**Phonetics**

From Wikipedia, the free encyclopedia

**Phonetics** (from the [Greek](http://en.wikipedia.org/wiki/Greek_language): [φωνή](http://en.wiktionary.org/wiki/%CF%86%CF%89%CE%BD%CE%AE), *phōnē*, "sound, voice") is a branch of [linguistics](http://en.wikipedia.org/wiki/Linguistics) that comprises the study of the [sounds](http://en.wikipedia.org/wiki/Sound) of human [speech](http://en.wikipedia.org/wiki/Speech), or—in the case of [sign languages](http://en.wikipedia.org/wiki/Sign_languages)—the equivalent aspects of sign.[[1]](http://en.wikipedia.org/wiki/Phonetics#cite_note-0) It is concerned with the physical properties of speech sounds or signs ([phones](http://en.wikipedia.org/wiki/Phone_%28phonetics%29)): their physiological production, acoustic properties, auditory perception, and neurophysiological status. [Phonology](http://en.wikipedia.org/wiki/Phonology), on the other hand, is concerned with the abstract, grammatical characterization of systems of sounds or signs.

The field of phonetics is a multiple layered subject of [linguistics](http://en.wikipedia.org/wiki/Linguistics) that focuses on speech. In the case of oral languages there are three basic areas of study:

* *Articulatory phonetics*: the study of the production of speech sounds by the articulatory and vocal tract by the speaker
* *Acoustic phonetics*: the study of the physical transmission of speech sounds from the speaker to the listener
* *Auditory phonetics*: the study of the reception and perception of speech sounds by the listener

These areas are inter-connected through the common mechanism of sound, such as wavelength ([pitch](http://en.wikipedia.org/wiki/Pitch_%28music%29)), amplitude, and harmonics.

|  |
| --- |
| **Contents**   * [1 History](http://en.wikipedia.org/wiki/Phonetics#History) * [2 Phonetic transcription](http://en.wikipedia.org/wiki/Phonetics#Phonetic_transcription) * [3 The difference between phonetics and phonology](http://en.wikipedia.org/wiki/Phonetics#The_difference_between_phonetics_and_phonology) * [4 Relation to phonology](http://en.wikipedia.org/wiki/Phonetics#Relation_to_phonology) * [5 Subfields](http://en.wikipedia.org/wiki/Phonetics#Subfields) * [6 Transcription](http://en.wikipedia.org/wiki/Phonetics#Transcription) * [7 Applications](http://en.wikipedia.org/wiki/Phonetics#Applications) * [8 See also](http://en.wikipedia.org/wiki/Phonetics#See_also) * [9 Notes](http://en.wikipedia.org/wiki/Phonetics#Notes) * [10 References](http://en.wikipedia.org/wiki/Phonetics#References) * [11 External links](http://en.wikipedia.org/wiki/Phonetics#External_links) |

**History**

Phonetics was studied as early as 500 BC in [ancient India](http://en.wikipedia.org/wiki/Indian_subcontinent), with [Pāṇini](http://en.wikipedia.org/wiki/P%C4%81%E1%B9%87ini)'s account of the [place](http://en.wikipedia.org/wiki/Place_of_articulation) and [manner of articulation](http://en.wikipedia.org/wiki/Manner_of_articulation) of consonants in his 5th century BC treatise on [Sanskrit](http://en.wikipedia.org/wiki/Sanskrit). The major [Indic alphabets](http://en.wikipedia.org/wiki/Brahmic_family) today order their consonants according to Pāṇini's classification. The [Ancient Greeks](http://en.wikipedia.org/wiki/Ancient_Greeks) are credited as the first to base a writing system on a phonetic alphabet. Modern phonetics began with [Alexander Melville Bell](http://en.wikipedia.org/wiki/Alexander_Melville_Bell), whose [*Visible Speech*](http://en.wikipedia.org/wiki/Visible_Speech) (1867) introduced a system of precise notation for writing down speech sounds.[[2]](http://en.wikipedia.org/wiki/Phonetics#cite_note-1)

**Phonetic transcription**

The [International Phonetic Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet)(IPA) is used as the basis for the phonetic transcription of speech. It is based on the Latin alphabet and is able to transcribe most features of speech such as consonants, vowels, and suprasegmental features. Every documented phoneme available within the known languages in the world is assigned its own corresponding symbol.

**The difference between phonetics and phonology**

Phonology concerns itself with systems of [phonemes](http://en.wikipedia.org/wiki/Phoneme), abstract *cognitive* units of speech sound or sign which distinguish the words of a language. Phonetics, on the other hand, concerns itself with the production, transmission, and perception of the *physical* phenomena which are abstracted in the mind to constitute these speech sounds or signs.

Using an Edison phonograph, [Ludimar Hermann](http://en.wikipedia.org/wiki/Ludimar_Hermann) investigated the spectral properties of vowels and consonants. It was in these papers that the term [*formant*](http://en.wikipedia.org/wiki/Formant) was first introduced. Hermann also played back vowel recordings made with the Edison phonograph at different speeds in order to test [Willis'](http://en.wikipedia.org/wiki/Wilfrid_Willis) and [Wheatstone's](http://en.wikipedia.org/wiki/Sir_Charles_Wheatstone) theories of vowel production.

**Relation to phonology**

In contrast to phonetics, [phonology](http://en.wikipedia.org/wiki/Phonology) is the study of how sounds and gestures pattern in and across languages, relating such concerns with other levels and aspects of language. Phonetics deals with the articulatory and acoustic properties of speech sounds, how they are produced, and how they are perceived. As part of this investigation, phoneticians may concern themselves with the physical properties of meaningful sound contrasts or the social meaning encoded in the speech signal (e.g. [gender](http://en.wikipedia.org/wiki/Gender), [sexuality](http://en.wikipedia.org/wiki/Human_sexuality), [ethnicity](http://en.wikipedia.org/wiki/Ethnicity), etc.). However, a substantial portion of research in phonetics is not concerned with the meaningful elements in the speech signal.

While it is widely agreed that phonology is grounded in phonetics, phonology is a distinct branch of linguistics, concerned with sounds and gestures as abstract units (e.g., [features](http://en.wikipedia.org/w/index.php?title=Features_%28linguistics%29&action=edit&redlink=1), [phonemes](http://en.wikipedia.org/wiki/Phoneme), [mora](http://en.wikipedia.org/wiki/Mora_%28linguistics%29), [syllables](http://en.wikipedia.org/wiki/Syllables), etc.) and their conditioned variation (via, e.g., [allophonic rules](http://en.wikipedia.org/wiki/Allophonic_rule), constraints, or [derivational rules](http://en.wikipedia.org/wiki/Phonological_rule)).[[3]](http://en.wikipedia.org/wiki/Phonetics#cite_note-2) Phonology relates to phonetics via the set of [distinctive features](http://en.wikipedia.org/wiki/Distinctive_feature), which map the abstract representations of speech units to articulatory gestures, acoustic signals, and/or perceptual representations.[[4]](http://en.wikipedia.org/wiki/Phonetics#cite_note-3)[[5]](http://en.wikipedia.org/wiki/Phonetics#cite_note-4)[[6]](http://en.wikipedia.org/wiki/Phonetics#cite_note-5)

**Subfields**

Phonetics as a research discipline has three main branches:

* [articulatory phonetics](http://en.wikipedia.org/wiki/Articulatory_phonetics) is concerned with the articulation of speech: The position, shape, and movement of *articulators* or [speech organs](http://en.wikipedia.org/wiki/Speech_organ), such as the lips, tongue, and [vocal folds](http://en.wikipedia.org/wiki/Vocal_folds).
* [acoustic phonetics](http://en.wikipedia.org/wiki/Acoustic_phonetics) is concerned with [acoustics](http://en.wikipedia.org/wiki/Acoustics) of speech: The spectro-temporal properties of the [sound waves](http://en.wikipedia.org/wiki/Sound_wave) produced by speech, such as their [frequency](http://en.wikipedia.org/wiki/Frequency), [amplitude](http://en.wikipedia.org/wiki/Amplitude), and [harmonic structure](http://en.wikipedia.org/wiki/Harmonic_structure).
* [auditory phonetics](http://en.wikipedia.org/wiki/Auditory_phonetics) is concerned with [speech perception](http://en.wikipedia.org/wiki/Speech_perception): the [perception](http://en.wikipedia.org/wiki/Perception), [categorization](http://en.wikipedia.org/wiki/Categorization), and [recognition](http://en.wikipedia.org/wiki/Recall_%28memory%29) of speech sounds and the role of the [auditory system](http://en.wikipedia.org/wiki/Auditory_system) and the [brain](http://en.wikipedia.org/wiki/Human_brain) in the same.

**Transcription**

Main article: [Phonetic transcription](http://en.wikipedia.org/wiki/Phonetic_transcription)

[Phonetic transcription](http://en.wikipedia.org/wiki/Phonetic_transcription) is a system for transcribing sounds that occur in [spoken language](http://en.wikipedia.org/wiki/Spoken_language) or [sign language](http://en.wikipedia.org/wiki/Sign_language). The most widely known system of phonetic transcription, the [International Phonetic Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet) (IPA), uses a one-to-one mapping between phones and written symbols.[[7]](http://en.wikipedia.org/wiki/Phonetics#cite_note-O.27Grady-p17-6)[[8]](http://en.wikipedia.org/wiki/Phonetics#cite_note-7) The standardized nature of the IPA enables its users to transcribe accurately and consistently the phones of different languages, [dialects](http://en.wikipedia.org/wiki/Dialect), and [idiolects](http://en.wikipedia.org/wiki/Idiolect).[[7]](http://en.wikipedia.org/wiki/Phonetics#cite_note-O.27Grady-p17-6)[[9]](http://en.wikipedia.org/wiki/Phonetics#cite_note-Ladefoged.2C_Peter_1975-8)[[10]](http://en.wikipedia.org/wiki/Phonetics#cite_note-9) The IPA is a useful tool not only for the study of phonetics, but also for language teaching, professional acting, and [speech pathology](http://en.wikipedia.org/wiki/Speech_pathology).[[9]](http://en.wikipedia.org/wiki/Phonetics#cite_note-Ladefoged.2C_Peter_1975-8).....

**Applications**

Application of phonetics include:

* [forensic phonetics](http://en.wikipedia.org/wiki/Forensic_linguistics): the use of phonetics (the science of speech) for forensic (legal) purposes.
* [Speech Recognition](http://en.wikipedia.org/wiki/Speech_Recognition): the analysis and transcription of recorded speech by a computer system.

**See also**

* [Experimental phonetics](http://en.wikipedia.org/wiki/Experimental_phonetics)
* [Index of phonetics articles](http://en.wikipedia.org/wiki/Index_of_phonetics_articles)
* [International Phonetic Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet)
* [Speech processing](http://en.wikipedia.org/wiki/Speech_processing)
* [Acoustics](http://en.wikipedia.org/wiki/Acoustics)
* [Biometric word list](http://en.wikipedia.org/wiki/Biometric_word_list)
* [Phonetics departments at universities](http://en.wikipedia.org/wiki/Phonetics_departments_at_universities)
* [X-SAMPA](http://en.wikipedia.org/wiki/X-SAMPA)
* [ICAO spelling alphabet](http://en.wikipedia.org/wiki/ICAO_spelling_alphabet)
* [Buckeye Corpus](http://en.wikipedia.org/wiki/Buckeye_Corpus)

**Notes**

* 1. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-0) O'Grady (2005) p.15
  2. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-1) [Alexander Melville Bell 1819-1905](http://www.acsu.buffalo.edu/%7Educhan/new_history/hist19c/subpages/mbell.html) . University at Buffalo, The State University of New York.
  3. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-2) Kingston, John. 2007. *The Phonetics-Phonology Interface*, in The Cambridge Handbook of Phonology (ed. Paul DeLacy), Cambridge University Press.
  4. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-3) Halle, Morris. 1983. *On Distinctive Features and their articulatory implementation*, Natural Language and Linguistic Theory, p. 91 - 105
  5. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-4) Jakobson, Roman, Gunnar Fant, and Morris Halle. 1976. Preliminaries to Speech Analysis: The Distinctive Features and their Correlates, MIT Press.
  6. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-5) Hall, T. Allen. 2001. Phonological representations and phonetic implementation of distinctive features, Mouton de Gruyter.
  7. ^ [***a***](http://en.wikipedia.org/wiki/Phonetics#cite_ref-O.27Grady-p17_6-0) [***b***](http://en.wikipedia.org/wiki/Phonetics#cite_ref-O.27Grady-p17_6-1) O'Grady (2005) p.17
  8. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-7) International Phonetic Association (1999) Handbook of the International Phonetic Association. Cambridge University Press.
  9. ^ [***a***](http://en.wikipedia.org/wiki/Phonetics#cite_ref-Ladefoged.2C_Peter_1975_8-0) [***b***](http://en.wikipedia.org/wiki/Phonetics#cite_ref-Ladefoged.2C_Peter_1975_8-1) Ladefoged, Peter (1975) A Course in Phonetics. Orlando: Harcourt Brace. 5th ed. Boston: Thomson/Wadsworth 2006.
  10. [**^**](http://en.wikipedia.org/wiki/Phonetics#cite_ref-9) Ladefoged, Peter & Ian Maddieson (1996) The Sounds of the World’s Languages. Oxford: Blackwell.

**References**

* O'Grady, William, et al. (2005). *Contemporary Linguistics: An Introduction* (5th ed.). Bedford/St. Martin's. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0312419368](http://en.wikipedia.org/wiki/Special:BookSources/0312419368).

**Wikipedia:IPA/Introduction**

From Wikipedia, the free encyclopedia

< [Wikipedia:IPA](http://en.wikipedia.org/wiki/Wikipedia:IPA)

This is an **introduction to the International Phonetic Alphabet** ([**IPA**](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet)) for English-speaking Wikipedians. Its purpose is to orient English speakers to the IPA's basic principles. Its purpose is to indicate how a word or name actually sounds. Wikipedia uses IPA because it's the global standard used by professionals.

