<u>Lexemes</u>

Marios Andreou

Heinrich Heine University, Düsseldorf, Germany

Introduction

A lexeme is a theoretical construct that stands for the unitary meaning and shared syntactic properties of a group of word forms. A lexeme is stripped of any inflectional endings. Thus *play*, *plays*, *played*, and *playing* are all inflected forms of the lexeme *play*. In a similar vein, *cat* and *cats* are inflected forms of the same lexeme, i.e. *cat*. Although inflection creates forms of the same lexeme, derivation creates new lexemes. Thus, *player* is not a form of the lexeme *play* but is considered a different lexeme; it has its own meaning and lexical category; it is a noun. With respect to their morphology, lexemes can be either simple or complex. For example, *door* is simple and *lemon-tree* is complex. With respect to their semantics, lexemes can be monosemous or polysemous. For instance, names for days such as *Saturday* have only one meaning and are monosemous, whereas a lexeme such as *university* is polysemous; it has more than one meaning. Polysemy is often contrasted with homonymy, under which the same form is associated with two or more unrelated meanings. Polysemous senses are attributed to the same lexeme, homonyms are considered as different lexemes. Although it is usually possible to identify a phonological form that is basic to all the forms of a lexeme, this is not always the case. For example, the form *kiss* is shared by all forms of the lexeme *kiss* (e.g. *kiss*, *kisses*, *kissed*, *kissing*). There is no common phonological form, however, between all forms of the lexeme *go* (e.g. *go*, *went*). Lexemes and their citation form should be kept distinct since the way a lexeme is cited is merely a convention and does not bear on any crucial phonological

property of that lexeme. In certain traditions, as for example in Greek, the first person singular is used as a citation form for verbs, whereas in other traditions, as for example in French, the infinitive is used. For nouns, the nominative singular is used. The terms lexeme, lexical unit, lexical item, word, and lemma are often used interchangeably in the relevant literature and in different linguistic fields.

Textbooks

Since Matthews 1965 and Lyons 1968, which offer the first treatments of the term lexeme in modern linguistic theory, the lexeme has become a fundamental unit in morphological and semantic analysis. As such, textbooks on morphology and semantics often address issues that relate to its definition and properties. Textbooks on morphology include Spencer 1991, Carstairs-McCarthy 1992, Plag 2003, Lieber 2010, and Aronoff and Fudeman 2011. Cruse 2000 and Löbner 2013 are introductions to semantics.

Aronoff, Mark, and Kirsten Fudeman. 2011. *What is morphology?* 2nd ed. Malden, MA: Wiley-Blackwell.
> An introduction to morphology that includes a very useful discussion of the term lexeme on pp. 43-46.

Carstairs-McCarthy, Andrew. 1992. *Current morphology*. New York: Routledge.
> A comprehensive survey that presents a number of approaches to morphology. It is very useful since it analyses each approach critically. On the lexeme, pp. 182-187.

Cruse, Alan, D. 2000. *Meaning in language: an introduction to semantics and pragmatics*. Oxford: Oxford University Press.
> An introduction to both semantics and pragmatics. Best for undergraduate students. On the lexeme, pp. 88-89.

Lieber, Rochelle. 2010. *Introducing morphology*. Cambridge, UK: Cambridge University Press.
> An introduction to morphology that does not require previous knowledge of the field. On the lexeme, pp. 4-5.

Löbner, Sebastian. 2013. *Understanding Semantics*. 2nd ed. New York, London: Routledge.
> An introduction to semantics. Presents and critically comments on a number of theoretical approaches to semantics. On the lexeme, pp. 40-42.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
> One of the first works on the term lexeme. The author defines the term lexeme as an abstract unit which occurs in different inflectional forms.

Matthews, Peter H. 1965. "The inflectional component of a word-and-paradigm grammar." *Journal of Linguistics* 1: 139-171.

> One of the first works on the term lexeme. The author defines the term word by its assignment to a specific vocabulary element (i.e. lexeme) and by the assignment to it of a specific set of morphosyntactic properties.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

> An introduction to morphology that targets both undergraduate and advanced students. On the lexeme, p. 9, p. 14, pp. 18–19, p. 28, p. 103, p. 117, p. 128, p. 193, and p. 194.

Spencer, Andrew. 1991. *Morphological theory: an introduction to word structure in generative grammar*. Oxford: Blackwell.

> An introduction to morphology that aims at showing how morphological theories have been developed, criticized, and revised. Best introduction for advanced students. On the lexeme, pp. 45-49.

