Word

Although many language users intuitively know what a 'word' is, an unequivocal and clear-cut linguistic definition cannot be given. The concept of 'word' is highly ambiguous and refers to different phenomena. On the one hand, words are language units at the intersection of two linguistic levels of description: morphology (which is the arrangement of meaningful sound sequences, or morphemes, in words) and syntax (which is the arrangement of words in sentences). Accordingly, lexicology is the linguistic branch that deals with words, which thus implies both morphological and syntactic aspects. On the other hand, words are regarded as basic units in the psychological reality of language acquisition, production, and processing: words in the mind. An attempt to produce a system reflecting the versatility of the notion of 'word' would need to describe words as both 'linguistic structures' (although of different kinds) and as parts of the mental lexicon.

The linguistic notion of 'word' captures three fundamentally different, but closely related, concepts that will be indicated in the following as $word_1$, $word_2$, and $word_3$. To begin with, so-called 'word forms' (i.e. $word_1$) refer to the physical substance of words in spoken and written language: phonological word forms in speech, and orthographic word forms in writing. For example, /bɔɪ/ and /bɔɪz/ are two phonological word forms, i.e. specific combinations of sounds in the phonic medium. On the other hand, *boy*, *boy's*, *boys*, and *boys'* represent four corresponding orthographic word forms, or combinations of letters (and apostrophes) in the written medium. The six word forms provide the inventory of possible word-form realizations of one underlying abstract unit, the lexeme *BOY*.

The lexeme ($word_2$), usually given in capitalized letters, is what all the actual word forms have in common at an abstract level. It is this abstract unit of a lexeme that is at the basis of a dictionary since all word forms of one and the same lexeme are subsumed under the same dictionary entry.

While word forms represent merely physical forms in terms of sounds or letters, lexemes—in a first approximation—are examples of linguistic signs. That is to say, a lexeme connects potential forms (i.e. the signifier) to an abstract meaning encoded by these forms (i.e. the signified), such as the phonological word forms /τεɪβλ/ and /τεɪβλζ/ (i.e. the signifier) to the meaning of 'piece of furniture consisting of a flat top supported on one or more legs' (i.e. the signified). This arbitrary, but fixed, form-meaning relation is inherent to the lexeme as a linguistic sign. However, a lexeme actually forms a set of several linguistic signs: some word forms of a lexeme may be linked to specific (sub-) meanings of the lexeme and vice versa. For example, all orthographic word forms of the verb YIELD, i.e. *yield*, *yields*, *yielded*, *yielding*, are linked to the possible meanings 'give way' and 'lead to'. On the other hand, a third meaning of the verb 'produce' is strongly associated with the first two aforementioned word forms. To be able to refer to such subsets of word forms with a specific range of meaning, the concept of 'lexical unit' or 'lexical linguistic sign' has been introduced.

Apart from word forms and lexemes (including lexical units), the concept of 'word' can also be

approached from a grammatical perspective. A grammatical word (i.e. word$_3$) is immediately relevant to syntax and has specific morphological features. The distinction between word forms and grammatical words is important because one word form may represent different grammatical words (a phenomenon called 'syncretism'). For example, the phonological word form /β I ζ/ can be regarded as a realization of the plural form, the singular genitive, and the plural genitive. In applying the concept of grammatical words, linguists are particularly interested in the question of how words, as minimal syntactic units, are arranged in grammatical structures. Irrespective of the specific grammatical principles at work, grammatical words have a certain ability to change position in a sentence (i.e. 'positional mobility'). Thus, it is often possible to shift a grammatical word, in its entirety, to a different position in a given sentence without jeopardizing the grammatical integrity of that sentence. It goes without saying that in English, which has a comparatively fixed word order, the overall degree of positional mobility is lower than, for example, in German, which allows for much more word-order variation. In terms of their internal structure, however, grammatical words are stable, in that the sequence of morphemes is, in principle, not subject to variation: for instance, *boy-ish* is a possible adjective in English, but the reversed sequence, i.e. **ish-boy*, is not permitted.

