

# Morpheme

Carmen Saldana<sup>a</sup> and Borja Herce<sup>b</sup>

<sup>a</sup>Universitat Pompeu Fabra, Department of Translation and Language Sciences, Roc Boronat 138, ES-08018 Barcelona

<sup>b</sup>University of Zurich, Department of Comparative Language Science, Affolternstrasse 56, CH-8050 Zurich

This chapter will be published in Nesi, H. & Milin, P. (Eds.) *International Encyclopedia of Language and Linguistics, 3rd Edition*. The abstract is already available online within the *Reference Collection in Social Science*, <https://doi.org/10.1016/B978-0-323-95504-1.00027-2>. © 2024 Elsevier.

## Key points

- Introduce the term and concept *morpheme*.
- Discuss the evolution of the concept of the morpheme through different theoretical models, with a focus on inflectional morphology.
- Introduce the historical and current debates surrounding the morpheme.
- Raise awareness about the complexity of the morphological component in language, and its interface with syntax, semantics and phonology.

## Glossary

**affix** a formative that is combined with stems, words or phrases to derive or inflect forms. The most common types of affixes are prefixes, suffixes and infixes.

**agglutination** process in which words are formed by concatenating exponents, each of which corresponds to a single morphosyntactic value.

**allomorph** any of two or more formatives which manifest the same morphosyntactic value.

**exponent** the degree to which different morphosyntactic categories are grouped together in single, indivisible formatives.

**exponent** a phonological manifestation of a morphosyntactic value.

**formative** an exponent of inflectional information. Formatives comprise not only affixes, phrasal affixes and particles but also non-segmental processes.

**fusion** the degree to which formatives are phonologically intertwined with their host.

**grammar** a language's set of rules for the correct arrangement or deployment of vocabulary items to form words, phrases and sentences.

**inflection** a process whereby a word is changed to express a specific morphosyntactic value. It is also used to refer to formatives wherewith the word is changed.

**lexeme** a unit of lexical meaning that underlies a set of words that are related through inflection.

**lexicon** a language's inventory of vocabulary items. It is frequently opposed to a language's grammar.

**root** indivisible part of a word, with lexical content, that does not contain affixes.

**segmental morphology** morphology which consists of the addition of stable segmental material in a concatenative manner. Morphology which does not conform to this ideal is referred to as non-segmental or non-concatenative.

**stem** the root, together with any derivational affixes, to which inflectional affixes are added.

**synthesis** the degree to which various formatives and lexical roots are bound together in unsegmentable ways.

## Abstract

The term *morpheme* is often used to refer to minimal structural units of language. It is one of the most widely used in the language sciences; however, as it is often the case, its definition is also one of the most disputed. The term has come to acquire various related technical senses in different theoretical models—some focus on the formal aspects of such minimal structural units, others on the meaning, and others on the form-meaning mapping—which are sometimes mutually incompatible, and frequently make it challenging to understand one another. In this chapter we will discuss the history of the concept of the morpheme and its role in morphological theory.

**Keywords:** morpheme; morphology; inflectional morphology; morphosyntax; morphosemantics; linguistic theory.

## 1 Introduction

The term *morpheme* is one of the most widely used in the language sciences; however, its definition is also one of the most disputed (Haspelmath, 2020, 2021). It has come to acquire various related technical senses in different theoretical models that are sometimes mutually incompatible and frequently make it challenging to understand one another. In this chapter we will discuss the history of the concept of *morpheme* and its role in morphological theory, with a focus on inflectional morphology.

## 2 Defining the morpheme

Historically, the term *morpheme* was coined by Baudouin de Courtenay towards the end of the 19<sup>th</sup> century to refer to the smallest structural unit of language (Baudouin de Courtenay, 1895); in particular, “that part of a word which is endowed with psychological autonomy and is for the very same reasons not further divisible” (Baudouin de Courtenay, 1972). By contrast to his definition of *phoneme*, we can understand that a morpheme is “psychologically autonomous” because it belongs to “the sphere of morphology which exists in the mind”, however, the nature of the mental representation is unspecified. There is no doubt, however, on what formally constitutes a morpheme in his view: *roots*, all types of *affixes* (e.g., prefixes and suffixes), and any type of exponent of syntactic relationships.

