International Encyclopedia of Language and Linguistics, Third Edition

Article Title

Dictionaries and Encyclopedias: Relationship

Author Contact Information

Keith Allan keith.allan@monash.edu

Abstract

The lexicon/dictionary is properly that part of an encyclopedia which stores information about the formal, morphosyntactic, and semantic specifications of listemes. The most important thing the lexicon/dictionary does is identify the concepts that comprise the salient properties of its typical denotation. Etymological and stylistic information, for instance, are encyclopedic data that are not strictly a part of the lexicon even though they must be closely networked with it. Similarly, encyclopedic data on the denotata of listemes must be closely networked with the lexicon/dictionary entries. An encyclopedia functions as a structured data-base containing exhaustive information on many, potentially all, branches of knowledge. To reflect reality, no language has one single unique all-encompassing lexicon/encyclopedia, instead being a network of modules.

Keywords

dictionary, encyclopedia, lexicon, listeme, network, decision procedure, wetware

Key points

- The functions of a dictionary/lexicon and those of an encyclopedia are defined.
- The characteristics of dictionaries and encyclopedias are described.
- The manner in which the dictionary/lexicon networks with the encyclopedia is elaborated and exemplified first from the audience perspective and then from the speaker/writer's perspective.

Glossary

Listeme: language expression whose meaning is not determinable from the meanings, if any, of its constituent morphemes or lexemes.

Morphosyntactic specification: morphological and syntactic properties of the listeme that identify what is traditionally called its 'part of speech'.

Denotatum: what the language expression is normally used to refer to in some real or imagined world.

Model: theoretical construct that represents a real (cognitive) process and its constituents.

Neurons: fundamental units of the brain and nervous system; cells responsible for receiving sensory input from the external world and for sending commands out.

Wetware: the human brain.

Synapses: each neuron has a few to hundreds of thousands of synaptic connections with other neurons.

Decision process: step-by-step process of choosing amongst alternative possibilities.

Introduction: Dictionary and lexicon and the relationship with an encyclopedia

Dictionaries such as the *Oxford English Dictionary* are publications created (not necessarily consciously) as partial models of the mental lexicon; thus, *dictionary* is to be understood as "(partial) model of the mental lexicon". (For a comprehensive brief history of dictionaries, see Hanks 2013.) A lexicon (sc. mental lexicon) is a bin for storing the meanings of so-called *listemes* (the term was first used by Di Sciullo and Williams 1987), those language expressions whose meaning is not determinable from the meanings, if any, of its constituent morphemes or lexemes. An encyclopedia functions as a structured data-base containing exhaustive information on many, potentially all, branches of knowledge, i.e. it represents a large data set. Consequently, Wikipedia (https://www.wikipedia.org) and the *Encyclopedia Britannica* are (partial) models of human knowledge.

A listeme's dictionary entry normally includes (a–c):

- a. Formal (phonological and graphological) specifications.
- b. Morphosyntactic specifications: properties such as the inherent morphosyntactic (lexical) category of the item; necessary subcategorization such as the conjugations of verbs; regularities and irregularities (e.g. the past tense of English strong verbs such as *drank* and *thought*); constraints on range (e.g. –ize is suffixed to nouns (atomize) and adjectives (legalize)).
- c. Semantic specifications identify the senses of a listeme; a sense describes the concepts that comprise the salient characteristics of the typical denotatum of that item (see Prototype semantics). The form of the semantic specification depends on the chosen metalanguage (see Cognitive Semantics; Dynamic Semantics; Frame Semantics; Lexicon Project; Lexical Conceptual Structure; Natural Semantic Metalanguage).

The denotation of a language expression α is what α is normally used to refer to in some possible (not necessarily real but imaginable) world. The reference of a language expression is what the

speaker/writer/signer is using the language expression to talk about – be it intensional, extensional, or non-existent. Thus, the sense of *I crashed my car yesterday* is "the speaker did something which caused severe damage to his or her automobile the day before the utterance was made". 'I' denotes the speaker; 'my car' denotes some kind of automobile driven by the speaker, etc. If, on Saturday 3 May 2025, Max says *I crashed my car yesterday* and his car is a Tesla, then Max is referring to himself, his Tesla, and the event of his crashing the Tesla on Friday 2 May 2025.

