

of the eternal problem of how to specify what an individual sense of a word 'means.'

The study of lexical semantics is less 'autonomous' than that of, say, phonology, or syntax. Especially if one takes a cognitive linguistic view, there is no clear dividing line between lexical semantics and the study of conceptual categories within cognitive psychology; and advances in one field tend to have repercussions in the other. Advances in certain areas of psycholinguistics can also be expected to throw light on word meaning. For instance, there is currently a developing body of work on the time course of semantic activation. The meaning of a word is not activated all at once when a word is recognized, and the details of the activation process cannot fail to have relevance to our understanding of the internal structure of a word's meaning.

One area of practical concern, which is poised for a major take-off, but is currently held back by lexical semantic problems, is the automatic processing of natural language by computational systems. The main problems are the complexity of natural meanings and their contextual variability. The work currently being done in this area can be expected to spill over, not only into general lexicography, but also into the linguistic study of word meanings.

See also: Dementia, Semantic; Lexical Processes (Word Knowledge): Psychological and Neural Aspects; Lexicology and Lexicography; Lexicon; Semantic Knowledge: Neural Basis of; Semantics

Bibliography

- Cruse D A 1986 *Lexical Semantics*. Cambridge University Press, Cambridge, UK
- Cruse D A 2000 *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press, Oxford, UK
- Levin B, Hovav Rappaport M 1992 Wiping the slate clean: A lexical semantic exploration. In: Levin B, Pinker S (eds.) *Lexical and Conceptual Semantics*. Blackwell, Oxford, UK, pp. 123–52
- Lyons J 1963 *Structural Semantics*. Cambridge University Press, Cambridge, UK
- Lyons J 1968 *Introduction to Theoretical Linguistics*. Cambridge University Press, Cambridge, UK
- Lyons J 1977 *Semantics*. Cambridge University Press, Cambridge, UK
- Nida E A 1975 *Componential Analysis of Meaning: An Introduction to Semantic Structures*. Mouton, The Hague, The Netherlands
- Taylor J R 1989 *Linguistic Categorization: Prototypes in Linguistic Theory*. Clarendon Press, Oxford, UK
- Ungerer F, Schmid H-J 1996 *An Introduction to Cognitive Linguistics*. Longman, London
- Wierzbicka A 1996 *Semantics: Primes and Universals*. Oxford University Press, Oxford, UK

D. A. Cruse

Lexicology and Lexicography

1. Introduction

Each language has a lexicon and a grammar, i.e., a set of elementary expressions and a set of rules according to which complex expressions are constructed from simpler ones. Some of these rules form complex words; others operate beyond the boundaries of the word, thus producing phrases and sentences. These distinctions, familiar from the days of the Greek grammarians, are not always clear cut, for at least two reasons. First, the notion of 'word' is not very well-defined (see also *Word Classes and Parts of Speech*). Second, there are complex expressions, whose meaning is more or less predictable from the meaning of its components, whereas this is not true for other complex expressions. The former are said to be 'compositional,' whereas the latter are 'lexicalized'; slightly different terms to characterize this opposition are 'productive' vs. 'idiomatic,' and 'free' vs. 'fixed'; in each case, the distinction is gradual. Lexicalization is rarely observed for inflected words (a possible exception are 'participles' such as *crooked in a crooked street*), but very frequent for compound words, such as *landlord* or *(to) withdraw*, or phrases such as *to kick the bucket*, which has a compositional as well as a lexicalized reading. Do lexicalized expressions belong to the lexicon of a language or to its grammar? There is no straightforward answer; their form is complex and rule-based, their meaning is not. Therefore, it is useful to take the term 'lexicon' in a somewhat broader sense; it contains all elementary expressions (lexicon in the narrower sense) as well as those expressions which are compound in form but not accordingly in meaning (see also *Lexicon*). The scientific investigation of the lexicon in this sense is usually called lexicology; it includes, for example, the historical development of the lexicon, its social stratification, its quantitative composition or the way in which some subfield is encoded in lexical items (e.g., 'terminology of hunting,' 'verbs of movement'). Lexicography, by contrast, deals with the compilation of dictionaries. There is considerable overlap between both disciplines, and in fact, not all authors make such a terminological distinction.