Unlike Baudouin's phoneme, however, the term *morpheme* only became popular with Bloomfield's work in the 20's in the 30's

(Bloomfield, 1933). Bloomfield defined it as “a recurrent (meaningful) form which cannot in turn be analyzed into smaller recurrent (meaningful) forms” (Bloomfield, 1926, p. 155), and importantly, this form should bear “no partial phonetic-semantic resemblance to any other form” (Bloomfield, 1933, p.161); in this definition he makes sure to classify as different morphemes those phonetically similar units which do not share any semantic resemblance (e.g., homonyms). That is, he mentioned explicitly a necessary semantic (i.e. meaning) component in the definition of this minimal morphological unit.<sup>1</sup>

In Bloomfield’s own use of the term, however, only formal constituents (but not necessarily the meaning constituents) of complex words could constitute morphemes. The use of morpheme in practice would thus be restricted to recurring minimal meaning-bearing forms *which can be identified by segmentation* (see also Haspelmath & Sims, 2013). On the one hand, *cat-s* and *remember-ed* were considered grammatically complex words composed of more than one morpheme. On the other, grammatically complex words such as *mice* or *sang* were analyzed as single morphemes because they are not formally segmentable in further recurrent meaningful units—even though their meaning is complex as they are the plural (PL) of *mouse* and the past tense (PST) of *sing* respectively. Bloomfield (1933) nevertheless describes them as *secondary morpheme-words*, in particular, as single morphemes with an additional *substitution feature*. This substitution feature is an alternative to their segmentable counterparts and they are equally meaningful, but are not considered morphemes (Bloomfield, 1933). Secondary morpheme-words can also be accompanied by a *zero-feature* (e.g., in the plural of sheep *sheep*), *suppletion* (e.g., the past tense of the verb ‘to go’ *went*), or a *minus-feature* (e.g., in the past tense of ‘to make’ where we find [meɪd] instead of \*[meɪkt] and thus the phonological segment [k] is lost). They are all considered single morphemes.

Bloomfield’s original definition makes a clear distinction between morphemes and non-morphemic processes (i.e., segmental vs non-segmental), and between meaningful and non-meaningful processes (e.g., morphologically conditioned vs phonologically conditioned alternations). In the reductionist spirit of the years following Bloomfield (1933), the impulse was to reevaluate those distinctions (e.g., Harris, 1942; Hockett, 1968). The focus was shifted to account for all aspects of the meaning as well as the form of complex words through both the forms and meanings of their component parts. Linguists started to analyze non-segmental processes such the alternation between *mouse* and *mice* as morphemes as well (e.g. Harris, 1942; Hockett, 1968; Jakobson, 1939). Therefore, all words involving morphological processes such as substitution, zero-marking or suppletion would no longer be defined as single morphemes. Along with this type of reanalysis, the nature of the morpheme as a meaning-bearing segmental unit altogether was questioned. Morphemes came to be understood as more abstract units on the grammatical level—just as phonemes are on the phonological level (Hockett, 1987). Akin to the relationship between the abstract unit phoneme and its various phonetic instantiations (allophones), the different realizations of the same meaning in morphology were considered morphological variants (allomorphs)

<sup>1</sup>It is nonetheless worth noting that the concept of meaning in his definition remained largely unspecified—even forms with no straightforward semantic association or function could also be designated as morphemes (e.g., *-end-* in Latin verbs like *prendere, pendere, rendere, attendere*, etc.; Bloomfield, 1926, p. 163).