Lexemes and Words

The term word is usually associated with three distinct uses, that is, phonological word, grammatical word, and lexeme. The phonological word is a string of sounds that serves as a unit for the phonology. The grammatical word stands for the different forms of a single word that depend on the syntactic context. The lexeme is an abstract unit that generalizes over inflected forms. Problems with the definition of the term word are presented in Dixon and Aikhenvald 2002. Matthews 1991 discusses words, word forms, and lexemes. See also Spencer 1991, Carstairs-McCarthy 1992, Plag 2003, Lieber 2010, Aronoff and Fudeman 2011, and Cruse 2000 cited under *Textbooks*. In the literature, it is often the case that lexeme and word are used interchangeably. Bally 1950 discusses the ambiguity of the term 'word'. For the way several word-forms are correlated with a given lexeme see Ackerman and Stump 2004. As far as the grammatical dimension is concerned, the distinction between the terms 'lexeme' and 'grammeme' is also relevant, as in Mel'čuk 1993 and Fradin 2003.

Ackerman, Farrell, and Gregory Stump. 2004. "Paradigms and periphrastic expression: a study in realization-based lexicalism." In *Projecting morphology*, edited by Andrew Spencer and Louisa Sadler, 111-57. Stanford: CSLI Publications.

> Develops an inferential-realizational conception of periphrasis and shows that rules of morphology define the realization of a lexeme's pairing with a particular set of morphosyntactic properties.

Bally, Charles. 1950. *Linguistique générale et linguistique française*. 3rd ed. Berne: Francke.

> Discusses the ambiguity of the term word (French *mot*) on pp. 287-289.

Dixon, Robert M. W., and Alexandra Aikhenvald Y., eds. 2002. *Word: a crosslinguistic typology*. Cambridge, UK: Cambridge University Press.

> On words and lexemes, pp. 6-7.

Fradin, Bernard. 2003. *Nouvelles approches en morphologie.* Paris: Presses Universitaires de France.

> Provides a definition of grammeme and discusses several variants of this notion pp. 92-95.

Matthews, Peter H. 1991. *Morphology*. 2d ed. Cambridge, UK: Cambridge University Press.

> On words and lexemes, pp. 24-41.

Mel'čuk Igor A. 1993. *Cours de morphologie générale. Première partie: le mot. 5 vols. Vol. 1.* Montréal: Presses de l'Université de Montréal - CNRS Editions. pp. 264-265

> Mel'čuk's provides a definition of grammeme as the morphosyntactic content expressed by inflected lexical units and contrasts it with the semantic content associated with morphologically derived lexical units.

Lexemes and Morphology

There are two main types of morphological frameworks the relevant criterion being whether they use the lexeme or the morpheme as their basic unit. Morpheme-based approaches are based on the isomorphism between sound and meaning, whereas lexeme-based approaches are usually separationist, in that they do not presuppose there is a direct link between the phonology, the syntax, and the semantics of an item. Lexeme-based approaches are found in Aronoff 1994, Beard 1995, Stump 2001, Blevins 2006, and Booij 2010. In Beard 1995, the distinction between lexemes and grammatical morphemes is crucial. Bonami and Crysmman 2016 offers an overview of lexeme formation in Lexical Functional Grammar and Head-driven Phrase Structure Grammar. Andreou 2017 offers a treatment of lexical rules in Frame Semantics.

Andreou, Marios. 2017. "Stereotype negation in frame semantics." *Glossa: a journal of general linguistics* 2, no. 1: 79. https://doi.org/10.5334/gjgl293.

> Introduces lexical rules into the framework of Frame Semantics and treats stereotype negation as a change in the value of an attribute of the base lexeme.

Aronoff, Mark. 1994. *Morphology by itself: stems and inflectional classes*. Cambridge, MA: MIT Press.

> A seminal work on lexeme-based morphology.

Beard, Robert. 1995. *Lexeme-morpheme base morphology*. Albany: State Univ. of New York Press.

> Proposes a separation between lexemes (i.e. major lexical categories such as noun, verb, and adjective) and affixes.

Blevins, James P. 2006. "Word-based morphology." *Journal of Linguistics* 42: 531–573.

    Develops a morphological theory which takes words to be the basic units of analysis.

Bonami, Olivier, and Crysmann Berthold. 2016. "Morphology in constraint-based lexical approaches to grammar." In *Cambridge Handbook of Morphology*, edited by Andrew Hippisley and Gregory T. Stump, 609-656. Cambridge: Cambridge University Press.