Important for linguistic semantics is the question: How much information is it necessary to include in a complete semantic representation in a lexicon/dictionary? Comprehension of a text is often tested by having the audience summarize the text and/or answer questions on it. Attempts in the field of artificial intelligence to program a machine to interpret a text so as to answer questions on it or to provide a summary for it reveal that the project requires input from what Schank and Abelson, 1977 call 'scripts', Lakoff, 1987 'idealized cognitive models', and Barsalou, 1992; Fillmore, 1975 & 1982; Fillmore and Atkins, 1992; Minsky, 1977 'frames'. These hypothetical constructs are not identical, but they - along with AI chatbots like ChatGPT, Bard, Co-pilot and other robotic processing automata – all call extensively upon encyclopedic knowledge. The normal practice before the 1980s was to favor parsimonious dictionary/lexical knowledge against elaborated encyclopedic knowledge, but things have changed. Haiman, 1980: 331 claimed 'Dictionaries are encyclopedias' which is certainly true of some existing dictionaries, for example The New Grove Dictionary of Jazz (Kernfield, 1994) is more encyclopedia than lexicon. New Oxford Dictionary of English (Pearsall & Hanks, 1998) includes 4500 place names, 4000 biographical entries, and 3000 other proper names. Such dictionaries function not only as an inventory of listemes but also as a cultural index to the language and the collective beliefs of its speakers – thus closely corresponding to human lexical abilities and traditional encyclopedias. Today most dictionaries are online but even their print versions can offer hypertext links to online encyclopedic information.

A dictionary/lexicon is an integral part of an encyclopedia

Because of a hearer's ability to 'shadow' a text very rapidly – that is, to begin understanding it and drawing appropriate inferences milliseconds after the speaker has uttered it (Marslen-Wilson, 1985 & 1989), there has to be a cognitive path from the listeme to directly access encyclopedic information about a denotatum. Consequently, the lexicon entry is one access point into the isomorphic set of encyclopedia entries, all of which are activated by recognition of the listeme. If the encyclopedia represents a large data set, then the lexicon constitutes an integral component of that encyclopedia.

Listemes characterize concepts

The most important thing the lexicon/dictionary does is identify the concepts that comprise the salient properties of its typical denotation (Allan, 2024). Information about denotata (potential referents) is stored in their encyclopedia entries. It is this information from which the senses of isomorphic listemes are abstracted. Such abstraction from particulars is evident in the ontogenetic development of listemes by children. Clark, 1973 reports a child's extension of bird to any moving creature (sparrows, cows, dogs, cats), moon to any round object, bow-wow to things that are bright, reflective and round (?based on the dog's eyes) such as a fur piece with glass eyes, pearl buttons, cuff links, a bath thermometer. The same process of metaphorical extension operates when adults name novel objects. For example, metaphorical extensions of body-part terms are recorded in many unrelated languages throughout the world (Allan, 1995; Anderson, 1978; Lakoff & Johnson, 1980; Johnson, 1987; Svorou, 1994; inter alios): foot of a mountain, head of school, back of a house, back seat, back of the hand, sole and heal of a shoe, fingers of a glove, and so on. When the brassiere was developed in the late nineteenth century, a term was required for the individual breast containers: in English their shape and function gave rise to [bra] cup; in Dutch kopje, diminutive of kop "cup"; in Polish miseczka, diminutive of miska "bowl"; in Czech kośíček, diminutive of koš "basket"; in Hungarian kosár "basket"; Spanish copa "drinking vessel"; in Turkish kap "container", or kup "bowl"; in French bonnet "cap". In all these languages, the particular term was adopted because it characterizes the configuration and function (salient properties) of the denotatum.

The idealized model of the encyclopedia and lexicon requires a heuristic updating facility. Suppose that Z has only ever encountered female bearers of the name *Beryl*, so that the semantic specification in Z's lexicon entry is "bearer of the name *Beryl*, normally a female". If Z comes across the name *Beryl* used of a cisgender man, not only is Z's encyclopedia expanded, but also Z's mental lexicon entry will be updated to "bearer of the name *Beryl*, normally a female, but attested for a male".