of the same morpheme. For example, the more abstract unit [pl] would be always considered a morpheme regardless of whether it was realized as a segmental unit or not; all different manifestations of plural [pl] would be considered allomorphs— e.g., any phonological realization of the plural suffix /s/ would be considered an allomorph, and so would alternations such as /au/ ⇒ /ai/ in *mouse* (MOUSE.SG) ⇒ *mice* (MOUSE.PL). Due to this shift in focus, within the Structuralist tradition, the phonological variants started to be designated as *morphs*, and the abstract units (e.g., [PL] or [PST]), as morphemes.<sup>2</sup>

With these developments, meaningful form and meaningful process would fall under the same umbrella of ‘morpheme’, shifting the focus from form to meaning, not only in the term, but also in morphology more generally. The fundamental opposition between the morphology (as a lexicon) and grammar that Bloomfield had carefully delineated started to blur. Within the generative tradition, a core property of theoretical frameworks such as *Distributed Morphology* (DM) developed in the 90’s is that syntax and morphology enter into the same types of constituent structures, and thus that morphology, at the end of the day, is just “syntax all the way down” (Halle & Marantz, 1993). DM defines morphemes in terms of syntax alone as the terminal nodes of syntactic trees regardless of whether they are phonologically realized or not. These terminal nodes are bundles of grammatical features, which can contain lexical (i.e., roots) and functional morphemes (i.e., morphosyntactic features, e.g., [PL])—corresponding approximately to the conventional division between functional and lexical categories but at a sub-word level. By redefining the morpheme in terms of meaning alone, DM deals with the problem of the coexistence of forms and processes as different realizations of the same morphosyntactic categories.

This shift from form to meaning allows for the comparable discussion of morpheme-based morphology across languages with more or less morphosyntactic categories expressed per unit (i.e., different degree of synthesis and exponence), and with more and less concatenative morphology (i.e., different degree of fusion). There are languages with chiefly cumulative exponence, where, for example, case and number are expressed within a single indivisible formative (e.g., Latin or Polish); the morphemic complexity of a word with such cumulative exponence would be comparable to that of a word with two separable formatives, one for case and another for number (e.g., as we find in Turkish or Hungarian). There are also languages which chiefly inflect words via non-segmental processes (e.g., Semitic languages such as Arabic or Hebrew) or which use tone to mark morphosyntactic distinctions (e.g., as many Bantu languages do; for further examples, see Bickel & Nichols, 2007); these non-concatenative units and processes would be treated equally with regards to morphological composition as prototypical morphemes in concatenative structures.

<sup>2</sup>Note that alternative uses of the term ‘morph’ also exist in more recent theory. For example, (Haspelmath, 2020) has recently advocated for using the term ‘morph’ instead to refer to the original sense of ‘morpheme’ (i.e. a meaning-bearing indivisible unit), and not only as a phonological form as suggested by earlier structuralists. Close to the Bloomfieldian use of ‘morpheme’, Haspelmath (2020) also advocates for a strict and tractable use of the term to refer to continuous and segmental forms with meaning exclusively, thus excluding processes, meaningful elements, or zeros.

### 3 Defying the morpheme

The Baudouin and Bloomfield models, later structuralist models, and generative models such as DM, even if with different proposals about the nature of morphemes and the relation between morphology and syntax, they all shared a morpheme-based approach: Words and phrases are defined by morphemic composition. These models would fall within what Hockett (1954) had coined *Item-and-Arrangement* (IA) models, and they made the morpheme as meaning-bearing unit (segmental or not) a central tenet to morphology and morphosyntax. In favor of such morpheme-centric approaches to morphology, there exists a vast body of quantitative and experimental research supporting a facilitatory role of morpheme-based morphology in language learning and transmission (e.g., Saldana et al., 2019; Mansfield et al., 2022), and suggesting the existence of typological and learning biases towards morpheme-based paradigmatic and syntagmatic relations (e.g., Mansfield et al., 2020; Saldana et al., 2021; Baerman et al., 2005; Bybee, 1985).