IPA's most daunting feature is that it has discrete [letters](http://en.wikipedia.org/wiki/Letter_%28alphabet%29) for almost all of the distinctive sounds found in the world's languages. (See [IPA#Chart](http://en.wikipedia.org/wiki/IPA#Chart).) Fortunately, using the IPA for English requires learning only a small subset of these:

* **Vowels**: English orthography uses 6 vowel *letters* (*a, e, i, o, u, y*) to represent some 15 vowel *sounds*. While the English system is compact, it is also ambiguous. The IPA is unambiguous, representing each vowel sound with a unique letter or sequence. (See the [vowel audio chart](http://en.wikipedia.org/wiki/IPA_vowels_chart_with_audio)). Note that what in English are called "[long vowels](http://en.wikipedia.org/wiki/Long_vowels)", *A, E, I, O, U*, are in the IPA transcribed with two letters apiece: /eɪ/, /iː/, /aɪ/, /oʊ/, and /juː/. The English digraphs *ee, oo, au, ei, ai, ou, ie, eu,* etc. are not used.
* **Consonants**: IPA consonants are mostly intuitive to an English speaker, with the same letter used for the same sound. Thus you already know /b, d, f, ɡ, h, k, l, m, n, p, r, s, t, v, w, z/, as long as you remember that these each have a single sound. For example, /ɡ/ always represents the sound of *get,* never of *gem,* and /s/ always the sound of *so,* never of *rose*. The letter which most confuses people is /j/, which has its Central-European value, a *y* sound as in English *Hallelujah*. Two English consonant sounds, *ch* in *chair* and *j* in *jump*, are transcribed with two IPA letters apiece, /tʃ/ and /dʒ/. The English digraphs *ch, ng, qu, sh, th* are not used. See and hear also [consonants audio chart](http://en.wikipedia.org/wiki/IPA_pulmonic_consonants_chart_with_audio)

|  |  |
| --- | --- |
| |  | | --- | | **Contents**   * [1 Vowels](http://en.wikipedia.org/wiki/Wikipedia:IPA/Introduction#Vowels) * [2 Consonants](http://en.wikipedia.org/wiki/Wikipedia:IPA/Introduction#Consonants) * [3 IPA's purpose and Wikipedia's use of IPA](http://en.wikipedia.org/wiki/Wikipedia:IPA/Introduction#IPA.27s_purpose_and_Wikipedia.27s_use_of_IPA) | |

**Vowels**

Main page: [Wikipedia:IPA for English#Key](http://en.wikipedia.org/wiki/Wikipedia:IPA_for_English#Key)

Long vowels

The first principle is to not use English [alphaphonemic](http://en.wikipedia.org/wiki/Alphaphonemic) pronunciations, as if you were reading the English alphabet. In the words below, the vowel letters are pronounced as in the English alphabet, but this is not a system found anywhere else:

* *A:* *r****a****ke*, ***a****ngel*
* *E* or *EE:* *s****ee****m, r****e****po*
* *I:* *r****i****ce*
* *O:* *d****o****te*
* *U:* ***u****se*

The English digraphs *ee, oo, au, ei, ai, ou, ie, eu,* etc. are not used.

Several of these sounds are actually two vowel sounds combined, rather than pure vowel sounds as they are in Spanish or Italian: The letter *A* is pronounced /eɪ/, *E, EE* is /iː/, *I* is /aɪ/, *O* is /oʊ/, and *U* is /juː/. In the IPA, the letter /j/ is used for the English *Y* sound, thus *you* and *ewe* are transcribed /juː/. (See [below](http://en.wikipedia.org/wiki/Wikipedia:IPA/Introduction#Consonants).) While transcribing in the IPA, you can write English alphaphonemic vowels as capitals: [rAk], [sEEm], [rIs], [dOt], [Uz], etc., and then convert from the conventions above:

* *A:* /eɪ/ *r****a****ke* /reɪk/ (not /raɪk/, which would be Germanic *reich*)
* *E:* /iː/ *s****ee****m* /siːm/
* *I:* /aɪ/ *r****i****ce* /raɪs/ (not /reɪs/, which would be *race*)
* *O:* /oʊ/ *d****o****te* /doʊt/ (not /daʊt/, which would be *doubt*)
* *U:* /juː/ ***u****se* /juːz/

**Notes**: English commonly requires *ea* or *ee* to write the /iː/ sound: *read, reed.*

A w-like sound can be heard at the end of *O* in words like *echoing* (say: *echo-echo-echoing*, and it may come out like *echo-wecho-wecho-wing*) and after the *co-* in *cooperate*; that is what the /ʊ/ in the transcription /oʊ/ captures.

There are a couple other long vowels and diphthongs in English: *OO* sound in *food* (but not *good*) is written /uː/: /fuːd/. That is, it is written like the vowel of *use* without the initial *y* sound /j/. As noted above, the *OW* sound of *doubt* or *cow* is written /aʊ/. There is also the *OY* sound /ɔɪ/ of *joy,* /dʒɔɪ/.

Short vowels

English short vowels are all transcribed by a single letter in the IPA.

Because English short vowels *a e i o u* are closer to the Classical pronunciation (still found in Spanish and Italian) than the long vowels are, it is the short vowels which are transcribed with IPA letters which resemble the English letters *a e i o u.* However, they are modified to show that they aren't exactly the Classical sounds. For the *a* sound of *cat,* the [Old English](http://en.wikipedia.org/wiki/Old_English) letter [ash](http://en.wikipedia.org/wiki/Ash_%28letter%29), ‹æ›, was resurrected: /kæt/. The *e i u* sounds of *pet, pit, put* (not *putt*) were originally written as capital letters, and that is sometimes still done with manual typewriters. However, small caps looked better, so they were for a time written E I U. These took more cursive forms over time, and are today written /ɛ ɪ ʊ/: *pet* /pɛt/, *pit* /pɪt/, *put* /pʊt/. The latter, of course, is also the short *oo* sound of *good* /ɡʊd/. The *u* vowel of *putt* or *cut*, sometimes written *a* in Indian words such as [Panjab](http://en.wikipedia.org/wiki/Panjab) = [Punjab](http://en.wikipedia.org/wiki/Punjab), is a small cap *A* without the cross-bar: *cut* /kʌt/.

The *a* sound in *bra* is written with a cursive *a.* Because it's long in many dialects, it's /ɑː/ in the IPA: /brɑː/. Likewise, the *aw* sound of *law* is long in many dialects, but different than the *bra* sound. It's written with an "open" *o* (just as /ɛ/ looks like an open *e*, since a small cap *o* looks just like a regular *o*ː *law* /lɔː/. (Many of you might not make this distinction, in which case you can think of these vowel letters as being the same when reading the IPA.) For those of you who distinguish it, there is a third similar sound, the *o* of *mop.* This is written with the *bra* vowel letter rotated 180°: *mop* /mɒp/. (A rather unusual IPA letter, as that's an unusual vowel, not found in many languages.)

Finally, there's the slurred [schwa](http://en.wikipedia.org/wiki/Schwa) sound found in many unstressed syllables, as at the end of *sofa.* This is written /ə/, a symbol used in many US dictionaries. The stressed syllable is marked with a tick: *sofa* /ˈsoʊfə/. Note that the letter /ə/ is never used for a stressed vowel; for words like *cut,* we use /ʌ/: *butter* /ˈbʌtər/, *cuppa* /ˈkʌpə/.

**Consonants**

Main page: [Wikipedia:IPA for English#Key](http://en.wikipedia.org/wiki/Wikipedia:IPA_for_English#Key)

While most IPA consonants are intuitive for English speakers, there are some caveats:

* The sound of the consonant *Y* is /j/, as in *yes* /ˈjɛs/ and *yellow* /ˈjɛloʊ/.

(This is the value the letter *J* has in central European languages like German and Polish. The IPA letter /y/ is used for a non-English vowel, the French *u*, German *ü*, and Swedish *y* sound.)

* The *NG* sound of *sing* is written by combining the letter *n* with the tail of the *g,* /ŋ/, as in *sing* /ˈsɪŋ/. This is not the same as the sound in *finger,* which has an extra *g* sound: /ˈfɪŋɡər/.
* The digraph *TH* is used for two sounds in English. Since the IPA uses a single letter for each sound, two new letters are required for these two sounds:
  + /θ/ for the *th* in *thick* /ˈθɪk/ (from the Greek letter [theta](http://en.wikipedia.org/wiki/Theta))
  + /ð/ for the *th* in *those* /ˈðoʊz/ (from the Old English letter [eth](http://en.wikipedia.org/wiki/Eth), which was used for the *th* sounds)
* The sound of the digraph *SH* is transcribed with the 'stretched' *S* seen in old books. It's used in its cursive form, /ʃ/, to make it easier to read, as in *push* /ˈpʊʃ/ and *shelf* /ˈʃɛlf/.
* There is a sound with no letter or digraph in English, though sometimes written *ZH* in foreign words. It's usually written *si,* as in *vision.* In the IPA, it's written with a 'stretched' *Z*, /ʒ/: *vision* /ˈvɪʒən/.
* As noted above, the digraph *CH* is a sequence of sounds, *T* plus *SH*. This may be hard for an English speaker to hear, but is obvious to a French speaker, which is why we get spellings like [*Tchaikovsky*](http://en.wikipedia.org/wiki/Tchaikovsky) but also *catch* in English. (Adding a *t* to *ch* doesn't make any difference, because the *ch* already has a *t* sound within it.) The IPA uses the same stretched *S* for this sound here as anywhere else: *itch* /ˈɪtʃ/.
* Similarly, the English consonant *J* is a sequence with a *d* sound in it. For instance, in *judge,* adding the *d* doesn't affect the consonant any, just the vowel. In the IPA, this is transcribed /dʒ/: *jump* /ˈdʒʌmp/, *judge* /ˈdʒʌdʒ/, or *Jesus* /ˈdʒiːzəs/.
* Finally, the IPA letter [r] is officially a [trill](http://en.wikipedia.org/wiki/Trill), as in Italian and Spanish. The rather unusual English *R* sound is transcribed with a turned *r,* [ɹ]. However, since this makes no difference within English, and not all English dialects actually use the [ɹ] sound, it's very common to see English *R* transcribed with a plain /r/, and that's the convention used on Wikipedia.

The English digraphs *ch, ng, qu, sh, th* are not used.

**IPA's purpose and Wikipedia's use of IPA**

IPA's purposes are to 1) represent the [phonetics](http://en.wikipedia.org/wiki/Phonetics) of words (how they sound), and 2) to give samples of the [phonology](http://en.wikipedia.org/wiki/Phonology) of a language (how the language as a whole sounds). The second purpose concerns only linguists. The first purpose concerns any interested reader, but only to a limited degree, as transcribing words into IPA does not need to be perfect or overly precise (something for fluent IPA users to consider). The word "transcribe" is used to distinguish this from normal writing or spelling, which has other purposes (such as preserving word etymologies and meaning).

The paradox with IPA is that it's complex enough to represent nearly anything, but high-fidelity transcriptions will use glyphs that are unfamiliar to English readers and unpracticed in English [phonology](http://en.wikipedia.org/wiki/Phonology). For example a transcription of something like the Icelandic name [*Eyjafjallajökull*](http://en.wikipedia.org/wiki/Eyjafjallaj%C3%B6kull) is pronounced [[ˈɛɪjaˌfjatl̥aˌjœkʏtl̥]](http://en.wikipedia.org/wiki/Wikipedia:IPA_for_Icelandic) ([http://upload.wikimedia.org/wikipedia/commons/thumb/2/21/Speaker_Icon.svg/13px-Speaker_Icon.svg.png](http://en.wikipedia.org/wiki/File:Eyjafjallaj%C3%B6kull-bjarmason.ogg) [listen](http://upload.wikimedia.org/wikipedia/commons/f/fb/Eyjafjallaj%C3%B6kull-bjarmason.ogg)), meaning *island-*[*mountain*](http://en.wikipedia.org/wiki/Mountain) *glacier,* may approximate Icelandic phonology, but such information will likely be too much for English readers, who may need to reference the name using what is at best an approximate pronunciation anyway. (Often an English version of a foreign name will try to employ translation in combination with transcription, hence *Eyja-fjalla glacier* (['eija-f'jala] *glacier*) suffices.)

**Phoneme**

From Wikipedia, the free encyclopedia

Jump to: [navigation](http://en.wikipedia.org/wiki/Phoneme#mw-head), [search](http://en.wikipedia.org/wiki/Phoneme#p-search)

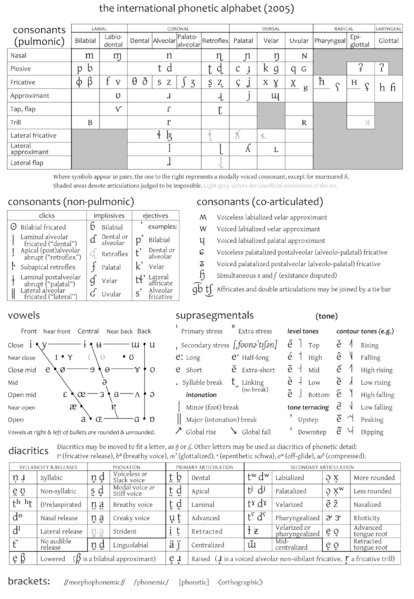
This article is about the speech unit. For the JavaME library, see [phoneME](http://en.wikipedia.org/wiki/PhoneME).