    An overview of lexeme formation rules and inheritance hierarchies in constraint-based lexical approaches.

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford Univ. Press.

    A morphological theory within the framework of construction grammar. It considers the lexeme as the basic unit of analysis.

Stump, Gregory T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge, UK: Cambridge Univ. Press.

    Develops a formal framework and treats inflectional morphology as an autonomous system of principles. Argues for the crucial role of paradigms in word formation.

Inflection, Derivation, Compounding

A word form of a lexeme that carries inflection is an item which realizes a set of morphosyntactic properties that may be relevant to syntax ("contextual inflection", Booij 1996) or not ("inherent inflection"). Given that lexemes are stripped of any inflectional endings, the use of the notion of a lexeme presupposes a distinction between inflection and derivation as noted in Spencer 2000. Beard 1988, Perlmutter 1988, and Anderson 1992 are proponents of the idea that inflectional rules belong to a post-syntactic component of spell-out rules. Booij 1996 offers evidence against this view. Bybee 1985 treats inflection and derivation as a continuum. As far as compounding is concerned, compounds are usually considered as new lexemes that consist of two or more lexemes. Problems with respect to defining compounding in terms of lexemes are discussed in Lieber and Štekauer 2009. The creation of complex lexemes via the lexicalization of syntactic phrases is discussed in Fradin 2009.

Anderson, Stephen R. 1992. *A-morphous morphology*. Cambridge, UK: Cambridge Univ. Press.

    Makes a case for the separation of inflection and derivation. Inflectional morphology is what is relevant to the syntax.

Beard, Robert. 1988. "On the separation of derivation from morphology: toward a lexeme-morpheme based morphology." *Quarderni di semantica* 9: 3-59.

    Makes a case for the separation of derivation from morphology.

Booij, Geert. 1996. "Inherent versus contextual inflection and the split morphology hypothesis." In *Yearbook of Morphology 1995*, edited by Geert Booij and Jaap van Marle, 1-16. Dordrecht: Kluwer.

    Introduces the distinction between inherent (e.g. plural form of nouns) and contextual inflection (e.g. case forms of nouns). Only inherent inflection can feed word formation.

Bybee, Joan. 1985. *Morphology: a study of the relation between meaning and form*. Amsterdam/Philadelphia: Benjamins.

    An account under which inflection and derivation form a continuum.

Fradin, Bernard. 2009. "IE, Romance: French." In *Oxford handbook of compounding*, edited by Rochelle Lieber and Pavol Štekauer, 417-35. Oxford: Oxford University Press.

    Discusses the creation of complex lexemes via morphological processes such as compounding, and via the lexicalization of syntactic phrases.

Perlmutter, David. 1988. "The split morphology hypothesis: Evidence from Yiddish." In *Theoretical morphology*, edited by Michael Hammond and Michael Noonan, 79-100. San Diego: Academic Press.

    An account under which derivation belongs to a pre-syntactic component of grammar and inflection is accounted for by a post-syntactic component of grammar.

Lieber, Rochelle and Pavol Štekauer. 2009. "Introduction: status and definition of compounding." In the *Oxford handbook of compounding*, edited by Rochelle Lieber and Pavol Štekauer, 3-18. Oxford: Oxford University Press.

    The authors discuss whether compounds are new lexemes and how one can identify a new lexeme.

Spencer, Andrew. 2000. "Inflection and the lexeme." *Acta Linguistica Hungarica* 47, no. 1-4: 335-344.

    Argues that a piece of morphology is inflectional if and only if it corresponds to a syntactic feature, and derivational if it is given direct semantic interpretation at the level of the lexeme.

Lexemes and Paradigms

The term lexeme is closely related to the notion of a paradigm, that is, a set of cells in which each cell is the pairing of an inflected form of the lexeme with a complete set of morphosyntactic properties for that lexeme. As shown in Stump 2016a and Spencer 2016, the notion of a paradigm is crucial for the analysis of a number of phenomena that manifest themselves at the morphology-syntax interface. Stump 2016b discusses heteroclisis, Baerman, et al. 2005 tackles syncretism, and Mel'čuk 1994 deals with suppletion. Ackerman, et al. 2009, Ackerman and Malouf 2013, and Sims 2015 examine the issue of predicting the form of one inflected word in the paradigm of a lexeme based on another form of the same lexeme.