Words also play an important role in the mental processes that underlie language use. It is reasonable to assume that words are stored in the mind in a highly structured way. This orderliness of the complex mental lexicon permits the amazing speed of lexical retrieval in natural speech. It is widely accepted that the mental lexicon presumably does not list each and every word in a random order. Rather, a list of lexical structures (words, word groups, word components) is supposed to be complemented with a set of rules and principles on how to combine these structures. This concept avoids, for instance, a listing of different words ending in the same suffix as separate entries: for example, *boyish* and *girlish* are derived from applying the same combinatorial rule to the base entries *boy* and *girl*, respectively. Such systematic combinations presuppose a mental capacity for analyzing complex words in speech processing, which is subsumed under the notion of 'morphological parsing'. In the aforementioned example, morphological parsing would refer to speakers' ability to strip the suffix *-ish* and to identify a general rule that makes it possible to derive an adjective from a noun by adding this suffix to the noun.

The importance of words for the linguistic competence in the mind and the highly structured design of the mental lexicon are also corroborated by research into language acquisition. The surprisingly rapid development of child language is to a large extent based on the acquisition of thousands of words in a comparatively short time. At a more or less specific stage of this development, children tend to overgeneralize morphological rules. For instance, they add the regular past tense morpheme (*-ed* in English) to all regular verbs (*learn* becomes *learned*) and irregular verbs (*go* becomes **goed*). This phenomenon reflects the tendency for a child's mental lexicon to comprise both basic elements (*learn*, *go*, *-ed*) and so-called rules of morphological parsing to combine these elements and analyze the combinations (verb_present $+ -ed =$ verb_past).

Lexical disorders, too, shed light on the structure of the mental lexicon. Aphasia—a condition of speechlessness that stroke patients often display when specific areas of the brain are affected—is a compelling example of such a disorder. Broca's aphasia, for example, leads to 'telegraphic speech', or the omission of function words (e.g. articles like *the*, prepositions like *of*), whereas patients suffering from Wernicke's aphasia show difficulty in lexical selection (i.e. they very often do not produce the words they want but other words with different meanings, which often renders their speech unintelligible).

In the light of their structural and mental significance, it is reasonable to assume that words are the most basic linguistic units. However, the overall importance of words for language to function is largely at odds with the yet unresolved problem of word identification. What is relatively simple in written language, where words are usually separated by spaces from each other (though not the case in all languages), is far more complicated in spoken language. Several attempts have been made to establish criteria for a reliable identification of words in the continuous stream of speech sounds. For example, words are said to be demarcated by positions at which pausing is, at least, possible. Another suggestion is to look for those stretches of sounds that could stand meaningfully on their own. Although these two approaches and others may be useful for the majority of words, other cases cannot be clarified, such as compounds, which might be considered as one word or as several words.

In all languages, linguists set out to group words into word classes. Generally, one can distinguish between 'open' and 'closed' word classes. Open word classes can always have new words freely added to them; in most languages, these classes include nouns, verbs, adjectives, and adverbs. In contrast, usually no new members can be added to closed word classes. In English, for example, prepositions (e.g. *under*), conjunctions (e.g. *because*), pronouns (e.g. *mine*), determiners (e.g. *the*), and numerals (e.g. *first*) are the

primary closed word classes. The basis for the categorization of words into classes is provided by their morphosyntactic behavior. Nouns, for instance, are almost always able to take a plural form (disregarding exceptions). As with word identification, word class categorization is not straightforward in all cases. An important example in English is, perhaps, the word class of adverbs. This class contains members that are extremely dissimilar in structural terms, such as the prototypical adverb *rapidly* and the negative particle *not*.

Words have an internal structure themselves and are at the same time the basic units of larger structures. The former aspect refers to the domain of 'derivational morphology', i.e. the linguistic description of processes and phenomena of word formation. The latter aspect captures the fact that words play a key role in syntactic structures, which they enter in a principle-guided fashion. Moreover, a given word in a specific position tends to predict, to a large extent, the range of words possible both before and after it. If the probability of two words occurring next to each other is significantly high, this phenomenon is called 'collocation' (from *co-location*).