A **phoneme** is a basic element of a given [language](http://en.wikipedia.org/wiki/Language) or [dialect](http://en.wikipedia.org/wiki/Dialect), from which words in that language or dialect are analyzed as being built up. The phoneme is defined by the [International Phonetic Association](http://en.wikipedia.org/wiki/International_Phonetic_Association) as "the smallest [segmental](http://en.wikipedia.org/wiki/Segment_%28linguistics%29) unit of sound employed to form meaningful contrasts between utterances".[[1]](http://en.wikipedia.org/wiki/Phoneme#cite_note-IPA-0)

Within [linguistics](http://en.wikipedia.org/wiki/Linguistics) there are differing views as to exactly what phonemes are and how a given language should be analyzed in phonemic terms. However a phoneme is generally regarded as an [abstraction](http://en.wikipedia.org/wiki/Abstraction) of a set (or [equivalence class](http://en.wikipedia.org/wiki/Equivalence_class)) of speech sounds ([*phones*](http://en.wikipedia.org/wiki/Phone_%28phonetics%29)) which are perceived as equivalent to each other in a given language. For example, in English, the "k" sounds in the words *kit* and *skill* are not identical (as described [below](http://en.wikipedia.org/wiki/Phoneme#Assignment_of_speech_sounds_to_phonemes)), but they are perceived as the same sound by speakers of the language, and are therefore both considered to represent a single phoneme, /k/. Different speech sounds representing the same phoneme are known as [allophones](http://en.wikipedia.org/wiki/Allophone). Thus phonemes are often considered to provide an [underlying representation](http://en.wikipedia.org/wiki/Underlying_representation) for words, while speech sounds make up the corresponding surface form.

The study of systems of phonemes is a major component of the branch of linguistics called [phonology](http://en.wikipedia.org/wiki/Phonology).

|  |
| --- |
| **Contents**   * [1 Notation](http://en.wikipedia.org/wiki/Phoneme#Notation) * [2 Assignment of speech sounds to phonemes](http://en.wikipedia.org/wiki/Phoneme#Assignment_of_speech_sounds_to_phonemes)   + [2.1 Minimal pairs](http://en.wikipedia.org/wiki/Phoneme#Minimal_pairs) * [3 Other features with phonemic status](http://en.wikipedia.org/wiki/Phoneme#Other_features_with_phonemic_status) * [4 Distribution of allophones](http://en.wikipedia.org/wiki/Phoneme#Distribution_of_allophones) * [5 Background and related ideas](http://en.wikipedia.org/wiki/Phoneme#Background_and_related_ideas) * [6 Restrictions on occurrence](http://en.wikipedia.org/wiki/Phoneme#Restrictions_on_occurrence) * [7 Biuniqueness](http://en.wikipedia.org/wiki/Phoneme#Biuniqueness) * [8 Neutralization and archiphonemes](http://en.wikipedia.org/wiki/Phoneme#Neutralization_and_archiphonemes) * [9 Morphophonemes](http://en.wikipedia.org/wiki/Phoneme#Morphophonemes) * [10 Numbers of phonemes in different languages](http://en.wikipedia.org/wiki/Phoneme#Numbers_of_phonemes_in_different_languages) * [11 Correspondence between letters and phonemes](http://en.wikipedia.org/wiki/Phoneme#Correspondence_between_letters_and_phonemes) * [12 Phonemes in sign languages](http://en.wikipedia.org/wiki/Phoneme#Phonemes_in_sign_languages) * [13 See also](http://en.wikipedia.org/wiki/Phoneme#See_also) * [14 References](http://en.wikipedia.org/wiki/Phoneme#References) |

****

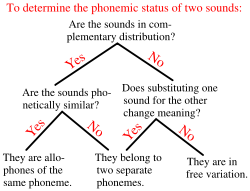
**Notation**

Phonemes are conventionally placed between [slashes](http://en.wikipedia.org/wiki/Slash_%28punctuation%29) in transcription, whereas speech sounds (phones) are placed between [square brackets](http://en.wikipedia.org/wiki/Square_bracket). Thus /pʊʃ/ represents a sequence of three phonemes /p/, /ʊ/, /ʃ/ (the word *push* in standard English), while [pʰʊʃ] represents the phonetic sequence of sounds [pʰ] ([aspirated](http://en.wikipedia.org/wiki/Aspiration_%28phonetics%29) "p"), [ʊ], [ʃ] (the usual pronunciation of *push*).

(Another similar convention is the use of angle brackets to enclose the units of [orthography](http://en.wikipedia.org/wiki/Orthography), namely [graphemes](http://en.wikipedia.org/wiki/Grapheme); for example, ⟨f⟩ represents the written letter (grapheme) *f*.)

The symbols used for particular phonemes are often taken from the [International Phonetic Alphabet](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet) (IPA), the same set of symbols that are most commonly used for phones. (For computer typing purposes, systems such as [X-SAMPA](http://en.wikipedia.org/wiki/X-SAMPA) and [Kirshenbaum](http://en.wikipedia.org/wiki/Kirshenbaum) exist to represent IPA symbols in [plain text](http://en.wikipedia.org/wiki/Plain_text).) However descriptions of particular languages may use different conventional symbols to represent the phonemes of those languages. For languages whose writing systems employ the [phonemic principle](http://en.wikipedia.org/wiki/Phonemic_principle), ordinary letters may be used to denote phonemes, although this approach is often hampered by the complexity of the relationship between orthography and pronunciation (see [Correspondence between letters and phonemes](http://en.wikipedia.org/wiki/Phoneme#Correspondence_between_letters_and_phonemes) below).

**Assignment of speech sounds to phonemes**

[](http://en.wikipedia.org/wiki/File:Phoneme-allophone-determination-chart.svg)

[http://bits.wikimedia.org/skins-1.20wmf1/common/images/magnify-clip.png](http://en.wikipedia.org/wiki/File:Phoneme-allophone-determination-chart.svg)

A simplified procedure for determining whether two sounds represent the same or different phonemes

A phoneme is a sound or a group of different sounds perceived to have the same function by speakers of the language or dialect in question. An example is the [English](http://en.wikipedia.org/wiki/English_language) phoneme /k/, which occurs in words such as ***c****at*, ***k****it*, *s****ch****ool*, *s****k****ill*. Although most native speakers do not notice this, in most English dialects the "c/k" sounds in these words are not identical: in ***c****at* and ***k****it* the sound is [aspirated](http://en.wikipedia.org/wiki/Aspiration_%28phonetics%29), while in *s****ch****ool* and *s****k****ill* it is unaspirated (listen to U.S. pronunciations of [About this sound](http://en.wikipedia.org/wiki/File:En-us-kit.ogg) [*kit*](http://upload.wikimedia.org/wikipedia/commons/8/85/En-us-kit.ogg) ([help](http://en.wikipedia.org/wiki/Wikipedia:Media_help)·[info](http://en.wikipedia.org/wiki/File:En-us-kit.ogg)) and [About this sound](http://en.wikipedia.org/wiki/File:En-us-skill.ogg) [*skill*](http://upload.wikimedia.org/wikipedia/commons/9/9b/En-us-skill.ogg) ([help](http://en.wikipedia.org/wiki/Wikipedia:Media_help)·[info](http://en.wikipedia.org/wiki/File:En-us-skill.ogg))). The words therefore contain different *speech sounds*, or [*phones*](http://en.wikipedia.org/wiki/Phone_%28phonetics%29), transcribed [kʰ] for the aspirated form, [k] for the unaspirated one. These different sounds are nonetheless considered to belong to the same phoneme, because if a speaker used one instead of the other, the meaning of the word would not change: using the aspirated form [kʰ] in *skill* might sound odd, but the word would still be recognized. By contrast, some other sounds would cause a change in meaning if substituted: for example, substitution of the sound [t] would produce the different word *s****t****ill*, and that sound must therefore be considered to represent a different phoneme (the phoneme /t/).

The above shows that in English, [k] and [kʰ] are [allophones](http://en.wikipedia.org/wiki/Allophones) of a single phoneme /k/. In some languages, however, [kʰ] and [k] are perceived by native speakers as different sounds, and substituting one for the other can change the meaning of a word; this means that in those languages, the two sounds represent different phonemes. For example, in [Icelandic](http://en.wikipedia.org/wiki/Icelandic_language), [kʰ] is the first sound of *kátur* meaning "cheerful", while [k] is the first sound of *gátur* meaning "riddles". Icelandic therefore has two separate phonemes /kʰ/ and /k/.

**Minimal pairs**

A pair of words like *kátur* and *gátur* (above) that differ only in one phone is called a [minimal pair](http://en.wikipedia.org/wiki/Minimal_pair) for the two alternative phones in question (in this case, [kʰ] and [k]). The existence of minimal pairs is a common test to decide whether two phones represent different phonemes or are allophones of the same phoneme. To take another example, the minimal pair ***t****ip* and ***d****ip* illustrates that in English, [t] and [d] belong to separate phonemes, /t/ and /d/; since these two words have different meanings, English speakers must be conscious of the distinction between the two sounds. In other languages, though, including [Korean](http://en.wikipedia.org/wiki/Korean_language), even though both sounds [t] and [d] occur, no such minimal pairs exist. The lack of minimal pairs distinguishing [t] and [d] in Korean provides evidence that in this language they are allophones of a single phoneme /t/. (The word /tata/ is pronounced [tada], for example. That is, when they hear this word, Korean speakers perceive the same sound in both the beginning and middle of the word, whereas an English speaker would perceive different sounds in these two locations.)

However, the absence of minimal pairs for a given pair of phones does not always mean that they belong to the same phoneme. They may be too dissimilar phonetically for it to be conceivable that speakers perceive them as the same sound; for example, English has no minimal pairs for the sounds [h] (as in ***h****at*) and [ŋ] (as in *ba****ng***), but they are so dissimilar that they are considered separate phonemes. There may also be "near minimal pairs" or other data which show that speakers of the language perceive two sounds as significantly different even if no exact minimal pairs exist in the lexicon.

**Other features with phonemic status**

While phonemes are normally conceived of as abstractions of discrete [segmental](http://en.wikipedia.org/wiki/Segment_%28linguistics%29) speech sounds (vowels and consonants), there are other features of pronunciation – principally [tone](http://en.wikipedia.org/wiki/Tone_%28linguistics%29) and [stress](http://en.wikipedia.org/wiki/Stress_%28linguistics%29) – which in some languages can change the meaning of words in the way that phoneme contrasts do, and are consequently called *phonemic* features of those languages.

*Phonemic stress* is encountered in languages such as English. For example, the word *invite* stressed on the second syllable is a verb, but when stressed on the first syllable (without changing any of the individual sounds) it becomes a noun. The position of the stress in the word affects the meaning, and therefore a full phonemic specification (providing enough detail to enable the word to be pronounced unambiguously) would include indication of the position of the stress: /ɪnˈvaɪt/ for the verb, /ˈɪnvaɪt/ for the noun. In other languages, such as [French](http://en.wikipedia.org/wiki/French_language), word stress cannot have this function (its position is generally predictable) and is therefore not phonemic (and is not usually indicated in dictionaries).

*Phonemic tones* are found in languages such as [Mandarin Chinese](http://en.wikipedia.org/wiki/Mandarin_Chinese), in which a given syllable can have five different tonal pronunciations. For example, *mā* (high level pitch) means "mom", *má* (rising pitch) means "hemp", *mǎ* (falling then rising) means "horse", *mà* (falling) means "scold", and *ma* (neutral tone) is an [interrogative particle](http://en.wikipedia.org/wiki/Interrogative_particle). The tone "phonemes" in such languages are sometimes called *tonemes*. Languages such as English do not have phonemic tone, although they use [intonation](http://en.wikipedia.org/wiki/Intonation_%28linguistics%29) for functions such as emphasis and attitude.

**Distribution of allophones**

When a phoneme has more than one [allophone](http://en.wikipedia.org/wiki/Allophone), the one actually heard at a given occurrence of that phoneme may be dependent on the phonetic environment (surrounding sounds) – allophones which normally cannot appear in the same environment are said to be in [complementary distribution](http://en.wikipedia.org/wiki/Complementary_distribution). In other cases the choice of allophone may be dependent on the individual speaker or other unpredictable factors – such allophones are said to be in [free variation](http://en.wikipedia.org/wiki/Free_variation).

**Background and related ideas**

The term *phonème* (from the [Greek](http://en.wikipedia.org/wiki/Greek_language): φώνημα, *phōnēma*, "a sound uttered") was reportedly first used by [A. Dufriche-Desgenettes](http://en.wikipedia.org/wiki/A._Dufriche-Desgenettes) in 1873, but it referred only to a speech sound. The term *phoneme* as an [abstraction](http://en.wikipedia.org/wiki/Abstraction) was developed by the Polish linguist [Jan Niecisław Baudouin de Courtenay](http://en.wikipedia.org/wiki/Jan_Niecis%C5%82aw_Baudouin_de_Courtenay) and his student [Mikołaj Kruszewski](http://en.wikipedia.org/wiki/Miko%C5%82aj_Kruszewski) during 1875–1895. The term used by these two was *fonema*, the basic unit of what they called *psychophonetics*. The concept of the phoneme was then elaborated in the works of [Nikolai Trubetzkoi](http://en.wikipedia.org/wiki/Nikolai_Trubetzkoi) and others of the [Prague School](http://en.wikipedia.org/wiki/Prague_School) (during the years 1926–1935), and in those of [structuralists](http://en.wikipedia.org/wiki/Structuralism) like [Ferdinand de Saussure](http://en.wikipedia.org/wiki/Ferdinand_de_Saussure), [Edward Sapir](http://en.wikipedia.org/wiki/Edward_Sapir), and [Leonard Bloomfield](http://en.wikipedia.org/wiki/Leonard_Bloomfield). Some structuralists wished to eliminate a cognitive or psycholinguistic function for the phoneme.[[*citation needed*](http://en.wikipedia.org/wiki/Wikipedia:Citation_needed)]

Later, it was also used in [generative linguistics](http://en.wikipedia.org/wiki/Generative_linguistics), most famously by [Noam Chomsky](http://en.wikipedia.org/wiki/Noam_Chomsky) and [Morris Halle](http://en.wikipedia.org/wiki/Morris_Halle), and remains central to many accounts of the development of modern [phonology](http://en.wikipedia.org/wiki/Phonology). As a theoretical concept or model, though, it has been supplemented and even replaced by others.