Further evidence for the link between dictionary/lexicon and encyclopedia

It is incontrovertible that encyclopedic data is called on in:

- a. metaphors like She's a gazelle / a tiger;
- the extension of a proper name like *Hoover* to denote vacuum cleaners and vacuum cleaning (I assume that, because many proper names are shared by different namebearers, there must be a stock of proper names located either partially or wholly in the dictionary);
- c. explaining the formation of the verb bowdlerize from the proper name Bowdler;
- d. making and understanding statements like (1) and (2):

- (1) Caspar Cazzo is no Pavarotti
- (2) Harry's boss is a bloody little Hitler!
- (1) implies that Caspar is not a great singer; we infer this because Pavarotti's salient characteristic was that he was a great singer. (2) is abusive because of the encyclopedic entry for the name *Hitler* cf. (3). Such comparisons draw on biodata that is appropriate in an encyclopedia entry for the person who is the standard for comparison, but not a dictionary, which should identify the salient characteristics of the typical name-bearer, cf. the *Oxford English Dictionary* for **Hitler** 'One who embodies the characteristics of Hitler; a dictatorial person' but not (contra Frege, 1892) encyclopedic information about a particular name bearer any more than the dictionary entry for *dog* should be restricted to a whippet or poodle rather than the genus as a whole). (3) offers a brief encyclopedia entry for *Hitler*:
- (3) **Hitler** *proper name* for **Adolf Hitler**, primarily responsible for World War II and castigated for being a fascist dictator who was ultimately responsible for the liquidation of six million Jews, and countless Slavs, Romani, homosexuals, and others whom he regarded as socially undesirable. Comparisons with Hitler imply a ruthless dictatorial manner, someone willing to murder millions of the people he disapproves of.

A more carefully composed encyclopaedic entry for *Hitler* than (3) would contain more genuine historical fact. It might also allow for a neo-Nazi assessment of Hitler, which would be far more positive than (3).

Significantly, different (groups of) individuals make different interpretations of facts, and the prejudices of language users are just as relevant to a proper account of language understanding as the true facts (if 'truth' can even be nonsubjectively determined, Allan, 2022 & 2023). This raises a problem: Will the lexicon/dictionary and encyclopedia have to institutionalize the mainstream stereotype? Or should the prejudices of different groups within the community be represented? The latter may be morally (and politically) dispreferred, but as a model of reality it is the preferable option.

And there is the outstanding question: Should the lexicon/dictionary contain every word in the language and the encyclopedia contain exhaustive information on all branches of knowledge, or should they be modular, rather like a collection of human minds? For instance, a medic's lexicon is full of medical jargon and a medic's encyclopedia contains medical knowledge unknown to the average patient; a botanist knows more about plants than most medics do and has the lexicon to talk about that knowledge; and so forth through the community for different interest groups. No one is sure what constitutes the common core of a lexicon/dictionary nor of an encyclopedia. Even if

such core data can be comprehensively identified, they will have to be connected to specialist modules with jargon dictionaries and specialist encyclopedias. And a modular system would need an archiving device to facilitate information incrementation and update to both lexicon/dictionary and encyclopedia. There are two reasons for favoring multiple lexicons and encyclopedias: first it would putatively model individual human capacities; second, it would divide data and processing into manageable chunks. As already shown in (3), the encyclopedia will contain entries for people under their proper names. In reality, as recognized above, individuals in a community will have different mental encyclopedias, i.e. partially different information (cf. Katz, 1977; Murphy, 2000).

Modeling the networked dictionary/lexicon and encyclopedia

The basic lexicon entry is a networked triple consisting of a formal representation, tagged here $_{f000}$, linked to a morphosyntactic category which is also linked with the semantic component of the triple, the latter tagged $_{5000}$: e.g. $_{f001}N_{51002}$ (where N=noun). For example, (4) models a partial entry for the common and proper names robin, Robin and Robyn. The subscripts are identity tags comparable with a strong password and, in the model, may be represented by any sequence of symbols (e.g. $Cj1^{\circ}3\%5@7^{\circ}2$) so long as they are consistent. These identity tags model neurons in the wetware. The following commands are constructs in the model decision process: & (conjunction), XOR (exclusive disjunction), ELSE (if not), ELIF (else if), GOTO, OUTPUT, TENABLE (this sense is/seems applicable to the denotatum, so OUPUT the data), NEXT ITEM. These commands model synapses in the wetware.