However, while theories such as Distributed Morphology dealt with the obstacle of the meaning without form—which had led to Bloomfield to define morphemes solely segmentally—alternative frameworks focused on the case of form without meaning (e.g., recurrent patterns without apparent semantic or functional association such the aforementioned *-end-* in Latin verbs, or empty morphs such as thematic vowels), and started to question the theoretical usefulness of the morpheme altogether. If one can have forms with no meaning, and meanings with no forms, one no longer needs meaning-bearing forms as elementary particles in morphology (Hockett, 1987; Anderson, 1992; Blevins, 2016). Several models were developed to shift the focus of morphology from units to processes and intend to do away with the morpheme. These models would fall within what Hockett (1954) had referred to as *Item-and-Process* (IP) models (Matthews, 1972; Anderson, 1992; Stump, 2001; Aronoff, 1994). Anderson (1992)'s A-morphous Morphology (AM) proposes that morphology is ultimately not about morphemes. Within AM, word structure is not morphemic composition but a system of relations governed by word formation rules, which (recursively) derive complex words (see also Aronoff, 1994). Anderson's model minimizes the non-phonological internal structure assigned to words and eliminates morphologically motivated boundary elements altogether. Moreover, properties and relations of lexical items are posited as independent from syntax and invisible to it, thus defining morphology as an autonomous component.

Stemming from the same concern regarding the usefulness of the morpheme, the Word and Paradigm (WP) models (e.g., Blevins, 2016; Hockett, 1987; Matthews, 1972) also propose to do away with morphemes as form-meaning mappings, and it is in many ways comparable to an Item-and-Process models such as AM (Bauer, 2001). However, while IP models are exclusively concerned with the composition and decomposition of individual word forms in isolation—just as IA models but from processes-based approach—and the relations between words fall outside their descriptive scope, WP models focus on such relations. Within the WP framework, only whole word forms (i.e., associated to their respective lexemes) have meanings, and smaller morphological elements just serve to generate contrasts between these whole words. Word and Paradigm morphology adopts a 'complex system' perspective (Blevins et al., 2016), which in the last decades has become popular in other aspects of the language sciences (e.g., Beckner et al.,

2009), according to which the properties of a system often cannot be fully captured successfully from the properties of its individual parts.

Amidst the dissection of the utility of the morpheme in morphology, Aronoff (1994) introduced the new term *morphome* to identify morphological units whose identity is established on purely morphological grounds, independent of phonology, semantics or syntax; the notion addresses the abstract and unmotivated (i.e., not functional) structure of the lexicon. A pattern or distribution which appears systematic—as it recurrently appears across various formatives—but which does not correspond to any conceivable morphosyntactic value is thus called morphomic. Examples of morphomes are thematic vowels in Latin or the plentiful patterns of stem alternations in Romance languages which extend over different word forms in verbal paradigms without any apparent morphosyntactic feature in common (Aronoff, 1994; Maiden, 2018).

Morphomes are not rare cross-linguistically and seem to be robustly transmitted (Maiden, 2011; Herce, 2023); their existence is thus often used to argue for an autonomous morphological component in language—as so did the aforementioned IP and WP models. However, more recent research suggests that the typological evidence does not support a strict dichotomy between morpheme and morphome, and question the motivation and usefulness of the distinction as a whole. Herce (2020) claims that, exploring the diachrony of morphological systems, no single property can be identified that consistently distinguishes morphomes from morphemes: Morphomes can have the same sources as morphemes and the same diachronic stability. Recent quantitative and experimental evidence (e.g., Saldana et al., 2022; Herce et al., 2023; Herce & Allasonnière-Tang, 2024) suggests that seemingly unmotivated morphomic patterns might not necessarily be so, and that the morpheme-morphome distinction might be gradient rather than a dichotomous one. Saldana et al. (2022) and Herce et al. (2023) show a gradient in the cross-linguistic recurrence and learnability—through artificial language learning experiments—across different types of morphomic patterns, from more to less morphemic-like. They propose that this gradient in cross-linguistic recurrence and learnability reflects a general bias towards patterns with the same identity to share higher semantic similarity, even if morphomic. This work thus questions the need to posit an autonomous morphological component due to the existence of morphomic and not only prototypically morphemic patterns.