Some linguists (such as [Roman Jakobson](http://en.wikipedia.org/wiki/Roman_Jakobson), [Morris Halle](http://en.wikipedia.org/wiki/Morris_Halle), and [Noam Chomsky](http://en.wikipedia.org/wiki/Noam_Chomsky)) consider phonemes to be further decomposable into [features](http://en.wikipedia.org/wiki/Distinctive_feature), such features being the true minimal constituents of language. Features overlap each other in time, as do [suprasegmental](http://en.wikipedia.org/wiki/Suprasegmental) phonemes in oral language and many phonemes in sign languages. Features could be designated as [acoustic](http://en.wikipedia.org/wiki/Acoustic_phonetics) (Jakobson) or [articulatory](http://en.wikipedia.org/wiki/Articulatory_phonetics) (Halle & Chomsky) in nature.

In some languages, the term [chroneme](http://en.wikipedia.org/wiki/Chroneme) may be used for contrastive length or *duration* of phonemes. In languages in which [tones](http://en.wikipedia.org/wiki/Tone_%28linguistics%29) are phonemic, the tone phonemes may be called [tonemes](http://en.wikipedia.org/wiki/Toneme). Not all scholars working on such languages use these terms.

By analogy with the phoneme, linguists have proposed other sorts of underlying objects, giving them names with the suffix *-eme*, such as [*morpheme*](http://en.wikipedia.org/wiki/Morpheme) and [*grapheme*](http://en.wikipedia.org/wiki/Grapheme). These are sometimes called [emic units](http://en.wikipedia.org/wiki/Emic_unit). The latter term was first used by [Kenneth Pike](http://en.wikipedia.org/wiki/Kenneth_Pike), who also generalized the concepts of [emic and etic](http://en.wikipedia.org/wiki/Emic_and_etic) description (from *phonemic* and *phonetic* respectively) to applications outside linguistics.

**Restrictions on occurrence**

Main article: [Phonotactics](http://en.wikipedia.org/wiki/Phonotactics)

Languages do not generally allow words or [syllables](http://en.wikipedia.org/wiki/Syllable) to be built of any arbitrary sequences of phonemes; there are [phonotactic](http://en.wikipedia.org/wiki/Phonotactics) restrictions on which sequences of phonemes are possible and in which environments certain phonemes can occur. Phonemes that are significantly limited by such restrictions may be called *restricted phonemes*. Examples of such restrictions in English include:

* /ŋ/, as in *si****ng***, occurs only at the end of a syllable, never at the beginning (in many other languages, such as [Swahili](http://en.wikipedia.org/wiki/Swahili_language) or [Thai](http://en.wikipedia.org/wiki/Thai_language), /ŋ/ can appear word-initially).
* /h/ occurs only before vowels and at the beginning of a syllable, never at the end (a few languages, such as [Arabic](http://en.wikipedia.org/wiki/Arabic_language), or [Romanian](http://en.wikipedia.org/wiki/Romanian_Language) allow /h/ syllable-finally).
* In many American dialects with the [cot–caught merger](http://en.wikipedia.org/wiki/Phonological_history_of_the_low_back_vowels#Cot.E2.80.93caught_merger), /ɔ/ occurs only before /r/ and /l/ (and in the [diphthong](http://en.wikipedia.org/wiki/Diphthong) [ɔɪ] if this is not interpreted as a single phoneme).
* In [non-rhotic dialects](http://en.wikipedia.org/wiki/Rhotic_and_non-rhotic_accents), /r/ can only occur before a vowel, never at the end of a word or before a consonant.
* /w/ and /j/ occur only before a vowel, never at the end of a syllable (except in interpretations where a word like *boy* is analyzed as /bɔj/).

Some phonotactic restrictions can alternatively be analyzed as cases of neutralization. See [Neutralization and archiphonemes](http://en.wikipedia.org/wiki/Phoneme#Neutralization_and_archiphonemes) below, particularly the example of the occurrence of the three English nasals before stops.

**Biuniqueness**

**Biuniqueness** is a requirement of classic [structuralist](http://en.wikipedia.org/wiki/Structuralism_%28linguistics%29) phonemics. It means that a given [phone](http://en.wikipedia.org/wiki/Phone_%28phonetics%29), wherever it occurs, must unambiguously be assigned to one and only one phoneme. In other words, the mapping between phones and phonemes is required to be many-to-one rather than [many-to-many](http://en.wikipedia.org/wiki/Many-to-many). The notion of biuniqueness was controversial among some pre-[generative](http://en.wikipedia.org/wiki/Generative_linguistics) linguists and was prominently challenged by [Morris Halle](http://en.wikipedia.org/wiki/Morris_Halle) and [Noam Chomsky](http://en.wikipedia.org/wiki/Noam_Chomsky) in the late 1950s and early 1960s.

An example of the problems arising from the biuniqueness requirement is provided by the phenomenon of [flapping](http://en.wikipedia.org/wiki/Flapping) in [North American English](http://en.wikipedia.org/wiki/North_American_English). This may cause either /t/ or /d/ (in the appropriate environments) to be realized with the phone [ɾ] (an [alveolar flap](http://en.wikipedia.org/wiki/Alveolar_flap)). For example, the same flap sound may be heard in the words *hi****tt****ing* and *bi****dd****ing*, although it is clearly intended to realize the phoneme /t/ in the first word and /d/ in the second. This appears to contradict biuniqueness.

For further discussion of such cases, see the next section.

**Neutralization and archiphonemes**

Phonemes that are contrastive in certain environments may not be contrastive in all environments. In the environments where they do not contrast, the contrast is said to be **neutralized**. In these positions it may become less clear which phoneme a given phone represents. Some [phonologists](http://en.wikipedia.org/wiki/Phonology) prefer not to specify a unique phoneme in such cases, since to do so would mean providing redundant or even arbitrary information – instead they use the technique of [underspecification](http://en.wikipedia.org/wiki/Underspecification). An **archiphoneme** is an object sometimes used to represent an underspecified phoneme.

An example of neutralization is provided by the Russian vowels /a/ and /o/. These phonemes are contrasting in [stressed](http://en.wikipedia.org/wiki/Stress_%28linguistics%29) syllables, but in unstressed syllables the contrast is lost, since both are [reduced](http://en.wikipedia.org/wiki/Vowel_reduction) to the same sound, usually [ə] (for details, see [Vowel reduction in Russian](http://en.wikipedia.org/wiki/Vowel_reduction_in_Russian)). In order to assign such an instance of [ə] to one of the phonemes /a/ and /o/, it is necessary to consider [morphological](http://en.wikipedia.org/wiki/Morphology_%28linguistics%29) factors (such as which of the vowels occurs in other forms of the words, or which [inflectional](http://en.wikipedia.org/wiki/Inflection) pattern is followed). In some cases even this may not provide an unambiguous answer. A description using the approach of underspecification would not attempt to assign [ə] to a specific phoneme in some or all of these cases, although it might be assigned to an archiphoneme, written something like |A|, which reflects the two neutralized phonemes in this position.

A somewhat different example is found in English, with the three [nasal](http://en.wikipedia.org/wiki/Nasal_consonant) phonemes /m, n, ŋ/. In word-final position these all contrast, as shown by the minimal triplet *sum* /sʌm/, *sun* /sʌn/, *sung* /sʌŋ/. However, before a [plosive](http://en.wikipedia.org/wiki/Plosive) such as /p, t, k/ (provided there is no [morpheme](http://en.wikipedia.org/wiki/Morpheme) boundary between them), only one of the nasals is possible in any given position: /m/ before /p/, /n/ before /t/ or /d/, and /ŋ/ before /k/, as in *limp, lint, link* ( /lɪmp/, /lɪnt/, /lɪŋk/). The nasals are therefore not contrastive in these environments, and according to some theorists this makes it inappropriate to assign the nasal phones heard here to any one of the phonemes (even though, in this case, the phonetic evidence is unambiguous). Instead they may analyze these phones as belonging to a single archiphoneme, written something like |N|, and state the [underlying representations](http://en.wikipedia.org/wiki/Underlying_representation) of *limp, lint, link* to be |lɪNp|, |lɪNt|, |lɪNk|.

This latter type of analysis is often associated with [Nikolai Trubetzkoy](http://en.wikipedia.org/wiki/Nikolai_Trubetzkoy) of the [Prague school](http://en.wikipedia.org/wiki/Prague_school). Archiphonemes are often notated with a capital letter within pipes, as with the examples |A| and |N| given above. Other ways the second of these might be notated include |m-n-ŋ|, {m, n, ŋ}, or |n\*|.

Another example from English, but this time involving complete phonetic convergence as in the Russian example, is the flapping of /t/ and /d/ in some American English (described above under [Biuniqueness](http://en.wikipedia.org/wiki/Phoneme#Biuniqueness)). Here the words *betting* and *bedding* might both be pronounced [ˈbɛɾɪŋ], and if a speaker applies such flapping consistently, it would be necessary to look for morphological evidence (the pronunciation of the related forms *bet* and *bed*, for example) in order to determine which phoneme the flap represents. As in the previous examples, some theorists would prefer not to make such a determination, and simply assign the flap in both cases to a single archiphoneme, written (for example) |D|.

For a special kind of neutralization proposed in [generative](http://en.wikipedia.org/wiki/Generative_linguistics) phonology, see [absolute neutralization](http://en.wikipedia.org/wiki/Absolute_neutralization).

**Morphophonemes**

Main article: [Morphophonology](http://en.wikipedia.org/wiki/Morphophonology)

A **morphophoneme** is a theoretical unit at a deeper level of abstraction than traditional phonemes, and is taken to be a unit from which [morphemes](http://en.wikipedia.org/wiki/Morpheme) are built up. A morphophoneme within a morpheme can be expressed in different ways in different [allomorphs](http://en.wikipedia.org/wiki/Allomorph) of that morpheme (according to [morphophonological](http://en.wikipedia.org/wiki/Morphophonological) rules). For example, the English plural morpheme *-s* appearing in words such as *cats* and *dogs* can be considered to consist of a single morphophoneme, which might be written (for example) //z// or |z|, and which is pronounced as [s] after most [voiceless consonants](http://en.wikipedia.org/wiki/Voiceless_consonant) (as in *cat****s***) and [z] in most other cases (as in *dog****s***).

**Numbers of phonemes in different languages**

A given language will use only a small subset of the many possible [sounds](http://en.wikipedia.org/wiki/Speech_sound) that the human [speech organs](http://en.wikipedia.org/wiki/Speech_organs) can produce, and (because of [allophony](http://en.wikipedia.org/wiki/Allophony)) the number of distinct phonemes will generally be smaller than the number of identifiably different sounds. Different languages vary considerably in the number of phonemes they have in their systems (although apparent variation may sometimes result from the different approaches taken by the linguists doing the analysis). The total phonemic inventory in languages varies from as few as 11 in [Rotokas](http://en.wikipedia.org/wiki/Rotokas) to as many as 112 in [!Xóõ](http://en.wikipedia.org/wiki/%21X%C3%B3%C3%B5) (including four tones).

The number of phonemically distinct [vowels](http://en.wikipedia.org/wiki/Vowel) can be as low as two, as in [Ubyx](http://en.wikipedia.org/wiki/Ubyx_language) and [Arrernte](http://en.wikipedia.org/wiki/Arrernte_language). At the other extreme, the [Bantu](http://en.wikipedia.org/wiki/Bantu_languages) language [Ngwe](http://en.wikipedia.org/wiki/Ngwe_language) has 14 vowel qualities, 12 of which may occur long or short, making 26 oral vowels, plus 6 nasalized vowels, long and short, making a total of 38 vowels; while [!Xóõ](http://en.wikipedia.org/wiki/%21X%C3%B3%C3%B5_language) achieves 31 pure vowels, not counting its additional variation by vowel length, by varying the [phonation](http://en.wikipedia.org/wiki/Phonation). As regards [consonant](http://en.wikipedia.org/wiki/Consonant) phonemes, [Rotokas](http://en.wikipedia.org/wiki/Rotokas_language) has only six, while !Xóõ has somewhere in the neighborhood of 77, and [Ubyx](http://en.wikipedia.org/wiki/Ubykh_phonology) 81. The [English language](http://en.wikipedia.org/wiki/English_language) uses a rather large set of 13 to 21 vowel phonemes, including diphthongs, although its 22 to 26 consonants are close to average.

Some languages, such as [French](http://en.wikipedia.org/wiki/French_language), have no phonemic [tone](http://en.wikipedia.org/wiki/Tone_%28linguistics%29) or [stress](http://en.wikipedia.org/wiki/Stress_%28linguistics%29), while several of the [Kam–Sui languages](http://en.wikipedia.org/wiki/Kam%E2%80%93Sui_languages) have nine tones, and one of the [Kru languages](http://en.wikipedia.org/wiki/Kru_languages), [Wobe](http://en.wikipedia.org/wiki/Wobe_language), has been claimed to have 14, though this is disputed.

The most common vowel system consists of the five vowels /i/, /e/, /a/, /o/, /u/. The most common consonants are /p/, /t/, /k/, /m/, /n/. Relatively few languages lack any of these, although it does happen: for example, [Arabic](http://en.wikipedia.org/wiki/Arabic) lacks /p/, [standard Hawaiian](http://en.wikipedia.org/wiki/Hawaiian_language) lacks /t/, [Mohawk](http://en.wikipedia.org/wiki/Mohawk_language) and [Tlingit](http://en.wikipedia.org/wiki/Tlingit_language) lack /p/ and /m/, [Hupa](http://en.wikipedia.org/wiki/Hupa_language) lacks both /p/ and a simple /k/, colloquial [Samoan](http://en.wikipedia.org/wiki/Samoan_language) lacks /t/ and /n/, while [Rotokas](http://en.wikipedia.org/wiki/Rotokas_language) and [Quileute](http://en.wikipedia.org/wiki/Quileute_language) lack /m/ and /n/.