2. The Lexicon

The lexicon of a language is stored primarily in the head of its speakers, and for most of the history of mankind, it was only stored there. We do not know what form the 'mental lexicon' has (see also *Psycholinguistics: Overview*). There is agreement, however, that it consists of individual lexical units which are somehow interrelated to each other. There is no generally accepted term for lexical units. The familiar

term 'word' is both too broad and too narrow; one would not want to consider *goes* as a lexical unit, although it is a word, whereas expressions such as *(to) cut up* or *red herring* are lexical units but consist of several words. Other terms occasionally found are 'lexeme,' 'lemma,' or 'lexical entry,' but since these are also used in other ways, it is probably best to speak of lexical units.

It is important to distinguish between a lexical unit and the way in which it is named. The word *house* in a dictionary, followed by all sorts of explanations, is not the lexical unit—it is a name for such a unit. The lexical unit itself is a bundle of various types of properties. These include:

(a) phonological properties, which characterize how the lexical unit is pronounced; they include sounds, syllabic structure, lexical accent and, in some languages, lexical tone;

(b) graphematic properties, which characterize how the lexical unit is written (see also *Spelling*);

(c) morphosyntactic properties, which characterize how the unit can become part of more complex expressions; typically, they concern inflectional paradigm, word class, government relations, and others;

(d) semantic properties, which concern the 'lexical meaning' of the unit, i.e., the contribution which it makes to the meaning of the construction in which it occurs.

Some of these properties may be absent. This is most obvious for graphematic properties, since not all languages are written. There are a few lexical units without lexical meaning, such as the expletive *there* in English. Many linguists also stipulate 'zero elements,' i.e., units with morphosyntactic and semantic properties but without phonological properties (such as 'empty pronouns'); but these are normally treated in the grammar rather than in the lexicon.

Whereas these four types of properties are the defining characteristics of a lexical unit, other information may be associated with it, for example, its etymology, its frequency of usage, its semantic counterpart in other languages, or encyclopedic knowledge (thus, it is one thing to know the meaning of *bread* and a different thing to know various sorts of bread, how it is made, its price, its role in the history of mankind, etc.).

The lexical units of a lexicon are in many ways interrelated. They may share some phonological properties (for example, they may rhyme with each other), they may belong to the same inflectional paradigm, they may have the opposite meaning ('antonyms,' such as *black* and *white*), approximately the same meaning ('synonyms,' such as *to begin* and *to start*), or when complex in form they may follow the same construction pattern. Lexicological research is often oriented towards these interrelations, whereas lexicography tends to give more weight to the lexical unit in itself. In general, there is much more lexicographical than lexicological work (for a survey of the latter, see

Schwarze and Wunderlich 1985); in fact, if there is any piece of linguistic description for some language, it is probably an elementary bilingual dictionary. The depth of this work varies massively not only across languages, but also with respect to the particular lexical properties. Whereas the phonological, graphematic and morphosyntactic features of the lexicon in Latin, English, French, and some dozen other languages with a comparable research tradition are fairly well described, there is no theoretically and empirically satisfactory analysis of the semantics of the lexicon for any language whatsoever. This has three interrelated reasons. First, there is no well-defined descriptive language which would allow the researcher to represent the meaning of some lexical unit, be it simple or compound; the most common practice is still to paraphrase it by an expression of the same language. Second, there is no reliable and easily applicable method of determining the lexical meaning of some unit; the most common way is to look at a number of occurrences in ongoing text and to try to understand what it means. Third, the relation between a particular form and a particular meaning is hardly ever straightforward; this is strikingly illustrated by a look at what even a medium-sized English dictionary has to say about the meaning of, for example, *on*, *sound*, *eye* or *(to) put up*. As a rule, there is not just one lexical meaning, but a whole array of uses which are more or less related to each other. This is not merely a practical problem for the lexicographer; it also casts some doubt on the very notion of 'lexical unit' itself (see also *Lexical Semantics*).

3. Making Dictionaries

Lexicographers often consider their work to be more of an art or a craft than a science (see, e.g., Landau 1984, Svensén 1993). This does not preclude a solid scientific basis, but it reflects the fact that their concrete work depends largely on practical skills such as being 'a good definer,' on one hand; and that it is to a great extent determined by practical, often commercial, concerns, on the other. Dictionaries are made for users, and they are intended to serve specific purposes. Their compilation requires a number of practical decisions.