### 4 Deploying the morpheme: Concluding remarks

The term and associated concept of the 'morpheme' has undergone continuous definitions and redefinitions across its long history. This is the product of active debate in the language sciences and indicates the central importance and usefulness of the concept to describe and analyze morphological structure in diverse languages. However, the proliferation of senses and technical terms in our discipline make it necessary for linguists to carefully explain their choices.

Smaller than word-level units of form and/or meaning need to be talked about, and therefore need terms for doing so. The ideal case would be to have separate and unambiguous terms to refer to a minimal unit of meaning, to a minimal unit of form, and to the cross-linguistically common configuration where a minimal unit of

form and meaning coincide. The last—e.g., *cat-s* (cat.PL) or *sister-s* (female-sibling.PL)—is often what is referred to as ‘morpheme’. While the usefulness of this (and other) notion(s) will vary depending on the theoretical focus and the typology of the target language (e.g., most useful in concatenative structures), history suggests that the ‘morpheme’ will retain a prominent role in formal and informal descriptions of morphological structure.

## Acknowledgments

CS is supported by the Beatriu de Pinós programme funded by the Generalitat de Catalunya (agreement No. 2022 BP 00101) and the European Union’s Horizon 2020 research and innovation programme under the MSCA grant agreement No 801370.

The outline of this chapter is adapted and expanded from Aronoff & Volpe (2006)’s chapter on this encyclopedia’s previous edition.