(4) f₁₄₉ **robin** /'rɒbɪn/ XOR f₁₅₀**Robin** /'rɒbɪn/ XOR f₁₅₁**Robyn** /'rɒbɪn/ & s₂₄₉ "small songbird": IF
TENABLE OUTPUT & GOTO NEXT ITEM, ELSE GOTO f_{150,151}N_{s250,251} & "bearer of the name
f₁₅₀Robin, either male or female"_{s250-251} IF TENABLE OUTPUT & GOTO NEXT ITEM, ELSE GOTO
"bearer of the name f₁₅₁Robyn, normally female"_{s251} IF TENABLE OUTPUT & GOTO NEXT ITEM
ELSE OUPUT 'Inappropriate Name' & GOTO NEXT ITEM.

To clarify: the meaning assigned *robin* as the output from 's249 "small songbird": IF TENABLE OUTPUT & GOTO NEXT ITEM' is 'small songbird'. If none of the available senses of /'rpbɪn/ is applicable, then the label is inappropriate, communication has foundered, and the next item is likely to seek to resolve this.

(4) lacks the encyclopedic information needed to identify what species of the bird *robin* is relevant, which will depend on the geographical context: European, American, and Australian robins are each a different species with different coloring and behavior. Encyclopedic data about a proper name may incorporate information about celebrated name bearers.

Now consider the basic lexicon/dictionary entry for *cup* in (5).

- (5) $_{f100}$ **cup** /k $_{\Lambda}$ p/ & $_{f100}$ N $_{s200}$ & $_{s200}$ " $_{f100}$ cup is a hollow hemispheroid usually with a diameter greater than or equal to its depth"
- (5) does not differentiate the many possible senses of the listeme *cup* which derive from the salient properties of different kinds of denotata. These properties are specified in the encyclopedic data linked to this listeme, see (6). se000 is a link from the semantic component to the encyclopedic component of the entry.
- (6) f100cup /kAp/ & f100Ns200 & s200 "f100 cup is a hollow hemispheroid usually with a diameter greater than or equal to its depth": IF TENABLE OUTPUT & GOTO se201i-204i, ELSE OUTPUT 'Inappropriate Name' & GOTO NEXT ITEM.
 - se201i "drinking vessel": GOTO se201a XOR se201b XOR se201c XOR se201d ELIF se202 "prize": GOTO se202 ELIF se203i "garment": GOTO se203a XOR se203b ELIF se204 "eukaryote": GOTO se204"
 - se201a "f100 cup is a flat-bottomed hollow oblate hemispheroidal drinking vessel, an impermeable physical artefact (entity) with a vertical handle and a container with a capacity of about 250ml, it is a typical Western style cup for containing drinks such as tea or coffee, that is typically accompanied by a matching saucer: IF TENABLE OUTPUT & GOTO NEXT ITEM XOR GOTO se201b"
 - impermeable physical artefact with a vertical handle and a container with a capacity of about 125ml, it is a typical Western espresso style coffee cup (demitasse) that is often accompanied by a matching saucer: IF TENABLE OUTPUT & GOTO NEXT ITEM XOR GOTO se201c"
 - se201c "f100 cup is a flat-bottomed hollow oblate hemispheroidal drinking vessel, an impermeable physical artefact, a container with a capacity of about 125 ml, it is a typical Chinese style tea cup and/or a typical Middle-Eastern style tea or coffee cup: IF TENABLE OUTPUT & GOTO NEXT ITEM XOR GOTO se201d"
 - se201d "f100 cup is a flat-bottomed hollow tapered cylindrical drinking vessel, a container which is an impermeable physical artefact having a diameter less than its depth and made of water-proof paper, plastic, polystyrene, or similar material with a capacity between approximately 250–500ml, it is a throw-away cup that typically lacks a handle and invariably lacks a saucer": IF TENABLE OUTPUT & GOTO NEXT ITEM
 - se202 "f100 cup is an impermeable physical artefact offered as a prize for a race or athletic contest": IF TENABLE OUTPUT & GOTO NEXT ITEM