3.1 Which Lexical Units are Included?

Languages are neither well-defined nor uniform entities; they change over time, and they vary with factors such as place, social class, or area talked about. A great deal of this variation is lexical. It is not possible nor would it be reasonable to cover this wealth in a single dictionary. Large dictionaries contain up to 300,000 'entries'; since idiomatic expressions are usually listed under one of their components (such as

to kick the bucket under (to) kick), they contain many more lexical units, perhaps up to 1 million. But, even so, they are by no means exhaustive. The second edition of the *Deutsches Wörterbuch* (see Sect. 5), the largest dictionary of German, covers less than 25 per cent of the lexical units found in the sources, and these sources are quite restricted themselves.

3.2 Which Lexical Properties are Described?

Just as it is impossible to include all lexical units of a language in a dictionary, it is neither possible nor desirable to aim at a full description of those which are included. Since a dictionary is normally a printed book, the graphematic properties of the unit (its 'spelling') are automatically given. Among the other defining properties, meaning is traditionally considered to be most important. Samuel Johnson's dictionary from 1755 (see Sect. 5) defines 'dictionary' as 'A book containing the words of any language in alphabetical order, with explanations of their meaning.' But Johnson also noted which syllable carries the main stress, and he gave some grammatical hints. In general, however, information on phonological properties was rare up to the end of the nineteenth century, and information on grammatical properties is usually still very restricted in nonspecialized dictionaries. But there are, of course, also dictionaries which specifically address these properties as well as some of the nondefining properties associated with a lexical entry, such as its origin (etymological dictionary) or, above all, its equivalent in other languages ('bilingual dictionary').

3.3 What is the Description Based Upon?

Usually, two types of sources are distinguished: 'primary sources' are samples of text in which the unit is used, 'secondary sources' refers to prior work of other lexicographers (and lexicologists). In fact, there is a third source, normally not mentioned in the theory of lexicography (sometimes called 'meta-lexicography'): this is the lexicographer's own knowledge of the language to be described, including his or her views on what is 'good' language. In practice, the bulk of a new dictionary is based on older dictionaries. This is always immoral and often illegal, if these are simply copied; but on the other hand, it would be stupid and arrogant to ignore the achievements of earlier lexicographers.

3.4 How is the Information Presented?

A dictionary consists of lexical entries arranged in some conventional order. Normally, an entry combines several lexical units under a single 'head word';

thus, all lexical units which include the word *put* may be listed under this head word, forming a kind of nest with an often very complex microstructure. We are used to alphabetically-ordered dictionaries; but there are other possibilities, for example, by thematic groups or by first appearance in written documents. Languages without alphabetic writing require different principles; in Chinese, for example, entries are usually arranged by subcomponents of the entire character and by the number of strokes.

These four questions can be answered in very different ways, resulting in very different types of dictionaries (see the survey in Hausmann et al. 1991, pp. 968–1573).

4. History

The first lexicographic documents are lists of Sumerian words (up to 1400) with their Akkadian equivalents, written in cuneiform script on clay tablets about 4,700 years ago. The practice compiling such word lists was continued throughout Antiquity and the Middle Ages; thus, the oldest document in German, the *Abrogans* (written around 765), is an inventory of some Latin words with explanations in German. Usually, these 'glossaries' did not aim at a full account of the lexicon; they simply brought together a number of words which, for one reason or another, were felt to be 'difficult,' and explained them either by a more familiar word in the same language or by a translation. Words were ordered alphabetically, by theme, or not at all. But there are also more systematic attempts, such as the *Catholicon*, a mixture of encyclopedia and dictionary which, compiled around 1250, was the first printed lexical work in Europe (Mainz 1460).

In the sixteenth century, two developments led to major changes. The first of these was the invention of printing by Gutenberg. By 1500, virtually all classical authors were available in print, thus offering a solid basis for systematic lexical accounts of Latin and Greek, such as Calepinus' *Dictionarium* (1502), soon to be followed by two early masterpieces: Robert Etienne's *Dictionarium seu Latinae Linguae Thesaurus* (Paris 1531) and Henri Etienne's *Thesaurus Graecae Linguae* (Paris 1572). The second major development was the slow but steady rise of national languages. Since early Italian, French, English, or German were hardly codified, a major aim of the first dictionaries in these languages was to give them clear norms. In some countries, national Academies were founded to this end. The outcome were dictionaries with a strongly normative, often puristic, stance, such as the *Vocabulario degli Accademici della Crusca* (Venice 1612), the *Dictionnaire de l'Académie Française* (Paris 1694) and the *Diccionario de autoridades publicado por la Real Academia Espanola* (1726–1739).