**Correspondence between letters and phonemes**

Main article: [Phonemic orthography](http://en.wikipedia.org/wiki/Phonemic_orthography)

Phonemes are considered to be the basis for [alphabetic](http://en.wikipedia.org/wiki/Alphabet) writing systems. In such systems the written symbols ([graphemes](http://en.wikipedia.org/wiki/Grapheme)) represent, in principle, the phonemes of the language being written. However, because changes in the spoken language are often not accompanied by changes in the established [orthography](http://en.wikipedia.org/wiki/Orthography) (as well as other reasons, including [dialect](http://en.wikipedia.org/wiki/Dialect) differences, the effects of [morphophonology](http://en.wikipedia.org/wiki/Morphophonology) on orthography, and the use of foreign spellings for some [loanwords](http://en.wikipedia.org/wiki/Loanword)), the correspondence between spelling and pronunciation in a given language may be highly distorted; this is the case with English, for example. (Occasionally, though, such discrepancies are reduced through the establishment of [spelling pronunciations](http://en.wikipedia.org/wiki/Spelling_pronunciation).)

The correspondence between symbols and phonemes in alphabetic writing systems is not necessarily a [one-to-one correspondence](http://en.wikipedia.org/wiki/One-to-one_correspondence). A phoneme might be represented by a combination of two or more letters ([digraph](http://en.wikipedia.org/wiki/Digraph_%28orthography%29), [trigraph](http://en.wikipedia.org/wiki/Digraph_%28orthography%29), etc.), like <sh> in English or <sch> in [German](http://en.wikipedia.org/wiki/German_language) (both representing phonemes /ʃ/). Also a single letter may represent two phonemes, as the Russian letter [я](http://en.wikipedia.org/wiki/%D0%AF) in some positions. There may also exist spelling/pronunciation rules (such as those for the pronunciation of <c> in [Italian](http://en.wikipedia.org/wiki/Italian_language)) that further complicate the correspondence of letters to phonemes, although they need not affect the ability to predict the pronunciation from the spelling and vice versa, provided the rules are known.

**Phonemes in sign languages**

In [sign languages](http://en.wikipedia.org/wiki/Sign_language), the basic elements of gesture and location were formerly called [*cheremes*](http://en.wikipedia.org/wiki/Chereme) or *cheiremes* but they are now generally referred to as phonemes, as with spoken languages.

Sign language phonemes may be classified as *Tab* (elements of location, from Latin *tabula*), *Dez* (the hand shape, from *designator*), *Sig* (the motion, from *signation*), and with some researchers, *Ori* (orientation). Facial expressions and mouthing are also phonemic.

[Stokoe notation](http://en.wikipedia.org/wiki/Stokoe_notation) is used by researchers to denote the phonemes of sign languages. Originally developed for [American Sign Language](http://en.wikipedia.org/wiki/American_Sign_Language), it has also been applied to [British Sign Language](http://en.wikipedia.org/wiki/British_Sign_Language) by Kyle and Woll, and to [Australian Aboriginal sign languages](http://en.wikipedia.org/wiki/Australian_Aboriginal_sign_languages) by Adam Kendon. Other sign notations, such as the [Hamburg Notation System](http://en.wikipedia.org/wiki/Hamburg_Notation_System) and [SignWriting](http://en.wikipedia.org/wiki/SignWriting), are phonetic scripts capable of writing any sign language. However, because they are not constrained by [phonology](http://en.wikipedia.org/wiki/Phonology), they do not yield a specific spelling for a sign. The SignWriting form, for example, will be different depending on whether the signer is left or right-handed, even though this makes no difference to the meaning of the sign.

**See also**

|  |  |
| --- | --- |
| * [Alphabetic principle](http://en.wikipedia.org/wiki/Alphabetic_principle) * [Alternation (linguistics)](http://en.wikipedia.org/wiki/Alternation_%28linguistics%29) * [Complementary distribution](http://en.wikipedia.org/wiki/Complementary_distribution) * [Diphone](http://en.wikipedia.org/wiki/Diphone) * [Diaphoneme](http://en.wikipedia.org/wiki/Diaphoneme) * [Emic and etic](http://en.wikipedia.org/wiki/Emic_and_etic) * [Free variation](http://en.wikipedia.org/wiki/Free_variation) * [Initial-stress-derived noun](http://en.wikipedia.org/wiki/Initial-stress-derived_noun) * [Minimal pair](http://en.wikipedia.org/wiki/Minimal_pair) * [Morphophonology](http://en.wikipedia.org/wiki/Morphophonology) | * [Phone](http://en.wikipedia.org/wiki/Phone_%28phonetics%29) * [Phonemic differentiation](http://en.wikipedia.org/wiki/Phonemic_differentiation) * [Phonemic orthography](http://en.wikipedia.org/wiki/Phonemic_orthography) * [Phonology](http://en.wikipedia.org/wiki/Phonology) * [Sphoṭa](http://en.wikipedia.org/wiki/Spho%E1%B9%ADa) * [Phonotactics](http://en.wikipedia.org/wiki/Phonotactics) * [Toneme](http://en.wikipedia.org/wiki/Tone_%28linguistics%29) * [Triphone](http://en.wikipedia.org/wiki/Triphone) * [Viseme](http://en.wikipedia.org/wiki/Viseme) |

**References**

|  |  |
| --- | --- |
| http://upload.wikimedia.org/wikipedia/en/thumb/4/4a/Commons-logo.svg/30px-Commons-logo.svg.png | Wikimedia Commons has media related to: [***Phonemes***](http://commons.wikimedia.org/wiki/Category:Phonemes) |

* 1. [**^**](http://en.wikipedia.org/wiki/Phoneme#cite_ref-IPA_0-0) International Phonetic Association (1999), ["Phonetic description and the IPA chart"](http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=0521637511), *Handbook of the International Phonetic Association: a guide to the use of the international phonetic alphabet*, Cambridge University Press, [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [9780521637510](http://en.wikipedia.org/wiki/Special:BookSources/9780521637510)

**Syllable**

From Wikipedia, the free encyclopedia

Jump to: [navigation](http://en.wikipedia.org/wiki/Syllable#mw-head), [search](http://en.wikipedia.org/wiki/Syllable#p-search)

For the computer operating system, see [Syllable (operating system)](http://en.wikipedia.org/wiki/Syllable_%28operating_system%29).

A **syllable** is a unit of organization for a sequence of [speech](http://en.wikipedia.org/wiki/Speech_communication) sounds. For example, the word *water* is composed of two syllables: *wa* and *ter*. A syllable is typically made up of a [syllable nucleus](http://en.wikipedia.org/wiki/Syllable_nucleus) (most often a [vowel](http://en.wikipedia.org/wiki/Vowel)) with optional initial and final margins (typically, [consonants](http://en.wikipedia.org/wiki/Consonant)).

Syllables are often considered the [phonological](http://en.wikipedia.org/wiki/Phonology) "building blocks" of [words](http://en.wikipedia.org/wiki/Word). They can influence the rhythm of a [language](http://en.wikipedia.org/wiki/Language), its [prosody](http://en.wikipedia.org/wiki/Prosody_%28linguistics%29), its [poetic](http://en.wikipedia.org/wiki/Poetry) meter and its [stress](http://en.wikipedia.org/wiki/Lexical_stress) patterns.

[Syllabic writing](http://en.wikipedia.org/wiki/Syllabic_writing) began several hundred years before the [first letters](http://en.wikipedia.org/wiki/Middle_Bronze_Age_alphabets). The earliest recorded syllables are on tablets written around 2800 BC in the [Sumerian](http://en.wikipedia.org/wiki/Sumer) city of [Ur](http://en.wikipedia.org/wiki/Ur). This shift from [pictograms](http://en.wikipedia.org/wiki/Pictogram) to syllables has been called "the most important advance in the [history of writing](http://en.wikipedia.org/wiki/History_of_writing)".[[1]](http://en.wikipedia.org/wiki/Syllable#cite_note-0)

A word that consists of a single syllable (like [English](http://en.wikipedia.org/wiki/English_language) *dog*) is called a **monosyllable** (and is said to be *monosyllabic*). Similar terms include **disyllable** (and *disyllabic*) for a word of two syllables; **trisyllable** (and *trisyllabic*) for a word of three syllables; and **polysyllable** (and *polysyllabic*), which may refer either to a word of more than three syllables or to any word of more than one syllable.

|  |
| --- |
| **Contents**   * [1 Syllable structure](http://en.wikipedia.org/wiki/Syllable#Syllable_structure)   + [1.1 Onset](http://en.wikipedia.org/wiki/Syllable#Onset)   + [1.2 Nucleus](http://en.wikipedia.org/wiki/Syllable#Nucleus)   + [1.3 Coda](http://en.wikipedia.org/wiki/Syllable#Coda)   + [1.4 Rhyme](http://en.wikipedia.org/wiki/Syllable#Rhyme)   + [1.5 Medial and final](http://en.wikipedia.org/wiki/Syllable#Medial_and_final)   + [1.6 Tone](http://en.wikipedia.org/wiki/Syllable#Tone)   + [1.7 Syllable weight](http://en.wikipedia.org/wiki/Syllable#Syllable_weight)   + [1.8 A classical definition](http://en.wikipedia.org/wiki/Syllable#A_classical_definition) * [2 Syllables and suprasegmentals](http://en.wikipedia.org/wiki/Syllable#Syllables_and_suprasegmentals) * [3 Syllables and phonotactic constraints](http://en.wikipedia.org/wiki/Syllable#Syllables_and_phonotactic_constraints) * [4 Syllabification](http://en.wikipedia.org/wiki/Syllable#Syllabification) * [5 Syllable division and ambisyllabicity](http://en.wikipedia.org/wiki/Syllable#Syllable_division_and_ambisyllabicity) * [6 Syllables and stress](http://en.wikipedia.org/wiki/Syllable#Syllables_and_stress) * [7 Syllables and vowel tenseness](http://en.wikipedia.org/wiki/Syllable#Syllables_and_vowel_tenseness) * [8 Syllable-less languages](http://en.wikipedia.org/wiki/Syllable#Syllable-less_languages) * [9 See also](http://en.wikipedia.org/wiki/Syllable#See_also) * [10 References](http://en.wikipedia.org/wiki/Syllable#References) * [11 Sources and recommended reading](http://en.wikipedia.org/wiki/Syllable#Sources_and_recommended_reading) * [12 External links](http://en.wikipedia.org/wiki/Syllable#External_links) |

**Syllable structure**

[](http://en.wikipedia.org/wiki/File:Syllable.svg)

[http://bits.wikimedia.org/skins-1.20wmf1/common/images/magnify-clip.png](http://en.wikipedia.org/wiki/File:Syllable.svg)

Tree representation of a syllable

In most theories of phonology, the general structure of a syllable (σ) consists of three segments:

[Onset](http://en.wikipedia.org/wiki/Syllable#Onset) (ω)

consonant, obligatory in some languages, optional or even restricted in others

[Nucleus](http://en.wikipedia.org/wiki/Syllable#Nucleus) (ν)

sonorant, obligatory in most languages

[Coda](http://en.wikipedia.org/wiki/Syllable#Coda) (κ)

consonant, optional in some languages, highly restricted or prohibited in others

The syllable is usually considered right-branching, i.e. nucleus and coda are grouped together as a "rime" and are only distinguished at the second level. However, in some traditional descriptions of certain languages[[*specify*](http://en.wikipedia.org/wiki/Wikipedia:Citing_sources)], the syllable is considered left-branching, i.e. onset and nucleus group below a higher-level unit, called a "body" or "core":

[Rime](http://en.wikipedia.org/wiki/Syllable#Rime) (ρ)

right branch, contrasts with onset, splits into nucleus and coda

[Body](http://en.wikipedia.org/wiki/Syllable#Body) or core

left branch, contrasts with coda, splits into onset and nucleus

In some theories the onset is strictly consonantal, thus necessitating another segment before the nucleus:

[Initial](http://en.wikipedia.org/wiki/Syllable#initial) (ι)

often termed *onset*, but leaving out semi-vowels

[Medial](http://en.wikipedia.org/wiki/Syllable#medial) (μ)

glide between initial, if any, and nucleus or rime

[Final](http://en.wikipedia.org/wiki/Syllable#final) (φ)

contrasts with initial, extended rime

Although every syllable has supra-segmental features, these are usually ignored if not semantically relevant, e.g. in [tonal languages](http://en.wikipedia.org/wiki/Tonal_language).

[Tone](http://en.wikipedia.org/wiki/Tone_%28linguistics%29) (τ)

may be carried by the syllable as a whole or by the rime

In some theories of phonology, these syllable structures are displayed as [tree diagrams](http://en.wikipedia.org/wiki/Tree_diagram) (similar to the trees found in some types of syntax). Not all phonologists agree that syllables have internal structure; in fact, some phonologists doubt the existence of the syllable as a theoretical entity.[[2]](http://en.wikipedia.org/wiki/Syllable#cite_note-1)

The *nucleus* is usually the vowel in the middle of a syllable. The *onset* is the sound or sounds occurring before the nucleus, and the *coda* (literally 'tail') is the sound or sounds that follow the nucleus. They are sometimes collectively known as the *shell*. The term *rime* covers the nucleus plus coda. In the one-syllable English word *cat*, the nucleus is *a* (the sound that can be shouted or sung on its own), the onset *c*, the coda *t*, and the rime *at*. This syllable can be abstracted as a *consonant-vowel-consonant* syllable, abbreviated *CVC*. Languages vary greatly in the restrictions on the sounds making up the onset, nucleus and coda of a syllable, according to what is termed a language's [phonotactics](http://en.wikipedia.org/wiki/Phonotactics).

**Onset**

Most syllables have an onset. Some languages restrict onsets to be only a single consonant, while others allow multiconsonant onsets according to various rules. For example, in English, onsets such as *pr-*, *pl-* and *tr-* are possible but *tl-* is not (except very marginally in foreign words such as [*Tlingit*](http://en.wikipedia.org/wiki/Tlingit)), and *sk-* is possible but *ks-* is not. In [Greek](http://en.wikipedia.org/wiki/Greek_language), however, both *ks-* and *tl-* are possible onsets, while contrarily in [Classical Arabic](http://en.wikipedia.org/wiki/Classical_Arabic) no multiconsonant onsets are allowed at all.