- se203a "f100 cup is a hollow hemispheroidal physical artefact of textile fabric that is one of a pair which constitute the principal parts of a brassiere, each cup being shaped to contain and support one of a woman's breasts": IF TENABLE OUTPUT & GOTO NEXT ITEM
- _{203b}"_{f100} *cup* is a hollow oblate hemispheroidal physical artefact worn as a shield by sportsmen to contain and protect male genitals": IF TENABLE OUTPUT & GOTO NEXT ITEM
- se204"f100cup is a hollow oblate hemispheroid that forms the woody seat of a naturally occurring entity that contains an acorn (its cupule), it is an acorn-cup": IF TENABLE OUTPUT & GOTO NEXT ITEM ELSE OUTPUT 'Inappropriate Name' & GOTO NEXT ITEM

There are in fact several more senses of *cup* (*Oxford English Dictionary*, 2024) so _{se204} is not the last possible sense of *cup*, but the point has been made as how the model of a listeme entry might work. (Matching morphosyntactic categories like determiners, prepositions, conjunctions, etc. with denotata is more complicated than for nouns, verbs, adjectives, and adverbs, but space does not allow for exemplification here).

The procedure demonstrated in (4), (5), and (6) starts from the formal specification and models audience procedure. A speaker, writer or signer proceeds from a concept of the denotatum in some such way as is modeled in (7).

- (7) IF denotatum x is "drinking vessel": GOTO se201a XOR se201b XOR se201c XOR se201d
 - se201a "x is a flat-bottomed hollow oblate hemispheroidal drinking vessel, an impermeable physical artefact (entity) with a vertical handle and a container with a capacity of about 250ml, it is a typical Western style cup for containing drinks such as tea or coffee, that is typically accompanied by a matching saucer: IF TENABLE OUTPUT f100 cup & GOTO NEXT ITEM ELIF GOTO se201b"
 - se201b "x is a flat-bottomed hollow oblate hemispheroidal drinking vessel, an impermeable physical artefact with a vertical handle and a container with a capacity of about 125ml, it is a typical Western espresso style coffee cup (demitasse) that is often accompanied by a matching saucer: IF TENABLE OUTPUT f100 cup & GOTO NEXT ITEM ELIF GOTO se201c"
 - se201c "x is a flat-bottomed hollow oblate hemispheroidal drinking vessel, an impermeable physical artefact, a container with a capacity of about 125 ml, it is a typical Chinese style tea cup and/or a typical Middle-Eastern style tea or coffee cup: IF TENABLE OUTPUT f100 cup & GOTO NEXT ITEM ELIF GOTO se201d"

se201d "x is a flat-bottomed hollow tapered cylindrical drinking vessel, a container which is an impermeable physical artefact having a diameter less than its depth and made of water-proof paper, plastic, polystyrene, or similar material with a capacity between approximately 250–500ml, it is a throw-away cup that typically lacks a handle and invariably lacks a saucer": IF TENABLE OUTPUT f100 cup & GOTO NEXT ITEM

And so forth.

To summarize

The encyclopedia is a large data base of general knowledge that includes lexical information traditionally found in dictionaries. A lexicon/dictionary is that part of an encyclopedia which stores information about the formal, morphosyntactic, and semantic specifications of listemes; for other views, see Boguraev & Briscoe, 1989; Butterfield, 2003; Hanks, 2013; Hanks, Lenarčič & McClure, 2022; Hüllen & Schulze, 1988; Jackendoff, 1975; 1995; 1997; Lyons, 1977; Mel'čuk, 1992; Pearsall & Hanks, 1998. To reflect reality, no language has one single unique all-encompassing lexicon/encyclopedia, instead being a network of modules.

References

- Allan, K (2022). 'Truth is what the context makes of it (La verdad es lo que el contexto hace de ella).'