The bulk of lexicographic work, however, was always done by enterprising publishers and engaged

individuals, such as Dr Samuel Johnson. Helped by six assistants, he produced *A Dictionary of the English Language* (London 1755), the first scholarly description of the English vocabulary, in less than eight years. It surpassed all its predecessors, including Bailey's *Dictionarium Britannicum* from 1736, which Johnson took as his point of departure, by the systematic use of quotations, taken from the best writers, and by his brilliant, sometimes somewhat extravagant, definitions (not everybody would dare to characterize patriotism as 'the last refuge of a scoundrel'). Less known, much less witty, but broader in coverage is the first comprehensive dictionary of German, Johann Christoph Adelung's *Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der hochdeutschen Mundart* (Leipzig 1774–86).

The rise of historical-comparative linguistics in the early nineteenth century led to an enormous increase in grammatical and lexical knowledge. The first dictionary which tried to cover this knowledge was the *Deutsches Wörterbuch* by Jacob Grimm and (to a much lesser extent) his brother Wilhelm Grimm. Its first fascicle appeared in 1852, after about ten years of preparatory work, in which the Grimms were helped by about 100 scholars providing excerpts ('covering my desk like snowflakes,' Jacob Grimm). At that time, it was already clear that the original plan of 6–7 volumes, to be finished within 10–12 years, was unrealistic. The Grimms finished only letters A–(most of) F, and the final folio volume (of altogether 32) appeared in 1960. This long duration, as well as the varying talents and preferences of the contributors, has led to many inconsistencies; some entries got out of balance (no less than 60 pages are devoted to the single word *Geist*); still, it is an incommensurable source of lexical information.

The work of the Grimms inspired a number of similar ventures, such as Emile Littré's masterly *Dictionnaire de la langue française* (1863–1873), which is much shorter, but also much more consistent; Matthias de Vries and his numerous successors' voluminous *Woordenboek der Nederlandsche Taal* (1864–1998), and finally *A New English Dictionary on a Historical Basis* (1884–1928), generally referred to as the 'Oxford English Dictionary' (OED). It was initiated in 1857 by the philologist and churchman Richard Trench; in 1860, members of the Philological Society started to collect excerpts; in 1879, the Clarendon Press appointed James Murray as the Principal Editor. The first fascicle appeared in 1882, and the whole work was completed in 1928, 13 years after Murray's death. More than 200 scholars were involved in its production, more than 2,000 people are known to have contributed excerpts. The OED is not without flaws, even in its revised edition, which appeared in 1989 in print and in 1992 on CD-ROM; but among all attempts to describe the lexicon of a language, it comes closest to falsify what Dr Johnson stated in the preface to his own dictionary: 'Every

other author may aspire to praise; the lexicographer can only hope to escape reproach.' (For a comprehensive survey of lexicographic work across languages, see Hausmann et al. 1991, pp. 1679–2710, 2949–3119).

5. *The Use of Computers*

We tend to think of dictionaries as the normal, if not the only possible way to compile and to present lexical information. But the dawn of the computer has provided us with a very different and in many ways more efficient tool. Computers can be used in at least three ways in lexicography. It is possible to transfer an existing dictionary to a computer, as has often been done over the last 20 years. Such a transfer offers several advantages: search is faster and more exhaustive; it is easier to revise and update the dictionary; and it is possible to add information not available in book format, for example, spoken sound instead of phonetic transcriptions. But essentially, the format of the printed book is maintained. Next, computers are a powerful tool in the production of a new dictionary. Rather than having a number of people read through books and newspapers and make excerpts of all occurrences which look interesting, it is now possible to compile huge text corpora that cover all varieties of a language, to scan these texts for all occurrences of words or word combinations, to sort these occurrences by various criteria, to link them to other occurrences, to add as much context as needed, etc. (see *Corpus Linguistics*).

The OED, is based on about 5 million excerpts, mostly handwritten on paper slips. A computer can easily process corpora of several hundred million words, i.e., several hundred million occurrences; new sources can rapidly be added. This allows a much broader and much more representative coverage of a lexicon than ever. But electronic corpora only provide the raw material; it still awaits lexical analysis. This analysis can be facilitated by computer tools, also; but no computer can tell us what a word means in a particular context. But even if only one minute is devoted to each occurrence in a one-hundred million corpus, it would take 10 lexicographers 100 years to go through it. This means that printed dictionaries can never reflect the wealth of information accessible in large corpora, since they presuppose that the lexicographer has finished the analysis. Therefore, the only way to make full use of large corpora is by means of lexical retrieval systems. They consist of (a) a computer-accessible and expandable corpus, (b) a set of tools, which allow, for example, not only the search for certain items but also statistical analysis or the determination of the first occurrence, and (c) a selective but steadily proceeding lexical analysis of the corpus. Thus, it is possible to add spoken forms in various dialects, information about word classes, or

the semantic analysis of some subset of lexical units, say all prepositions or all morphologically simple verbs. Similarly, translation equivalents can be added. Unlike printed dictionaries, such a lexical retrieval system will never come to an end, it is steady work-in-progress to which many can contribute and which will give us a deeper and broader understanding of the lexicon than any other method.