## References

- Anderson, S. R. (1992). *A-morphous morphology*. Cambridge, UK: Cambridge University Press.
- Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes* (Vol. 22). Cambridge, MA: MIT press.
- Aronoff, M., & Volpe, M. (2006). Morpheme. In K. Brown (Ed.), *Encyclopedia of language & linguistics (second edition)* (Second Edition ed., p. 274–276). Amsterdam, NL: Elsevier. doi: doi:10.1016/B0-08-044854-2/04245-0
- Baerman, M., Brown, D., & Corbett, G. G. (2005). *The syntax-morphology interface: A study of syncretism* (Vol. 109). Cambridge, UK: Cambridge University Press.
- Baudouin de Courtenay, J. (1895). *Versuch einer theorie phonetischer alternationen*. Strassburg: Commissionsverlag von Karl J. Trübner.
- Baudouin de Courtenay, J. (1972). *The beginnings of structural linguistics: A baudouin de courtenay anthology*. Bloomington, IN: Indiana University Press.
- Bauer, L. (2001). *Morphological productivity* (Vol. 95). Cambridge, UK: Cambridge University Press.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... others (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59, 1–26. doi: doi:10.1111/j.1467-9922.2009.00533.x
- Bickel, B., & Nichols, J. (2007). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description*. Cambridge, UK: Cambridge University Press. (2nd edition)
- Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford, UK: Oxford University Press.
- Blevins, J. P., Ackerman, F., Malouf, R., & Ramscar, M. (2016). Morphology as an adaptive discriminative system. *Morphological metatheory*, 271–302. doi: doi:10.1075/la.229.10ble
- Bloomfield, L. (1926). A set of postulates for the science of language. *Language*, 2(3), 153–164. doi: doi:10.2307/408741
- Bloomfield, L. (1933). *Language*. New York, NY: Henry Holt and Company.
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia, PA: John Benjamins.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20: Essays in linguistics in honour of sylvain bromberger* (pp. 111–176). Cambridge, MA: MIT Press.
- Harris, Z. S. (1942). Morpheme alternants in linguistic analysis. (3), 169–180. doi: doi:10.2307/409550
- Haspelmath, M. (2020). The morph as a minimal linguistic form. *Morphology*, 30(2), 117–134. doi: doi:10.1007/s11525-020-09355-5
- Haspelmath, M. (2021). Towards standardization of morphosyntactic terminology for general linguistics. *Linguistic categories, language description and linguistic typology*, 35–57. doi: doi:10.1075/tsl.132.02has
- Haspelmath, M., & Sims, A. (2013). *Understanding morphology*. London, UK: Routledge.
- Herce, B. (2020). On morphemes and morphemes: Exploring the distinction. *Word Structure*, 13(1). doi: doi:10.3366/word.2020.0159
- Herce, B. (2023). *The typological diversity of morphemes: A cross-linguistic study of unnatural morphology*. Oxford, UK: Oxford University Press.
- Herce, B., & Allasonnière-Tang, M. (2024). The meaning of morphemes: distributional semantics of Spanish stem alternations. *Linguistics Vanguard*(0). doi: doi:10.1515/lingvan-2023-0010
- Herce, B., Saldana, C., Mansfield, J., & Bickel, B. (2023). Naturalness is gradient in morphological paradigms: Evidence from positional splits. *Glossa*, 8(1), online. doi: doi:10.16995/glossa.9280
- Hockett, C. F. (1954). Two models of grammatical description. *Word*, 10(2-3), 210–234. doi: doi:10.1080/00437956.1954.11659524
- Hockett, C. F. (1968). *The state of the art* (Vol. 73). Berlin, DE: Walter de Gruyter.
- Hockett, C. F. (1987). *Refurbishing our foundations: Elementary linguistics from an advanced point of view*. Amsterdam, NL: John Benjamins.
- Jakobson, R. (1939). Signe zéro. In *Mélanges de linguistique offerts à Charles Bally* (pp. 143–152). Georg.
- Maiden, M. (2011). Morphophonological persistence. In *The Cambridge history of the romance languages* (pp. 155–215). Cambridge, UK: Cambridge University Press.
- Maiden, M. (2018). *The romance verb: Morphomic structure and diachrony*. Oxford, UK: Oxford University Press.
- Mansfield, J., Saldana, C., Hurst, P., Nordlinger, R., Stoll, S., Bickel, B., & Perfors, A. (2022). Category clustering and morphological learning. *Cognitive Science*, 46(2), e13107. doi: doi:10.1111/cogs.13107
- Mansfield, J., Stoll, S., & Bickel, B. (2020). Category clustering: A probabilistic bias in the morphology of verbal agreement marking. *Language*, 96(2), 255–293. doi: doi:10.1353/lan.2020.0021
- Matthews, P. H. (1972). *Inflectional morphology: A theoretical study based on aspects of latin verb conjugation* (Vol. 6). Cambridge, UK: Cambridge University Press.
- Saldana, C., Herce, B., & Bickel, B. (2022). More or less unnatural: Semantic similarity shapes the learning and cross-linguistic distributions of morphological paradigms. *Open Mind: Discoveries in Cognitive Science*. doi: doi:10.1162/opmi.a.00062
- Saldana, C., Kirby, S., Truswell, R., & Smith, K. (2019). Compositional hierarchical structure evolves through cultural transmission: an experimental study. *Journal of Language Evolution*, 4(2), 83–107. doi: doi:10.1093/jole/lzz002
- Saldana, C., Oseki, Y., & Culbertson, J. (2021). Cross-linguistic patterns of morpheme order reflect cognitive biases: An experimental study of case and number morphology. *Journal of Memory and Language*, 118, 104204. doi: doi:10.1016/j.jml.2020.104204
- Stump, G. T. (2001). *Inflectional morphology: A theory of paradigm structure* (Vol. 93). Cambridge, UK: Cambridge University Press.