Ackerman, Farrell and Robert Malouf. 2013. "Morphological organization: The low entropy conjecture." *Language* 89: 429–464. https://doi.org/10.1353/lan.2013.0054.
    Offers evidence for the central role of paradigms in morphological organization.

Ackerman, Farrell, James P. Blevins, and Robert Malouft. 2009. "Parts and wholes: Implicative patterns in inflectional paradigms." In *Analogy in Grammar: Form and Acquisition*, edited by James P. Blevins and Juliette Blevins, 54–82. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199547548.003.0003.
    The authors use conditional entropy to calculate the predictive value of particular words in relation to others.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2005. *The syntax-morphology interface: a study of syncretism*. Cambridge: Cambridge University Press.
    Includes a discussion of syncretism in a wide range of languages.

Mel'čuk, Igor. 1994. "Suppletion: Toward a logical analysis of the concept." *Studies in Language* 18, 2: 339–410.
    Discusses a number of types of suppletion (e.g. derivational vs. inflectional suppletion). Also examines suppletion from a historical perspective.

Sims, Andrea D. 2015. *Inflectional defectiveness*. Cambridge: Cambridge University Press.
    Examines gaps in the paradigm of lexemes.

Spencer, Andrew. 2016. "How are words related?" In *Morphological Metatheory*, edited by Daniel Siddiqi and Heidi Harley, 1–26. Amsterdam: John Benjamins.
    Argues in favor of the notions word, lexeme and paradigm, and against proposals to dispense with an autonomous morphological module of grammar.

Stump Gregory T. 2016a. "Paradigms at the interface of a lexeme's syntax and semantics with its inflectional morphology." In *Morphological Metatheory,* edited by Daniel Siddiqi and Heidi Harley, 27–58. Amsterdam: John Benjamins.
    Discusses cases of asymmetry between syntactico-semantic content and morphological form using evidence from the Old English conjugational system.

Stump Gregory T. 2016b. *Inflectional paradigms: content and form at the syntax-morphology interface*. Cambridge: Cambridge University Press.
    A comprehensive analysis of inflectional paradigms.

Lexeme and the Lexicon

Listedness is the property of being listed in the mental lexicon. Items listed in the mental lexicon are usually called lexical units or lexical items. More often than not, the terms lexical unit, lexical item, and

lexeme are used interchangeably in the literature. The property of being listed is not a crucial property of lexemes for two reasons. First, nonce words or newly formed words are arguably lexemes but are not listed in the lexicon. Second, not every lexical unit/item is a lexeme. Idioms are not lexemes for Matthews 1974. Lyons 1977 considers idioms to be lexemes. For Booij 2007, phrasal verbs are listemes in the sense of DiSciullo and Williams 1987 but not lexemes.

Booij, Geert. 2007. *The grammar of words: an introduction to linguistic morphology*. 2nd edition. Oxford: Oxford University Press.
> An introduction to morphology in which the author also examines whether phrasal verbs such as *look up* constitute lexemes and concludes that although they are lexical units (i.e. listemes), they are not lexemes.

DiSciullo, Anna Maria and Edwin Williams. 1987. *On the definition of word*. Cambridge, MA: MIT Press.
> This study discusses a number of uses of the notion word (syntactic atom, morphological object, phonological word) and introduces the term listeme, that is, an object that is memorized and listed in the lexicon.

Lyons, John. 1977. *Semantics*. Cambridge University Press, Cambridge.
> In this introduction to semantics, idioms belong to what Lyons calls "word-lexemes".

Matthews, Peter H. 1974. *Morphology*. Cambridge: Cambridge University Press.
> One of the works that established the term lexeme. In this work, the author does not consider idioms to be lexemes.

Lexeme and the Organization of the Lexicon

The terminological confusion with respect to the use of the term lexeme and other related terms is also reflected in studies on semantic relations (also, sense relations). As Murphy 2003 shows, it is not clear whether semantic relations relate words, lexemes, lexical items or lemmata. Lyons 1977 (cited under *Lexeme and the Lexicon*) defines hyponymy as a relation which holds between a subordinate lexeme and a superordinate lexeme. Antonymy is treated as a lexeme relation in Panther and Thornburg 2012. Synonymy is spoken of in relation to lexemes in Ci 1987. Cann 2011 offers an overview of sense relations among lexemes. Groups of lexemes with common semantics form lexical (also, semantic) fields. These fields that are usually modeled as frames or scripts offer a more holistic view of lexeme relations. Extensive discussion on lexical/semantic fields is found in Löbner 2013 (cited under *Textbooks*) and Lyons 1977 (cited under * Lexeme and the Lexicon *). In a number of approaches, the relations between lexemes are given as an organized hierarchy, as for example in Koenig 1999 and Davis 2001. Sag 2012 shows how derivation establishes a link between lexemes.