Some languages require all syllables to have an onset; in these languages a *null onset* such as in the English word "at" is not possible. This is less strange than it may appear at first, as most such languages allow syllables to begin with a phonemic [glottal stop](http://en.wikipedia.org/wiki/Glottal_stop) (the sound in the middle of English "uh-oh"). Furthermore, in English and most other languages, a word that begins with a vowel is automatically pronounced with an initial glottal stop when following a pause, whether or not a glottal stop occurs as a phoneme in the language. Consequently, few languages make a phonemic distinction between a word beginning with a vowel and a word beginning with a glottal stop followed by a vowel, since the distinction will generally only be audible following another word. (However, [Hawaiian](http://en.wikipedia.org/wiki/Hawaiian_language) and a number of other [Polynesian languages](http://en.wikipedia.org/wiki/Polynesian_languages) do make such a distinction; cf. Hawaiian /ahi/ "fire", /ʔahi/ "tuna".)

This means that the difference between a syllable with a null onset and one beginning with a glottal stop is often purely a difference of [phonological](http://en.wikipedia.org/wiki/Phonology) analysis, rather than the actual pronunciation of the syllable. In some cases, the pronunciation of a (putatively) vowel-initial word when following another word – particularly, whether or not a glottal stop is inserted – indicates whether the word should be considered to have a null onset. For example, many [Romance languages](http://en.wikipedia.org/wiki/Romance_languages) such as [Spanish](http://en.wikipedia.org/wiki/Spanish_language) never insert such a glottal stop, while [English](http://en.wikipedia.org/wiki/English_language) does so only some of the time, depending on factors such as conversation speed; in both cases, this suggests that the words in question are truly vowel-initial. But there are exceptions here, too. For example, [German](http://en.wikipedia.org/wiki/German_language) and [Arabic](http://en.wikipedia.org/wiki/Arabic_language) both require that a glottal stop be inserted between a word and a following, putatively vowel-initial word. Yet such words are said to begin with a vowel in German but a glottal stop in Arabic. The reason for this has to do with other properties of the two languages. For example, a glottal stop does not occur in other situations in German, e.g. before a consonant or at the end of word. On the other hand, in Arabic, not only does a glottal stop occur in such situations (e.g. Classical /saʔala/ "he asked", /raʔj/ "opinion", /dˤawʔ/ "light"), but it occurs in alternations that are clearly indicative of its phonemic status (cf. Classical /kaːtib/ "writer" vs. /maktuːb/ "written", /ʔaːkil/ "eater" vs. /maʔkuːl/ "eaten").

The writing system of a language may not correspond with the phonological analysis of the language in terms of its handling of (potentially) null onsets. For example, in some languages written in the [Latin alphabet](http://en.wikipedia.org/wiki/Latin_alphabet), an initial glottal stop is left unwritten; on the other hand, some languages written using non-Latin alphabets such as [abjads](http://en.wikipedia.org/wiki/Abjad) and [abugidas](http://en.wikipedia.org/wiki/Abugida) have a special [zero consonant](http://en.wikipedia.org/wiki/Zero_consonant) to represent a null onset. As an example, in [Hangul](http://en.wikipedia.org/wiki/Hangul), the alphabet of the [Korean language](http://en.wikipedia.org/wiki/Korean_language), a null onset is represented with ㅇ at the left or top section of a graph, as in 역 "station", pronounced *yeok*, where the [diphthong](http://en.wikipedia.org/wiki/Diphthong) *yeo* is the nucleus and *k* is the coda.

**Nucleus**

|  |  |
| --- | --- |
| **Examples of syllable nuclei** | |
| **Word** | **Nucleus** |
| c**a**t [kæt] | [æ] |
| b**e**d [bɛd] | [ɛ] |
| **o**de [oʊd] | [oʊ] |
| b**ee**t [bit] | [i] |
| b**i**te [baɪt] | [aɪ] |
| r**ai**n [reɪn] | [eɪ] |
| b**i**tt**e**n [ˈbɪt.ən] or [ˈbɪt.n] | [ɪ] [ə] or [n] |

Generally, every syllable requires a nucleus (sometimes called the *peak*), and the minimal syllable consists only of a nucleus, as in the English words "eye" or "owe". The syllable nucleus is usually a vowel, in the form of a [monophthong](http://en.wikipedia.org/wiki/Monophthong), [diphthong](http://en.wikipedia.org/wiki/Diphthong), or [triphthong](http://en.wikipedia.org/wiki/Triphthong), but sometimes is a [syllabic consonant](http://en.wikipedia.org/wiki/Syllabic_consonant). By far the most common syllabic consonants are [sonorants](http://en.wikipedia.org/wiki/Sonorant) like [l], [r], [m], [n] or [ŋ], but a few languages have so-called [*syllabic fricatives*](http://en.wikipedia.org/wiki/Syllabic_fricative), also known as *fricative vowels*. (In the context of [Chinese phonology](http://en.wikipedia.org/wiki/Chinese_phonology), the related but non-synonymous term *apical vowel* is commonly used.) [Mandarin Chinese](http://en.wikipedia.org/wiki/Mandarin_Chinese) is famous for having such sounds in at least some of its dialects, for example the [pinyin](http://en.wikipedia.org/wiki/Pinyin) syllables *sī shī rī*, sometimes pronounced [sź̩ ʂʐ̩́ ʐʐ̩́] respectively. A few languages, such as [Nuxalk](http://en.wikipedia.org/wiki/Nux%C3%A1lk_language) (Bella Coola), even allow [stop consonants](http://en.wikipedia.org/wiki/Stop_consonant) and [voiceless fricatives](http://en.wikipedia.org/wiki/Voiceless_fricative) as syllabic nuclei. However, linguists have analyzed this situation in various ways, some arguing that such syllables have no nucleus at all, and some arguing that the concept of "syllable" cannot clearly be applied at all to these languages. See the discussion below concerning [syllable-less languages](http://en.wikipedia.org/wiki/Syllable#Syllable-less_languages).

**Coda**

Main article: [Syllable coda](http://en.wikipedia.org/wiki/Syllable_coda)

A coda-less syllable of the form V, CV, CCV, etc. is called an **open syllable** (or *free syllable*), while a syllable that has a coda (VC, CVC, CVCC, etc.) is called a **closed syllable** (or *checked syllable*). Note that they have nothing to do with [open](http://en.wikipedia.org/wiki/Open_vowel) and [close vowels](http://en.wikipedia.org/wiki/Close_vowel). Almost all languages allow open syllables, but some, such as [Hawaiian](http://en.wikipedia.org/wiki/Hawaiian_language), do not have closed syllables.

In [English](http://en.wikipedia.org/wiki/English_language), consonants have been analyzed as acting simultaneously as the coda of one syllable and the onset of the following syllable, as in 'bellow' *bel-low,* a phenomenon known as **ambisyllabicity**. It is argued that words such as *arrow* [/ˈæroʊ/](http://en.wikipedia.org/wiki/Wikipedia:IPA_for_English) can't be divided into separately pronounceable syllables: neither /æ/ nor /ær/ is a possible independent syllable, and likewise with the other short vowels /ɛ ɪ ɒ ʌ ʊ/. However, Wells (1990) argues against ambisyllabicity in English, positing that consonants and consonant clusters are codas when after a stressed syllable followed by an unstressed syllable, or after a full vowel and followed by a reduced syllable, and are onsets in other contexts. (See [English phonology#Phonotactics](http://en.wikipedia.org/wiki/English_phonology#Phonotactics).)

**Rhyme**

Main article: [Syllable rhyme](http://en.wikipedia.org/wiki/Syllable_rhyme)

**Medial and final**

In the phonology of some East Asian languages, especially [Chinese](http://en.wikipedia.org/wiki/Chinese_language), the syllable structure is expanded to include an additional, optional segment known as a **medial**, which is located between the onset (often termed the *initial* in this context) and the rime. The medial is normally a [glide consonant](http://en.wikipedia.org/wiki/Glide_consonant), but reconstructions of [Old Chinese](http://en.wikipedia.org/wiki/Old_Chinese) generally include [liquid](http://en.wikipedia.org/wiki/Liquid_consonant) medials (/r/ in modern reconstructions, /l/ in older versions), and many reconstructions of [Middle Chinese](http://en.wikipedia.org/wiki/Middle_Chinese) include a medial contrast between /i/ and /j/, where the /i/ functions phonologically as a glide rather than as part of the nucleus. In addition, many reconstructions of both Old and Middle Chinese include complex medials such as /rj/, /ji/, /jw/ and /jwi/. The medial groups phonologically with the rime rather than the onset, and the combination of medial and rime is collectively known as the **final**.

**Tone**

In most languages, the pitch or pitch contour in which a syllable is pronounced conveys shades of meaning such as emphasis or surprise, or distinguishes a statement from a question. In [tonal languages](http://en.wikipedia.org/wiki/Tonal_languages), however, the pitch of a word affects the basic lexical meaning (e.g. "cat" vs. "dog") or grammatical meaning (e.g. past vs. present). In some languages, only the pitch itself (e.g. high vs. low) has this effect, while in others, especially East Asian languages such as [Chinese](http://en.wikipedia.org/wiki/Chinese_language), [Thai](http://en.wikipedia.org/wiki/Thai_language) or [Vietnamese](http://en.wikipedia.org/wiki/Vietnamese_language), the shape or contour (e.g. level vs. rising vs. falling) also needs to be distinguished.

**Syllable weight**

A [**heavy syllable**](http://en.wikipedia.org/wiki/Syllable_weight) is one with a *branching rime* or *branching nucleus* – this is a metaphor, based on the nucleus or coda having lines that branch in a tree diagram. In some languages, heavy syllables include both VV (branching nucleus) and VC (branching rime) syllables, contrasted with V, which is a **light syllable**. (A "branching nucleus" is a long vowel or [diphthong](http://en.wikipedia.org/wiki/Diphthong). A "branching rime" is a rime where the syllable ends in a consonant, also known as a *closed syllable*; generally, this means that either the nucleus is followed by two consonants or by a single, final consonant.) In other languages, only VV syllables are heavy, while both VC and V syllables are light. Some languages distinguish a third type of **superheavy syllable**, which consists of VVC syllables (with both a branching nucleus and rime) and/or VCC syllables (with a coda consisting of two or more consonants). In [moraic theory](http://en.wikipedia.org/wiki/Mora_%28linguistics%29), heavy syllables are said to have two moras, while light syllables are said to have one and superheavy syllables are said to have three. [Japanese](http://en.wikipedia.org/wiki/Japanese_language) is generally described this way.

Many languages forbid superheavy syllables, while a significant number forbid any heavy syllable. Some languages strive for consonant syllable weight; for example, in stressed, non-final syllables in [Italian](http://en.wikipedia.org/wiki/Italian_language), short vowels co-occur with closed syllables while long vowels co-occur with open syllables, so that all such syllables are heavy (not light or superheavy).

The difference between heavy and light frequently determines which syllables receive [stress](http://en.wikipedia.org/wiki/Stress_%28linguistics%29) – this is the case in [Latin](http://en.wikipedia.org/wiki/Latin) and [Arabic](http://en.wikipedia.org/wiki/Arabic_language), for example. The system of [poetic meter](http://en.wikipedia.org/wiki/Meter_%28poetry%29) in many classical languages, such as [Classical Greek](http://en.wikipedia.org/wiki/Classical_Greek), [Classical Latin](http://en.wikipedia.org/wiki/Classical_Latin) and [Sanskrit](http://en.wikipedia.org/wiki/Sanskrit_language), is based on syllable weight rather than stress (so-called *quantitative rhythm* or *quantitative meter*).

**A classical definition**

[Guilhem Molinier](http://en.wikipedia.org/w/index.php?title=Guilhem_Molinier&action=edit&redlink=1), a member of the [Consistori del Gay Saber](http://en.wikipedia.org/wiki/Consistori_del_Gay_Saber), which was the first literary academy in the world and held the [Floral Games](http://en.wikipedia.org/wiki/Floral_Games) to award the best [troubadour](http://en.wikipedia.org/wiki/Troubadour) with the *violeta d'aur* top prize, gave a definition of the syllable in his [*Leys d'amor*](http://en.wikipedia.org/w/index.php?title=Leys_d%27amor&action=edit&redlink=1) (1328–1337), a book aimed at regulating the then flourishing [Occitan](http://en.wikipedia.org/wiki/Occitan_language) poetry:

|  |  |
| --- | --- |
| *Sillaba votz es literals.* Segon los ditz gramaticals. En un accen pronunciada. Et en un trag: d'una alenada. | A syllable is the sound of several letters, According to those called grammarians, Pronounced in one accent And uninterruptedly: in one breath. |

**Syllables and suprasegmentals**

The domain of [suprasegmental](http://en.wikipedia.org/wiki/Suprasegmental) features is the syllable and not a specific sound, that is to say, they affect all the segments of a syllable:

* [Stress](http://en.wikipedia.org/wiki/Lexical_stress)
* [Tone](http://en.wikipedia.org/wiki/Tone_%28linguistics%29)

Sometimes [syllable length](http://en.wikipedia.org/wiki/Syllable_length) is also counted as a suprasegmental feature; for example, in some Germanic languages, long vowels may only exist with short consonants and vice versa. However, syllables can be analyzed as compositions of long and short phonemes, as in Finnish and Japanese, where consonant gemination and vowel length are independent.