With collocations, corpus linguistics has definitely broken new ground. In searching very large computer collections of text (so-called corpora) for co-occurrences, it is now possible to base the concept of collocation on an empirical foundation. By applying such corpus linguistic methods, it can be statistically determined to what extent a given word predicts another word in its immediate neighborhood. For example, *strong* and *argument* co-occur in significant frequency in authentic English language use. Their combination is thus not of a random nature but constitutes a collocation, i.e. a lexical co-selection. Furthermore, specific words tend to occur in a restricted range of grammatical patterns (a phenomenon called 'colligation', from *co-ligation*). For example, *naked eye* almost exclusively occurs in the grammatical pattern 'verb/adjective + preposition + definite article + *naked eye*' as in *visible to the naked eye*. Collocations and colligations reveal that lexical choices and grammatical choices are not independent of each other but that lexis and grammar are intricately intertwined. One could argue that the traditionally established distinction between lexis and grammar should eventually be overcome altogether and be replaced with an integrated lexicogrammar (or a 'pattern grammar').

The linguistic level of words, i.e. lexicology, is of particular interest for dictionary makers or lexicographers. They take an inventory of the words of a language but, by definition, are bound to fail to give an exhaustive account of all words in existence at a given time. Although it is impossible to list all words in a given language, the use of dictionaries has led to a high degree of standardization in orthographic and phonological word forms and are, of course, a helpful tool for language users in general and language learners in particular. Dictionaries should include orthographic, phonological, grammatical, and semantic information about words, thereby strongly resembling the composition of the mental lexicon as hypothesized by psycholinguists. One ought to bear in mind, however, that the mental lexicon is based on a list of entries and rules to combine them, whereas dictionaries usually contain only an alphabetical list of entries. Furthermore, large dictionaries represent enormous and idealized word lists of which no single native speaker has command: they approximate the so-called 'magnavocabulary' of a language in its entirety, but do not correspond to the specific vocabulary that an individual speaker has at his or her disposal.

The object of inquiry for lexicographers and lexicologists is in constant flux. Productivity and imagination in human languages are perhaps most obvious at the level of words.

References

Aitchison, Jean. 1994. Words in the mind: an introduction to the mental lexicon. Oxford: Blackwell.

Bauer, Laurie. 1983. English word-formation. Cambridge: Cambridge University Press.

Coates, Richard. 1999. Word structure. London: Routledge.

Cruse, D. Alan. 1986. Lexical semantics. Cambridge: Cambridge University Press.

Esser, Jürgen. 2000. Corpus linguistics and the linguistic sign. Corpus linguistics and linguistic theory: Proceedings of the 20th ICAME conference, ed. by Christian Mair and Marianne Hundt. Amsterdam: Rodopi.

Fischer, Roswitha. 1998. Lexical change in present-day English: a corpus-based study of the motivation, institutionalization, and productivity of creative neologisms. Tübingen: Narr.

Hunston, Susan, and Gill Francis. 2000. Pattern grammar: a corpus-driven approach to the lexical grammar of English. Amsterdam: John Benjamins.

Katamba, Francis. 1994. English words. London: Routledge.

Lipka, Leonhard. 1990. An outline of English lexicology: lexical structure, word semantics, and word-formation. Tübingen: Niemeyer.; 2nd edition, 1992.

Marchand, Hans. 1969. The categories and types of present-day English word-formation: a synchronic–diachronic approach. München: Beck.

Stubbs, Michael. 2001. Words and phrases: corpus studies of lexical semantics. Oxford: Blackwell.

Wierzbicka, Anna. 1985. Lexicography and conceptual analysis. Ann Arbor: Karoma.

JOYBRATO MUKHERJEE

See also **Aphasia; Corpus Linguistics**

Manuscript Queries

Title: Encyclopedia of Linguistics
Alphbet-W Chapter-9: WORD

Page	Query Number	Query
		NO QUERIES

Note: If your article contains special/phonetic characters, please ensure that they have been rendered in the correct form as some of these were missing in the electronic files submitted for production.