Cann, Ronnie. 2011. "Sense relations." In *Semantics: An International Handbook of Natural Language Meaning vol. 1*, edited by Claudia Maienborn, Klaus Von Heusinger and Paul Portner, 456-478. Berlin, Boston: Mouton de Gruyter.

    Discusses sense inclusion, sense exclusion, and identity of sense in relation to lexemes.

Ci, Jiwei. 1987. "Synonymy and polysemy." *Lingua* 72, no. 4: 315-331.

    Makes a case for three levels of synonymy, i.e. form-synonymy, lexeme-synonymy and sense-synonymy.

Davis, Anthony R. 2001. *Linking by types in the hierarchical lexicon*. Stanford: CSLI.

    Provides an account for the range of verb types via a hierarchical classification of verb meanings and the introduction of constraints.

Koenig, Jean-Pierre. 1999. *Lexical relations.* Stanford: CSLI.

    Proposes an organization of lexical knowledge as a model of lexical relations that builds on work in Head-Driven Phrase-Structure Grammar.

Murphy, Lynne M. 2003. *Semantic relations and the lexicon: antonymy, synonymy, and other paradigms*. Cambridge: Cambridge University Press.

    Offers a critical discussion of the literature on semantic relations and presents a pragmatic approach to these relations.

Panther, Klaus-Uwe, and Linda L. Thornburg. 2012. "Antonymy in language structure and use." In *Cognitive Linguistics Between Universality and Variation*, edited by Mario Brdar, Milena Žic Fuchs, and Ida Raffaelli, 161-188. Newcastle, UK: Cambridge Scholars.

    Proposes that antonymy is found on both the paradigmatic and the syntagmatic axes of language and language use.

Sag Ivan. 2012. "Sign-based construction grammar: an informal synopsis." In *Sign-based construction grammar*, edited by Hans Boas and Ivan Sag, 69-202. Stanford: CSLI.

    Shows that derivation establishes a relation between lexemes, while the link established by inflection is between lexeme and word-form.

Lemmas and Lexemes

The term lemma is used in the fields of psycholinguistics and computational linguistics in different ways. More often than not the lemma is identified with the citation form of a lexeme.

Lemmas and Lexemes in Psycholinguistics

In psycholinguistics, the terms lemma and lexeme are usually contrasted in models of speech production and lexical access. In particular, in the process of lexicalization (i.e. from meaning to

sound), the term lemma stands for the abstract form of a word in the lexicon and coves the syntactic and semantic properties of that word. The lemma does not carry phonological information. The term lexeme is often used for the specific form of a word (i.e. its sound). Thus, the lexeme is specified for phonology. Roelofs 1992, Roelofs 1998, and Levelt, et al. 1999 argue in favor of the lemma/lexeme distinction. Caramazza 1997 and Caramazza and Miozzo 1997 propose that there is a direct link between the lexical-semantic level and the lexeme level. Thus, they dispense with the lemma level. Laubstein 1999 uses blends to study the lemma/lexeme distinction. Kittredge, et al. 2008 shows how age of acquisition and lexical frequency have an effect on the lemma and lexeme levels in lexical access.

Caramazza, Alfonso. 1997. "How many levels of processing are there in lexical access?" *Cognitive Neuropsychology* 14, no. 1: 177–208.

    An accessible overview of the relevant literature. Proposes to abandon the lemma level.

Caramazza, Alfonso, and Michele Miozzo. 1997. "The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon." *Cognition* 64: 309-343.

    The authors present evidence from grammatical class deficits in anomic patients, tip-of-the-tongue states, and semantic substitution errors and make a case for abandoning the level of the lemma.

Kittredge Audrey K., Gary S. Dell, Jay Verkuilen, and Myrna F. Schwartz. 2008. "Where is the effect of frequency in word production? Insights from aphasic picture-naming errors." *Cognitive Neuropsychology* 25, no. 4: 463-492. https://doi.org/10.1080/02643290701674851.

    The authors present a regression analysis of aphasic picture naming errors and show the lexical frequency affects both the lemma level and the lexeme level.