 Claridades: Revista de Filosofía 14 (2), 15-33. Https://doi.org/10.24310/Claridadescrf.v14i2

 .13620.
- Allan, K (2023). 'Why truth is necessarily pragmatic.' *Intercultural Pragmatics* 20 (3), 251-267. Https://doi.org/10.1515/ip-2023-3003.
- Allan, K (2024). 'The characterization of concepts in a metalanguage for lexicographic semantics.' *Topoi* 43.
- Barsalou, L W (1992). 'Frames, concepts, and conceptual fields.' In Lehrer, A & Kittay, E (eds.) Frames, Fields, and Contrasts. Norwood NJ: Lawrence Erlbaum. 21–74.
- Boguraev, B & Briscoe, T (eds.) (1989). *Computational Lexicography for Natural Language Processing*. London: Longman.
- Butterfield, J (ed.) (2003). *Collins English Dictionary: Complete and Unabridged*. 6th edn. Bishopriggs (Glasgow): HarperCollins.
- Clark, EV (1973). 'What's in a word? On the child's acquisition of semantics in his first language.' In Moore TE (ed.) *Cognitive Development and the Acquisition of Language*. New York: Academic Press. 65–110.

- Di Sciullo, A-M & Williams, E (1987). On the Definition of Word. Cambridge MA: MIT Press.
- Fillmore, CJ (1975). 'An alternative to checklist theories of meaning.' In Cogen, C et al. (eds.)

 Proceedings of the First Annual Meeting of the Berkeley Linguistics Society. Berkeley: Berkeley
 Linguistics Society. 123–31.
- Fillmore, CJ (1982). 'Frame semantics.' In Linguistic Society of Korea (eds.) *Linguistics in the Morning Calm.* Seoul: Hanshin. 111–38.
- Fillmore, CJ & Atkins, BT (1992). 'Toward a frame-based lexicon: the semantics of RISK and its neighbors.' In Lehrer, A & Kittay, E (eds.) *Frames, Fields, and Contrasts*. Hillsdale: Lawrence Erlbaum. 75–102.
- Haiman, J (1980). 'Dictionaries and encyclopedias.' Lingua 50, 329-357.
- Hanks, P (2013). 'Lexicography from earliest times to the present.' In Allan, K (ed.) *Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press. 503-535.
- Hanks, P, Lenarčič, S & McClure, P (2022). *Dictionary of American Family Names (2 ed.)*. Oxford: Oxford University Press. Https://doi.org/10.1093/acref/9780190245115.001.0001.
- Hüllen, W & Schulze, R (eds.) (1988). *Understanding the Lexicon: Meaning, Sense, and World Knowledge in Lexical Semantics*. Tübingen: Max Niemeyer Verlag.
- Jackendoff, RS (1975). 'Morphological and semantic regularities in the lexicon.' *Language* 51, 639-671.
- Jackendoff, RS (1995). 'The boundaries of the lexicon.' In Everaert, M et al. (eds.) *Idioms: Structural and Psychological Perspectives*. Hillsdale NJ: Erlbaum. 133–65.
- Jackendoff, RS (1997). Architecture of the Language Faculty. Cambridge MA: MIT Press.
- Katz, JJ (1977). 'A proper theory of names.' Philosophical Studies 31, 1–80.
- Kernfield, BD (ed.) (1994). The New Grove Dictionary of Jazz. London: Macmillan.
- Lakoff, G (1987). Women, Fire, and Dangerous Things. Chicago: University of Chicago Press.
- Lyons, J (1977). Semantics. 2 vols. Cambridge: Cambridge University Press.
- Marslen-Wilson, W (1985). 'Speech shadowing and speech comprehension.' *Speech Communication* 4, 55–73.
- Marslen-Wilson, W (1989). 'Access and integration: projecting sound onto meaning.' In Marslen-Wilson, W (ed.) *Lexical Representation and Processing*. Cambridge MA: MIT Press. 3–24.

- Mel'čuk, IA (1992). 'Lexicon: An Overview.' In Bright W (ed.) *International Encyclopedia of Linguistics, Vol.2* New York: Oxford University Press.
- Minsky, M (1977). 'Frame-system theory.' In Johnson-Laird, PN & Wason, PC (eds.) *Thinking:* Readings in Cognitive Science. Cambridge: Cambridge University Press. 355–76.
- Murphy, ML (2000). 'Knowledge *of* words versus knowledge *about* words: the conceptual basis of lexical relations.' In Peeters, B (ed.) *The Lexicon-Encyclopedia Interface*. Oxford: Elsevier Science. 317-48.

Oxford English Dictionary (2024) http://www.oed.com

Pearsall, J & Hanks, P (eds.) (1998). New Oxford Dictionary of English. Oxford: Clarendon Press.

Schank, R & Abelson, RC (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale NJ: Lawrence Erlbaum.