**Syllables and phonotactic constraints**

[Phonotactic](http://en.wikipedia.org/wiki/Phonotactics) rules determine which sounds are allowed or disallowed in each part of the syllable. [English](http://en.wikipedia.org/wiki/English_language) allows very complicated syllables; syllables may begin with up to three consonants (as in *string* or *splash*), and occasionally end with as many as four (as in *prompts*). Many other languages are much more restricted; [Japanese](http://en.wikipedia.org/wiki/Japanese_language), for example, only allows /ɴ/ and a [chroneme](http://en.wikipedia.org/wiki/Chroneme) in a coda, and theoretically has no consonant clusters at all, as the onset is composed of at most one consonant.[[3]](http://en.wikipedia.org/wiki/Syllable#cite_note-2)

There are languages that forbid empty onsets, such as [Hebrew](http://en.wikipedia.org/wiki/Hebrew_language) and [Arabic](http://en.wikipedia.org/wiki/Arabic_language) (the names transliterated as "Israel", "Abraham", "Omar", "Ali" and "Abdullah", among many others, actually begin with semiconsonantic glides or with glottal or pharyngeal consonants). Conversely, some analyses of the [Arrernte language](http://en.wikipedia.org/wiki/Arrernte_language) of central Australia posit that no onsets are permitted at all in that language, all syllables being underlyingly of the shape VC(C).[[4]](http://en.wikipedia.org/wiki/Syllable#cite_note-3)

**Syllabification**

Main article: [Syllabification](http://en.wikipedia.org/wiki/Syllabification)

*Syllabification* is the separation of a word into syllables, whether spoken or written. In most languages, the actually spoken syllables are the basis of syllabification in writing too. Due to the very weak correspondence between sounds and letters in the spelling of modern English, for example, written syllabification in English has to be based mostly on etymological i.e. morphological instead of phonetic principles. English "written" syllables therefore do not correspond to the actually spoken syllables of the living language.

(*Syllabification* may also refer to the process of a consonant becoming a syllable nucleus.)

**Syllable division and ambisyllabicity**

Most commonly, a single consonant between vowels is grouped with the following syllable (i.e. /CV.CV/), while two consonants between vowels are split between syllables (i.e. /CVC.CV/). In some languages, however, such as [Old Church Slavonic](http://en.wikipedia.org/wiki/Old_Church_Slavonic), any group of consonants that can occur at the beginning of a word is grouped with the following syllable; hence, a word such as *pazdva* would be syllabified /pa.zdva/. (This allows the [phonotactics](http://en.wikipedia.org/wiki/Phonotactics) of the language to be defined as requiring open syllables.) Contrarily, in some languages, any group of consonants that can occur at the end of a word is grouped with the following syllable.

In English, it has been disputed whether certain consonants occurring between vowels (especially following a stressed syllable and preceding an unstressed syllable) should be grouped with the preceding or following syllable. For example, a word such as *better* is sometimes analyzed as /ˈbɛt.ər/ and sometimes /ˈbɛ.tər/. Some linguists have in fact asserted that such words are "ambisyllabic", with the consonant shared between the preceding and following syllables. However, Wells (2002)[[1]](http://www.phon.ucl.ac.uk/home/wells/syllabif.htm) argues that this is not a useful analysis, and that English syllabification is simply /ˈCVC(C).V/.

**Syllables and stress**

Syllable structure often interacts with stress. In [Latin](http://en.wikipedia.org/wiki/Latin), for example, stress is regularly determined by [syllable weight](http://en.wikipedia.org/wiki/Syllable_weight), a syllable counting as heavy if it has at least one of the following:

* a long vowel in its [nucleus](http://en.wikipedia.org/wiki/Syllable_nucleus)
* a [diphthong](http://en.wikipedia.org/wiki/Diphthong) in its nucleus
* one or more [coda](http://en.wikipedia.org/wiki/Syllable_coda)(e)

In each case the syllable is considered to have two [moras](http://en.wikipedia.org/wiki/Mora_%28linguistics%29).

**Syllables and vowel tenseness**

In most [Germanic languages](http://en.wikipedia.org/wiki/Germanic_language), [lax vowels](http://en.wikipedia.org/wiki/Tenseness) can occur only in closed syllables. Therefore, these vowels are also called [checked vowels](http://en.wikipedia.org/wiki/Checked_and_free_vowels), as opposed to the tense vowels that are called *free vowels* because they can occur even in open syllables.

**Syllable-less languages**

The notion of syllable is challenged by languages that allow long strings of consonants without any intervening vowel or sonorant. Languages of the Northwest coast of North America, including [Salishan](http://en.wikipedia.org/wiki/Salishan_languages) and [Wakashan languages](http://en.wikipedia.org/wiki/Wakashan_languages), are famous for this. For instance, these [Nuxálk](http://en.wikipedia.org/wiki/Nux%C3%A1lk_language) (Bella Coola) words contain only [obstruents](http://en.wikipedia.org/wiki/Obstruent):

[ɬχʷtɬtsxʷ] 'you spat on me'

[tsʼktskʷtsʼ] 'he arrived'

[xɬpʼχʷɬtɬpɬɬs] 'he had in his possession a bunchberry plant' *(Bagemihl 1991:589, 593, 627)*

[sxs] 'seal blubber'

In Bagemihl's survey of previous analyses, he finds that the word [tsʼktskʷtsʼ] would have been parsed into 0, 2, 3, 5, or 6 syllables depending which analysis is used. One analysis would consider all vowel and consonants segments as syllable nuclei, another would consider only a small subset as nuclei candidates, and another would simply deny the existence of syllables completely.

This type of phenomenon has also been reported in [Berber languages](http://en.wikipedia.org/wiki/Berber_languages) (such as Indlawn Tashlhiyt Berber), [Moroccan Arabic](http://en.wikipedia.org/wiki/Moroccan_Arabic) (apparently under Berber influence) and [Mon–Khmer languages](http://en.wikipedia.org/wiki/Mon%E2%80%93Khmer_languages) (such as [Semai](http://en.wikipedia.org/wiki/Semai_language), Temiar, [Kammu](http://en.wikipedia.org/wiki/Kammu_language)). This feature has also been reported in Ōgami, a Miyako Ryukyuan language.[[5]](http://en.wikipedia.org/wiki/Syllable#cite_note-4) Even in English there are a few utterances that have no vowels; for example, *shh* (meaning "be quiet") and *psst* (a sound used to attract attention).

*Indlawn Tashlhiyt Berber:*

[tftktst tfktstt] 'you sprained it and then gave it'

[rkkm] 'rot' (imperf.) *(Dell & Elmedlaoui 1985, 1988)*

*Semai:*

[kckmrʔɛːc] 'short, fat arms' *(Sloan 1988)*

**See also**

* [English phonology#Phonotactics](http://en.wikipedia.org/wiki/English_phonology#Phonotactics). Covers syllable structure in English.
* [Mora (linguistics)](http://en.wikipedia.org/wiki/Mora_%28linguistics%29)
* [List of the longest English words with one syllable](http://en.wikipedia.org/wiki/List_of_the_longest_English_words_with_one_syllable)
* [Phonology](http://en.wikipedia.org/wiki/Phonology)
* [Pitch accent](http://en.wikipedia.org/wiki/Pitch_accent)
* [Stress (linguistics)](http://en.wikipedia.org/wiki/Stress_%28linguistics%29)
* [Syllabary](http://en.wikipedia.org/wiki/Syllabary) writing system
* [Syllabic consonant](http://en.wikipedia.org/wiki/Syllabic_consonant)
* [Syllabification](http://en.wikipedia.org/wiki/Syllabification)
* [Timing (linguistics)](http://en.wikipedia.org/wiki/Timing_%28linguistics%29)
* [IPA symbols for syllables](http://en.wikipedia.org/wiki/International_Phonetic_Alphabet#Suprasegmentals)
* [Entering tone](http://en.wikipedia.org/wiki/Entering_tone)
* [Minor syllable](http://en.wikipedia.org/wiki/Minor_syllable)
* [Line (poetry)](http://en.wikipedia.org/wiki/Line_%28poetry%29)

**References**

* 1. [**^**](http://en.wikipedia.org/wiki/Syllable#cite_ref-0) [Geoffrey Blainey](http://en.wikipedia.org/wiki/Geoffrey_Blainey), [*A Short History of the World*](http://en.wikipedia.org/wiki/A_Short_History_of_the_World_%28Geoffrey_Blainey%29), p.87, citing J.T. Hooker et al., *Reading the Past: Ancient Writing from Cuneiform to the Alphabet*, [British Museum](http://en.wikipedia.org/wiki/British_Museum), 1993, Ch. 2
  2. [**^**](http://en.wikipedia.org/wiki/Syllable#cite_ref-1) See [CUNY Conference on the Syllable](http://www.cunyphonologyforum.net/syllable.php) for discussion of the theoretical existence of the syllable.
  3. [**^**](http://en.wikipedia.org/wiki/Syllable#cite_ref-2) Shibatani, Masayoshi (1987). "Japanese". In Bernard Comrie. *The World's Major Languages*. Oxford University Press. pp. 855–80. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-19-520521-9](http://en.wikipedia.org/wiki/Special:BookSources/0-19-520521-9).
  4. [**^**](http://en.wikipedia.org/wiki/Syllable#cite_ref-3) [Arrernte: a language with no syllable onsets](http://www.jstor.org/stable/4179048). Gavan Breen and Rob Pensalfini. *Linguistic Inquiry*. Vol. 30, No. 1 (1999), pp. 1-25. Massachusetts Institute of Technology.
  5. [**^**](http://en.wikipedia.org/wiki/Syllable#cite_ref-4) <http://halshs.archives-ouvertes.fr/docs/00/52/95/98/PDF/irl-ogami.pdf>

**Sources and recommended reading**

* [Bagemihl, Bruce](http://en.wikipedia.org/wiki/Bruce_Bagemihl) (1991). "Syllable structure in Bella Coola". *Linguistic Inquiry* **22**: 589–646.
* [Clements, George N.](http://en.wikipedia.org/wiki/George_N._Clements); [Keyser, Samuel J.](http://en.wikipedia.org/wiki/Samuel_Jay_Keyser). (1983). *CV phonology: A generative theory of the syllable*. Linguistic inquiry monographs (No. 9). Cambridge, MA: MIT Press. [ISBN 0-262-53047-3](http://en.wikipedia.org/wiki/Special:BookSources/0262530473) (pbk); [ISBN 0-262-03098-5](http://en.wikipedia.org/wiki/Special:BookSources/0262030985) (hb)
* Dell, François; Elmedlaoui, Mohamed (1985). "Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber". *Journal of African Languages and Linguistics* **7** (2): 105–130. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.1515/jall.1985.7.2.105](http://dx.doi.org/10.1515%2Fjall.1985.7.2.105).
* Dell, François; Elmedlaoui, Mohamed (1988). "Syllabic consonants in Berber: Some new evidence". *Journal of African Languages and Linguistics* **10**: 1–17. [doi](http://en.wikipedia.org/wiki/Digital_object_identifier):[10.1515/jall.1988.10.1.1](http://dx.doi.org/10.1515%2Fjall.1988.10.1.1).
* [Ladefoged, Peter](http://en.wikipedia.org/wiki/Peter_Ladefoged) (2001). *A course in phonetics* (4th ed.). [Fort Worth](http://en.wikipedia.org/wiki/Fort_Worth,_Texas): Harcourt College Publishers. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-15-507319-2](http://en.wikipedia.org/wiki/Special:BookSources/0-15-507319-2).

**Word**

From Wikipedia, the free encyclopedia

Jump to: [navigation](http://en.wikipedia.org/wiki/Word#mw-head), [search](http://en.wikipedia.org/wiki/Word#p-search)

For other uses, see [Word (disambiguation)](http://en.wikipedia.org/wiki/Word_%28disambiguation%29).

In [language](http://en.wikipedia.org/wiki/Language), a **word** is the smallest element that may be uttered in isolation with [semantic](http://en.wikipedia.org/wiki/Semantic) or [pragmatic](http://en.wikipedia.org/wiki/Pragmatics) content (with literal or practical [meaning](http://en.wikipedia.org/wiki/Meaning_%28linguistics%29)). This contrasts with a [morpheme](http://en.wikipedia.org/wiki/Morpheme), which is the smallest unit of meaning but will not necessarily stand on its own. A word may consist of a single morpheme (for example: *oh!, rock, red, quick, run, expect*), or several (*rocks, redness, quickly, running, unexpected*), whereas a morpheme may not be able to stand on its own as a word (in the words just mentioned, these are *-s, -ness, -ly, -ing, un-, -ed*).

A complex word will typically include a [root](http://en.wikipedia.org/wiki/Root_%28linguistics%29) and one or more [affixes](http://en.wikipedia.org/wiki/Affix) (*rock-s, red-ness, quick-ly, run-ning, un-expect-ed*), or more than one root in a [compound](http://en.wikipedia.org/wiki/Compound_word) (*black-board, rat-race*). Words can be put together to build larger elements of language, such as [phrases](http://en.wikipedia.org/wiki/Phrase) (*a red rock*), [clauses](http://en.wikipedia.org/wiki/Clause) (*I threw a rock*), and [sentences](http://en.wikipedia.org/wiki/Sentence_%28linguistics%29) (*He threw a rock too but he missed*).

The term *word* may refer to a spoken word or to a written word, or sometimes to the abstract concept behind either. [Spoken words](http://en.wikipedia.org/wiki/Spoken_words) are made up of units of sound called [phonemes](http://en.wikipedia.org/wiki/Phoneme), and written words of symbols called [graphemes](http://en.wikipedia.org/wiki/Grapheme), such as the letters of the English alphabet.

|  |
| --- |
| **Contents**   * [1 Definitions](http://en.wikipedia.org/wiki/Word#Definitions)   + [1.1 Semantic definition](http://en.wikipedia.org/wiki/Word#Semantic_definition)   + [1.2 Features](http://en.wikipedia.org/wiki/Word#Features)   + [1.3 Word boundaries](http://en.wikipedia.org/wiki/Word#Word_boundaries)     - [1.3.1 Orthography](http://en.wikipedia.org/wiki/Word#Orthography) * [2 Morphology](http://en.wikipedia.org/wiki/Word#Morphology) * [3 Philosophy](http://en.wikipedia.org/wiki/Word#Philosophy) * [4 Classes](http://en.wikipedia.org/wiki/Word#Classes) * [5 See also](http://en.wikipedia.org/wiki/Word#See_also) * [6 Notes](http://en.wikipedia.org/wiki/Word#Notes) * [7 References](http://en.wikipedia.org/wiki/Word#References) * [8 External links](http://en.wikipedia.org/wiki/Word#External_links) |

**Definitions**

Further information: [Lexeme](http://en.wikipedia.org/wiki/Lexeme) and [Lemma (morphology)](http://en.wikipedia.org/wiki/Lemma_%28morphology%29)

The ease or difficulty of deciphering a word depends on the language. [Dictionaries](http://en.wikipedia.org/wiki/Dictionaries) categorize a language's [lexicon](http://en.wikipedia.org/wiki/Lexicon) (i.e., its vocabulary) into [lemmas](http://en.wikipedia.org/wiki/Lemma_%28morphology%29). These can be taken as an indication of what constitutes a "word" in the opinion of the writers of that language.