Laubstein, Ann Stuart. 1999. "Lemmas and lexemes: the evidence from blends." *Brain and Language* 68, no. 1-2: 135-143. https://doi.org/10.1006/brln.1999.2091.

    Shows that word frequency effects occur at the lexeme level and not at the lemma level. This supports the distinction between the lemma level and the lexeme level.

Levelt, Willem J.M., Ardi Roelofs, and Antje S. Meyer. 1999. "A theory of lexical access in speech production." *Behavioral and Brain Sciences* 22:1-38.

    A highly influential paper on speech production that makes a case for the lemma level.

Roelofs, Ardi. 1992. "A spreading-activation theory of lemma retrieval in speaking." *Cognition* 42: 107-142.

    Makes a case for the lemma/lexeme distinction.

Roelofs, Ardi, Antje S. Meyer, and Willem J.M. Levelt. 1998. "A case for the lemma/lexeme distinction in models of speaking: comment on Caramazza and Miozzo (1997)." *Cognition* 69. 219-230.

> The authors extend their model of speaking to writing and the representation and production of morphologically complex items.

<u>Lemma and Lexeme in Computational Linguistics</u>

In computational linguistics, the term lemma is often used instead of the term lexeme or the citation form of a lexeme. Thus, *walk*, *walks*, and *walked* belong to the lemma *walk*. Lemmatization is the process of mapping from word-forms to a lemma. This process is context sensitive since it involves identifying the various senses of that lemma as well as the part of speech the lemma belongs to. The identification of lemmas and their senses is important for the creation of dictionary entries. Markó, et al. 2007 develops a method for the automatic acquisition of lexemes for a multilingual thesaurus. Chrupała, et al. 2008 and Müller, et al. 2015 develop systems that learn from morphologically annotated corpora. Bergmanis and Goldwater 2018 shows the importance of using contextual information for lemmatization. Roth et al. 2008 develops a system for lemmatization for a morphologically rich language, namely, Arabic. Leopold and Kindermann 2002 proposes to skip the lemmatization procedure for text categorization. Korenius, et al. 2004 compares lemmatization and stemming as word normalization methods for information retrieval. Gahl 2008 shows that lemma frequency has an impact on word durations.

Bergmanis, Toms and Sharon Goldwater. 2018. "Context sensitive neural lemmatization with Lematus." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 1391-1400. New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1126.

> The authors develop a context-sensitive lemmatizer with an improved performance on unseen and ambiguous words. They also critically comment on three other context-sensitive lemmatization systems.

Chrupała Grzegorz, Georgiana Dinu, and Josef van Genabith. 2008. "Learning morphology with morfette." In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, edited by Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias, 2362–2367. Marrakech: European Language Resources Association (ELRA).

> A probabilistic system which learns to perform joint morphological tagging and lemmatization from morphologically annotated corpora.

Gahl, Susanne. 2008. "Time and thyme are not homophones: the effect of lemma frequency on word durations in spontaneous speech." *Language* 84, no. 3: 474-496. https://doi.org/10.1353/lan.0.0035.

> The author shows that high-frequency words are shorter than their low-frequency homophones.

Korenius, Tuomo, Jorma Laurikkala, Kalervo Järvelin, and Martti Juhola. 2004. "Stemming and lemmatization in the clustering of Finnish text documents." In *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*, 625-633. ACM, New York.

    The authors show that, compared to stemming, lemmatization is a better word normalization method for information retrieval in Finnish.

Leopold, Edda, and Jörg Kindermann. 2002. "Text categorization with support vector machines. How to represent texts in input space?" *Machine Learning* 46: 423-444. https://doi.org/10.1023/A:1012491419635.

    The authors use support vector machines and show that the use of word-forms instead of lemmas leads to a substantial reduction of processing time without leading to a substantial loss of precision.

Markó, Kornél, Stefan Schulz, and Udo Hahn. 2007. "Automatic lexeme acquisition for a multilingual medical subword thesaurus." *International Journal of Medical Informatics* 76, no. 2-3: 184-189. https://doi.org/10.1016/j.ijmedinf.2006.05.032.

    The authors propose a method for the automated creation of new lexicon entries.

Roth, Ryan, Owen Rambow, Nizar Habash, Mona Diab, and Cynthia Rudin. 2008. "Arabic morphological tagging, diacritization, and lemmatization using lexeme models and feature ranking." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Short papers*, 117-120. Columbus, OH.

    The authors develop a system for morphological tagging, diacritization, and lemmatization for Arabic.