See also: Corpus Linguistics; Lexical Access, Cognitive Psychology of; Lexical Processes (Word Knowledge); Psychological and Neural Aspects; Lexical Semantics

Bibliography

- Hausmann F-J, Reichmann O, Wiegand H E, Zgusta L (eds.) 1991 *Wörterbücher—Dictionaries—Dictionnaires*. De Gruyter, Berlin
- Landau SA 1984 *Dictionaries: The Art and Craft of Lexicography*. Scribner, New York
- McArthur T 1986 *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer*. Cambridge University Press, Cambridge, UK
- Schwarze C, Wunderlich D (eds.) 1985 *Handbuch der Lexikologie*. Athenäum, Königstein, Germany
- Svensén B 1993 *Practical Lexicography. Principles and Methods of Dictionary Making*. Oxford University Press, Oxford, UK

W. Klein

Lexicon

1. The Notion of Lexicon

The lexicon is standardly viewed as a listing of all the morphemes of a language, with information indicating how each morpheme behaves in the components of grammar involving phonology, syntax, and semantics. In no small part, the shape and character of grammar is determined by what the lexicon contains for these other grammatical devices. Nevertheless, both historically and conventionally, the lexicon has been seen as the passive module in the system of grammar.

More recently, the model of the lexicon has undergone significant revision and maturation. In particular, two trends have driven the architectural concerns of lexical researchers: (a) a tighter integration of compositional operations of syntax and semantics with the lexical information structures that bear them; and (b) a serious concern with how lexical types reflect the underlying ontological commitments of the grammar. In the process, the field has moved towards addressing more encompassing problems in linguistic theory, such as those below:

(a) How can we explain the polymorphic nature of language?

(b) How can we capture the creative use of words in novel contexts?

(c) How can semantic types predictably map to syntactic representations?

(d) What are the 'atoms' of lexical knowledge, if they exist at all?

In this article, we first review the conventional view of the lexicon and then contrast this with the theories of lexical information that have emerged since around 1990.

By all accounts, the conventional model of the lexicon is that of a database of words, ready to act in the service of more dynamic components of the grammar. This view has its origins squarely in the generative tradition (Chomsky 1955) and has been an increasingly integral part of the concept of the lexicon ever since. While the 'Aspects' model of selectional features restricted the relation of selection to that between lexical items, work by Jackendoff (1972) and McCawley (1968), showed that selectional restrictions must be available to computations at the level of derived semantic representation rather than at deep structure. But where did this view come from? In order to understand both the classical model of the lexicon as a database and the current models of lexically encoded grammatical information, it is necessary to appreciate the structuralist distinction between 'syntagmatic processes' and 'paradigmatic systems' in language. The lexicon has emerged as the focal point communicating between these two components, and can be seen as a hook which links the information at these two levels. One can go further still and view the elements of the lexicon as not just the building blocks for the more active components of the grammar, but also as actively engaging the building principles themselves.

While syntagmatic processes refer to the influence of horizontal elements on a word or phrase, paradigmatic systems refer to vertical substitutions in a phrasal structure. Syntagmatics evolved into the theory of abstract syntax while paradigmatics was all but abandoned in generative linguistics. In an early discussion of syntagmatic dependencies, Hjelmslev (1943) uses the term 'selection' explicitly in the modern sense and notes the importance of integrating paradigmatic systems with the syntagmatic processes they participate in. For Hjelmslev, there are two possible types of relations that can exist between elements in a syntagmatic process: 'interdependence' and 'determination', the latter of which is related to the notion of selectional restriction as developed by Chomsky (1965). As Cruse (1986) notes 'One reason that selectional restrictions were not integrated into mechanisms of grammatical selection and description in the 1970s and 1980s is that, if they are imposed correctly, the grammar is forced to model two computations:

(a) the entailment relations between selectional