO Z39.19-1993:6)

Non-Preferred Term
'The synonym or quasi-synonym of a preferred term. A non-preferred term is not assigned to documents, but is provided as an entry point in a thesaurus or alphabetical index, the user being directed by an instruction (for example USE or SEE) to the appropriate preferred term; sometimes known as "non-descriptor." (ISO 5964-1985: 3)

Preferred Term
A term used consistently when indexing to represent a given concept; sometimes known as "descriptor." (ISO 5964-1985: 3)

Quasi-Synonym
'A term whose meaning is not exactly synonymous with that of another term, yet which may nevertheless be treated as its equivalent in a thesaurus.' (ANSI/NISO Z39.19-1993:37)

Related Terms
'Related terms are used for index entries that are laterally associated with the main term. They are provided to suggest search concepts for alternative or additional retrieval beyond the main term.'
http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=thesaurus&#HierarchyDisplay

Relationship Indicators
'A word, phrase, abbreviation, or symbol identifying a semantic relationship between terms. Examples of relationship indicators are USE, UF (used for), and RT (related term).' (ANSI/NISO Z39.19-1993:38)

Scope Note Display
'A scope note is used to restrict or expand the application of a descriptor, to distinguish between descriptors that have overlapping meanings in natural language, or to provide other advice on term usage to either the indexer or the searcher. A scope note should state the chosen meaning of a descriptor; it may also indicate other meanings that are recognized in natural

language, but which have been deliberately excluded from the controlled vocabulary.' ANSI/NISO Z39.19-1993:3.
'Scope notes are written by the database producer. They usually tell when the term was first used in indexing, how it has been applied, and other information relevant to your search. Important: scope notes are available only in databases that produce scope notes for their vocabulary. Not all databases will have scope notes available.'
http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=scope&

Situational Context
'Orients the concept and the term to the subject field or subfield, but doesn't contain explanatory or definitive information.' (Wright 2002:4)

Stopwords
'Certain words having questionable access value (such as *and*, *of*, *in*, etc.).' The *NASA Thesaurus, Volume 2*, page v. http://www.sti.nasa.gov/98Thesaurus/vol2.pdf.

Term
'One or more words designating a concept.' (ANSI/NISO Z39.19-1993:38)

Thesaurus Entry
'This is the term whose hierarchy is currently being displayed.'
http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=thesaurus&#HierarchyDisplay

Top Term
'The broadest descriptor in a thesaurus hierarchy, sometimes indicated by the abbreviation TT.' (ANSI/NISO Z39.19-1993:38)

Tree Hierarchy
'The Tree Display shows the selected term in the context of its conceptually broader and narrower terms in each vocabulary branch of the Tree.' 'Its broader terms appear above and at the left margin, and its narrower terms appear below and indented. The numbers in parentheses indicate the quantity of documents in the database that are indexed with each term.' http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=tree&

Used For Terms
'Used For Terms are non-postable synonyms to a main Thesaurus entry. They are displayed to provide information that may help in understanding the scope of a term. This is an informational entry only, and as such has no documents associated with it.' http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=thesaurus&#HierarchyDisplay

Use References
'Non-postable terms listed in the thesaurus will show a "Use" reference. This reference shows the term that is used for indexing the searched term, along with the number of documents [postings] associated with the referenced term.'
http://www.lib.utsystem.edu/ovidweb/ovidweb.cgi?S=&HC=thesaurus&#HierarchyDisplay

BIBLIOGRAPHY

ANSI/NISO Z39.19-1993. *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*. http://www.niso.org/standards/resources/Z39-19.html

Atkins, B.T.S. (ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators.* Tübingen: Niemeyer (Lexicographica Series Maior 88).

Béjoint, H. 1994. *Tradition and Innovation in Modern English Dictionaries*. Oxford: Clarendon Press.

Canadian Oxford Paperback Dictionary. 2000. Alex Bisset (ed.). Don Mills, ON:Oxford University Press.

Coward, David F. and Charles E. Grimes. 2000. *Making Dictionaries. A Guide to Lexicography and the Multi-Dictionary Formatter. SIL International:Waxhaw, North Carolina.* http://www.sil.org/computing/shoebox/MDF_2000.pdf. p. 67.

de Bessé, Bruno, Nkwenti-Azeh, Blaise and Juan C Sager. 1997. *Glossary of Terms Used in Terminology. Terminology* 4:1. John Benjamins**.** 1997. pp. 119-156. http://www.termilat.info/public/env408.rtf.

ISO 690-1987. Documentation - Bibliographic references - Content, form and structure.

ISO 690-2-1987. Information and documentation - Bibliographic references - Part 2: Electronic documents or parts thereof.

ISO 2788-1986. Documentation-Guidelines for the Establishment and Development of Monolingual Thesauri

ISO 5964-1985. Documentation-Guidelines for the establishment and development of multilingual thesauri.

Landau, Sidney I. 2001. *Dictionaries: The Art and Craft of Lexicography*. (Second edition.) Cambridge: Cambridge University Press.

NASA Thesaurus. 2003. http://www.sti.nasa.gov/98Thesaurus/vol1.pdf. NASA Center for AeroSpace Information. Hanover, MD.

Pearce-Moses, Richard. 1989. 'A rose is a rose is a rose: But a ferrotype isn't always a tintype or an energiatype'. Paper presented at the Society of American Archivists Annual Conference. http://members.cox.net/pearce-moses/rose.htm.

Svensén, B. 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.

Wright, Sue Ellen. 2002. *Web-Glossary FAQ*. Kent State University. Institute for Applied Linguistics. http://appling.kent.edu/ResourcePages/TSA-2003/TSAWeb/Final%20lineup/Terminology_Management_FAQ.htm.