Thomas Müller, Ryan Cotterell, Alexander Fraser, and Hinrich Schütze. 2015. "Joint lemmatization and morphological tagging with lemming." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2268-2274. Lisbon, Portugal: Association for Computational Linguistics.

    The authors develop a system that consists of a classifier for lemmatization and a sequence model for tagging and show that taking into consideration morphosyntactic features can help the lemmatization task.

Polysemous, Monosemous, and Homonymous Lexemes

Lexemes may be monosemous, that is, have one meaning, or polysemous, that is, have more than one meaning. Polysemous lexemes may be disambiguated in context as shown in Pustejovsky 1995 and Asher 2011. Apresjan 1974 is an influential study on polysemy. Polysemy is often contrasted with homonymy. Two lexemes are homonymous if they have unrelated meanings, but share a number of

other properties, such as their sound form. The lexemes *bank* "financial institution" and *bank* "edge of a river" are homonymous since they sound the same but have different and unrelated meanings.

Apresjan, Juri D. 1974. "Regular polysemy." *Linguistics* 14, no. 2: 5-32.

  Discusses cases in which the relation between the various senses of a polysemous lexeme is predictable (i.e. regular polysemy).

Asher, Nicholas. 2011. *Lexical meaning in context*. Cambridge: Cambridge University Press.

  Makes a case for treating regular polysemy in terms of coercion.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: The MIT Press.

  A highly influential framework that assumes that there is a core meaning and that the various senses of a polysemous lexeme are derived from it via lexical rules.

Tests for Monosemy and Polysemy

The classification of a lexeme as monosemous or polysemous is non-trivial as shown in the debate on whether the verb *keep* is monosemous or polysemous found in Jackendoff 1992 and Fodor 1998. A number of tests to classify a lexeme as monosemous or polysemous are found in Zwicky and Saddock 1975 and Cruse 2000 (cited under *Textbooks*). Geeraerts 1993 offers a review of these tests.

Fodor, Jerry A. 1998. *Concepts: where cognitive science went wrong*. Oxford: Oxford University Press.

  Proposes that *keep* is monosemous and that its various meanings are a matter of the context into which *keep* is embedded.

Geeraerts, Dirk. 1993. "Vagueness's puzzles, polysemy's vagaries." *Cognitive Linguistics* 4, no. 3: 223-272.

  The author reviews the various criteria for polysemy and shows that they can lead to contradictory results.

Jackendoff, Ray. 1992. "What is a concept, that a person may grasp it?" In *Languages of the Mind: Essays on Mental Representation,* edited by Ray Jackendoff, 21-52. Cambridge, MA: MIT Press.

  Makes a case for treating *keep* as polysemous, in that the various concepts *keep* expresses share the same functional structure.

Zwicky, Arnold, and Jerry Sadock. 1975. "Ambiguity tests and how to fail them." In *Syntax and Semantics Volume 4*, edited by John P. Kimball, 1-36. New York: Academic Press.

  The authors propose the identity test by conjunction reduction which requires identity of sense but not of reference.

Semantic Approaches to Polysemy

There are two main semantic theories on the representation of polysemy: (a) sense enumeration lexicons defended in Katz 1972, Lakoff 1987, and Brugman 1988, and (b) core meaning approaches as developed in Caramazza and Grober 1976, Ruhl 1989, and Pustejovsky 1995 (cited under *Ambiguity*). Geeraerts 2010 offers an overview of the various approaches. Falkum and Vicente 2015 is a collection of papers on polysemy. Lieber 2004 offers a treatment of polysemy in word formation. Rainer 2014 offers an overview of polysemy in word formation.

Brugman, Claudia. 1988. *The story of over: polysemy, semantics, and the structure of the lexicon*. New York: Garland.

> A treatment of prepositional semantics in Cognitive linguistics. Proposes that the different senses of a polysemous lexeme are stored in the long-term memory rather than generated.

Caramazza, Alfonso, and Ellen Grober. 1976. "Polysemy and the structure of the subjective lexicon." In *Semantics: Theory and Application*, edited by Cléa Rameh, 181-206. Washington: Georgetown University Round Table on Languages and Linguistics.

> The authors make a case for treating polysemy in terms of one core representation in the lexicon. The various senses of the polysemous lexeme are a matter of context and semantic mechanisms of sense extension.

Falkum, Ingrid L., and Agustin Vicente, eds. 2015. Polysemy: current perspectives and approaches. *Lingua* 157: 1-172.

> A special issue on polysemy. Includes contributions from leading scholars in the field.