**Semantic definition**

[Leonard Bloomfield](http://en.wikipedia.org/wiki/Leonard_Bloomfield) introduced the concept of "Minimal Free Forms" in 1926. Words are thought of as the smallest meaningful unit of [speech](http://en.wikipedia.org/wiki/Speech) that can stand by themselves.[[1]](http://en.wikipedia.org/wiki/Word#cite_note-0) This correlates phonemes (units of sound) to [lexemes](http://en.wikipedia.org/wiki/Lexeme) (units of meaning). However, some written words are not minimal free forms, as they make no sense by themselves (for example, *the* and *of*).[[2]](http://en.wikipedia.org/wiki/Word#cite_note-1)

Some semanticists have proposed a theory of so-called semantic primitives or [semantic primes](http://en.wikipedia.org/wiki/Semantic_primes), indefinable words representing fundamental concepts that are intuitively meaningful. According to this theory, semantic primes serve as the basis for describing the meaning, without circularity, of other words and their associated conceptual denotations.[[3]](http://en.wikipedia.org/wiki/Word#cite_note-wierzbicka96-2)

**Features**

In the [Minimalist](http://en.wikipedia.org/wiki/Minimalist_program) school of [theoretical syntax](http://en.wikipedia.org/wiki/Theoretical_syntax), words (also called *lexical items* in the literature) are construed as "bundles" of [linguistic features](http://en.wikipedia.org/wiki/Feature_%28linguistics%29) that are united into a structure with form and meaning.[[4]](http://en.wikipedia.org/wiki/Word#cite_note-3) For example, the word "bears" has semantic features (it denotes real-world objects, [bears](http://en.wikipedia.org/wiki/Bear)), [category](http://en.wikipedia.org/wiki/Lexical_category) features (it is a noun), [number](http://en.wikipedia.org/wiki/Grammatical_number) features (it is plural and must agree with verbs, pronouns, and demonstratives in its domain), [phonological](http://en.wikipedia.org/wiki/Phonology) features (it is pronounced a certain way), etc.

**Word boundaries**

The task of defining what constitutes a "word" involves determining where one word ends and another word begins—in other words, identifying word boundaries. There are several ways to determine where the word boundaries of spoken language should be placed:

* **Potential pause**: A speaker is told to repeat a given sentence slowly, allowing for pauses. The speaker will tend to insert pauses at the word boundaries. However, this method is not foolproof: the speaker could easily break up polysyllabic words, or fail to separate two or more closely related words.
* **Indivisibility**: A speaker is told to say a [sentence](http://en.wikipedia.org/wiki/Sentence_%28linguistics%29) out loud, and then is told to say the sentence again with extra words added to it. Thus, *I have lived in this village for ten years* might become *My family and I have lived in this little village for about ten or so years*. These extra words will tend to be added in the word boundaries of the original sentence. However, some languages have [infixes](http://en.wikipedia.org/wiki/Infix), which are put inside a word. Similarly, some have [separable affixes](http://en.wikipedia.org/wiki/Separable_affix); in the [German](http://en.wikipedia.org/wiki/German_language) sentence "Ich **komme** gut zu Hause **an**", the verb *ankommen* is separated.
* **Phonetic boundaries**: Some languages have particular rules of [pronunciation](http://en.wikipedia.org/wiki/Pronunciation) that make it easy to spot where a word boundary should be. For example, in a language that regularly [stresses](http://en.wikipedia.org/wiki/Lexical_stress) the last syllable of a word, a word boundary is likely to fall after each stressed syllable. Another example can be seen in a language that has [vowel harmony](http://en.wikipedia.org/wiki/Vowel_harmony) (like [Turkish](http://en.wikipedia.org/wiki/Turkish_language)):[[5]](http://en.wikipedia.org/wiki/Word#cite_note-4) the vowels within a given word share the same *quality*, so a word boundary is likely to occur whenever the vowel quality changes. Nevertheless, not all languages have such convenient phonetic rules, and even those that do present the occasional exceptions.
* **Orthographic boundaries**: See below.

**Orthography**

In languages with a [literary tradition](http://en.wikipedia.org/wiki/Writing), there is interrelation between [orthography](http://en.wikipedia.org/wiki/Orthography) and the question of what is considered a single word. [Word separators](http://en.wikipedia.org/wiki/Word_separator) (typically [spaces](http://en.wikipedia.org/wiki/Space_%28punctuation%29)) are common in modern orthography of languages using [alphabetic scripts](http://en.wikipedia.org/wiki/Alphabetic_script), but these are (excepting isolated precedents) a relatively modern development (see also [history of writing](http://en.wikipedia.org/wiki/History_of_writing)).

In [English orthography](http://en.wikipedia.org/wiki/English_orthography), [compound](http://en.wikipedia.org/wiki/Compound_%28linguistics%29) expressions may contain spaces. Examples are *ice cream*, *air raid shelter*, *get up*, and these must thus be considered as more than one word. (*Ice*, *cream*, *air* etc. indisputably exist as free forms, the case of *get* is less clear.) In contrast, *brownstone* is spelt as a single word and would thus be considered as such for most purposes even though *brown* and *stone* are free forms.

[Vietnamese](http://en.wikipedia.org/wiki/Vietnamese_language) orthography, although using the [Latin alphabet](http://en.wikipedia.org/wiki/Latin_alphabet), delimits monosyllabic morphemes, not words. East Asian orthography (languages using [CJK characters](http://en.wikipedia.org/wiki/CJK_characters)) also tend to delimit syllables (in the case of [Chinese characters](http://en.wikipedia.org/wiki/Chinese_characters)) or [morae](http://en.wikipedia.org/wiki/Mora_%28linguistics%29) (in the case of [kana](http://en.wikipedia.org/wiki/Kana)) rather than full words. [Hangul](http://en.wikipedia.org/wiki/Hangul) the [Korean](http://en.wikipedia.org/wiki/Korean_language) alphabet, delimits both syllables and words, by grouping graphemes into syllabic blocks but also adds spaces between words. Conversely, [synthetic languages](http://en.wikipedia.org/wiki/Synthetic_language) often combine many lexical morphemes into single words, making it difficult to boil them down to the traditional sense of words found more easily in [analytic languages](http://en.wikipedia.org/wiki/Analytic_language); this is especially difficult for [polysynthetic languages](http://en.wikipedia.org/wiki/Polysynthetic_language), such as [Inuktitut](http://en.wikipedia.org/wiki/Inuktitut) and [Ubykh](http://en.wikipedia.org/wiki/Ubykh_language), where entire sentences may consist of a single word.

**Morphology**

Main article: [Morphology (linguistics)](http://en.wikipedia.org/wiki/Morphology_%28linguistics%29)

Further information: [Inflection](http://en.wikipedia.org/wiki/Inflection)

In [synthetic languages](http://en.wikipedia.org/wiki/Synthetic_language), a single [word stem](http://en.wikipedia.org/wiki/Word_stem) (for example, *love*) may have a number of different forms (for example, *loves*, *loving*, and *loved*). However for some purposes these are not usually considered to be different words, but rather different forms of the same word. In these languages, words may be considered to be constructed from a number of [morphemes](http://en.wikipedia.org/wiki/Morpheme). In [Indo-European languages](http://en.wikipedia.org/wiki/Indo-European_languages) in particular, the morphemes distinguished are

* the [root](http://en.wikipedia.org/wiki/Root_%28linguistics%29)
* optional [suffixes](http://en.wikipedia.org/wiki/Suffixes)
* a [desinence](http://en.wikipedia.org/wiki/Desinence), or inflectional suffix.

Thus, the Proto-Indo-European *\*wr̥dhom* would be analyzed as consisting of

1. *\*wr̥-*, the [zero grade](http://en.wikipedia.org/wiki/Zero_grade) of the root *\*wer-*
2. a root-extension *\*-dh-* (diachronically a suffix), resulting in a complex root *\*wr̥dh-*
3. The [thematic suffix](http://en.wikipedia.org/wiki/Thematic_suffix) *\*-o-*
4. the [neuter gender](http://en.wikipedia.org/wiki/Neuter_gender) nominative or accusative singular desinence *\*-m*.

**Philosophy**

Philosophers have found words objects of fascination since at least the 5th century BC, with the foundation of the [philosophy of language](http://en.wikipedia.org/wiki/Philosophy_of_language). [Plato](http://en.wikipedia.org/wiki/Plato) analyzed words in terms of their origins and the sounds making them up, concluding that there was some connection between sound and meaning, though words change a great deal over time. [John Locke](http://en.wikipedia.org/wiki/John_Locke) wrote that the use of words "is to be sensible marks of ideas", though they are chosen "not by any natural connexion that there is between particular articulate sounds and certain ideas, for then there would be but one language amongst all men; but by a voluntary imposition, whereby such a word is made arbitrarily the mark of such an idea".[[6]](http://en.wikipedia.org/wiki/Word#cite_note-5) [Wittgenstein](http://en.wikipedia.org/wiki/Wittgenstein)'s thought transitioned from a word as representation of meaning to "the meaning of a word is its use in the language."[[7]](http://en.wikipedia.org/wiki/Word#cite_note-6)

**Classes**

Main article: [Lexical category](http://en.wikipedia.org/wiki/Lexical_category)

[Grammar](http://en.wikipedia.org/wiki/Grammar) classifies a language's lexicon into several groups of words. The basic bipartite division possible for virtually every [natural language](http://en.wikipedia.org/wiki/Natural_language) is that of [nouns](http://en.wikipedia.org/wiki/Noun) vs. [verbs](http://en.wikipedia.org/wiki/Verb).

The classification into such classes is in the tradition of [Dionysius Thrax](http://en.wikipedia.org/wiki/Dionysius_Thrax), who distinguished eight categories: [noun](http://en.wikipedia.org/wiki/Noun), [verb](http://en.wikipedia.org/wiki/Verb), [adjective](http://en.wikipedia.org/wiki/Adjective), [pronoun](http://en.wikipedia.org/wiki/Pronoun), [preposition](http://en.wikipedia.org/wiki/Preposition), [adverb](http://en.wikipedia.org/wiki/Adverb), [conjunction](http://en.wikipedia.org/wiki/Grammatical_conjunction) and [interjection](http://en.wikipedia.org/wiki/Interjection).

In Indian grammatical tradition, [Pāṇini](http://en.wikipedia.org/wiki/P%C4%81%E1%B9%87ini) introduced a similar fundamental classification into a nominal (nāma, suP) and a verbal (ākhyāta, tiN) class, based on the set of desinences taken by the word.

**See also**

* [Longest words](http://en.wikipedia.org/wiki/Longest_words)
* [Utterance](http://en.wikipedia.org/wiki/Utterance)

**Notes**

* 1. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-0) Katamba 11
  2. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-1) Fleming 77
  3. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-wierzbicka96_2-0) Wierzbicka 1996; Goddard 2002
  4. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-3) Adger (2003), pp. 36–7.
  5. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-4) Bauer 9
  6. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-5) ["Locke ECHU BOOK III Chapter II Of the Signification of Words"](http://www.rbjones.com/rbjpub/philos/classics/locke/ctb3c02.htm). Rbjones.com. Retrieved 13 March 2012.
  7. [**^**](http://en.wikipedia.org/wiki/Word#cite_ref-6) ["Ludwig Wittgenstein (Stanford Encyclopedia of Philosophy)"](http://plato.stanford.edu/entries/wittgenstein). Plato.stanford.edu. Retrieved 13 March 2012.

**References**

* Adger, David (2003). *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0199243700](http://en.wikipedia.org/wiki/Special:BookSources/0199243700).
* Barton, David (1994). *Literacy: An Introduction to the Ecology of Written Language*. Blackwell Publishing. p. 96.
* Bauer, Laurie (1983). *English Word-formation*. Cambridge: Cambridge University Press. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-521-28492-9](http://en.wikipedia.org/wiki/Special:BookSources/0-521-28492-9).
* Brown, Keith R. (Ed.) (2005) Encyclopedia of Language and Linguistics (2nd ed.). Elsevier. 14 vols.
* Crystal, David (1995). *The Cambridge Encyclopedia of the English Language* (1 ed.). Cambridge: Cambridge University Press. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-521-40179-8](http://en.wikipedia.org/wiki/Special:BookSources/0-521-40179-8).
* Fleming, Michael et al. (2001). *Meeting the Standards in Secondary English: A Guide to the ITT NC*. Routledge. p. 77. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0415233771](http://en.wikipedia.org/wiki/Special:BookSources/0415233771).
* Goddard, Cliff (2002). ["The search for the shared semantic core of all languages"](http://www.une.edu.au/lcl/nsm/pdf/Goddard_Ch1_2002.pdf). In Cliff Goddard and Anna Wierzbicka. *Meaning and Universal Grammar: Theory and Empirical Findings*. **Volume I**. Amsterdam: John Benjamins. pp. 5–40
* Katamba, Francis (2005). *English Words: Structure, History, Usage*. Routledge. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0415298938](http://en.wikipedia.org/wiki/Special:BookSources/0415298938).
* Plag, Ingo (2003). *Word-formation in English*. Cambridge: Cambridge University Press. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-521-52563-2](http://en.wikipedia.org/wiki/Special:BookSources/0-521-52563-2).
* Simpson, J.A. and E.S.C. Weiner, ed. (1989). *Oxford English Dictionary* (2 ed.). Clarendon Press. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-198-61186-2](http://en.wikipedia.org/wiki/Special:BookSources/0-198-61186-2).
* Wierzbicka, Anna (1996). *Semantics: Primes and Universals*. Oxford University Press. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [0198700024](http://en.wikipedia.org/wiki/Special:BookSources/0198700024).