Geeraerts, Dirk. 2010. *Theories of lexical semantics*. Oxford: Oxford University Press.

> An overview of the main traditions in the lexical semantics. Very useful for undergraduate and advanced students.

Katz, Jerrold J. 1972. *Semantic theory*. New York: Harper and Row.

> A decompositional model of the semantic component of grammar that contains a dictionary and a set of projection rules. Different senses of a lexeme are listed under a single lexical entry in the dictionary.

Lakoff, George. 1987. *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago, Illinois: The University of Chicago Press.

> Makes a case for the representation of polysemous senses as a network of senses centered around a prototypical sense.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University

Press.
    An influential lexical semantic framework for the treatment of polysemy in word formation.
Rainer, Franz. 2014. "Polysemy in derivation." In the *Oxford handbook of derivational morphology*, edited by Rochelle Lieber and Pavol Štekauer, 338-353. Oxford: Oxford University Press.
    An overview of polysemy in derivation.

Ruhl, Charles. 1989. *On monosemy: a study in linguistic semantics*. Albany, New York: State University of New York Press.
    Treats the multiple senses of a lexeme in terms of a single core meaning that is shared by all the different senses of that lexeme.

Computational Approaches to Polysemy

Copestake and Briscoe 1995 offers an analysis of polysemy within an implemented constraint based approach. Pustejovsky 1995 (cited under *Ambiguity*) analyzes polysemy within the generative lexicon approach. Ravin and Leacock 2000 is a volume on the various theoretical and computational approaches to polysemy. Wilson and Carston 2007 offers a treatment of polysemy in lexical pragmatics. Berez and Gries 2009 offers a corpus-based analysis of polysemy. McCarthy, et al. 2016 is a recent study on identifying the various senses of a polysemous lexeme computationally.

Berez, Andrea L. and Stefan Th. Gries. 2009. "In defense of corpus-based methods: A behavioral profile analysis of polysemous get in English." In *Proceedings of the 24th Northwest Linguistics Conference. University of Washington Working Papers in Linguistics Vol. 27*, edited by Steven Moran, Darren S. Tanner, and Michael Scanlon, 157−166. Seattle, WA: Department of Linguistics.
    The authors apply a behavioral profile analysis to the verb *get* and show how a corpus-based approach can be used to determine the number of senses a word has and what their similarities and differences are.

Copestake, Ann, and Ted Briscoe. 1995. "Semi-productive polysemy and sense extension." *Journal of Semantics*, 12, no. 1: 15-67.
    The authors propose a distinction between constructional polysemy (i.e. contextual specialization in context) and sense extension which predictably relates two or more senses. Constructional polysemy is treated as instantiation of an underspecified lexical entry and sense extension by use of lexical rules.

McCarthy, Diana, Marianna Apidianaki, and Katrin Erk. 2016. "Word sense clustering and clusterability." *Computational Linguistics* 42: 245–275. https://doi.org/10.1162/COLI_a_00247.
    A computational study on how to partition the various usages of a word into discrete senses.

Ravin, Yael and Claudia Leacock, eds. 2000. *Polysemy: theoretical and computational approaches*. Oxford: Oxford University Press.

A collective volume that includes contributions by some of the leading authors in the fields of semantics and computational linguistics.

Wilson, Deirdre, and Robyn Carston. 2007. "A unitary approach to lexical pragmatics: relevance, inference and ad hoc concepts." In *Pragmatics*, edited by Noel Burton-Roberts, 230-259. London: Palgrave.

An inferential account of the way lexical narrowing, approximation, and metaphorical extension work within the framework of relevance theory.

Psycholinguistic Research on Ambiguity

Evidence that there is a difference in the way polysemous and homonymous lexemes are stored is found in Klepousniotou 2002, Klepousniotou and Baum 2007, Rodd, et al. 2002, and Beretta, et al. 2005. Experimental research on the enumeration and core meaning approaches is found in Pylkkänen, et al. 2006, Foraker and Murphy 2012, and Klein and Murphy 2001.

Beretta Alan, Robert Fiorentino, David Poeppel. 2005. "The effects of homonymy and polysemy on lexical access: an MEG study." *Cognitive Brain Research* 24: 57-65.

The authors offer behavioral and neurophysiological evidence for the homonymy/polysemy distinction. The evidence supports the separate entry account for homonymy, and a single-entry model of polysemy.

Foraker, Stephani, and Gregory. L. Murphy. 2012. "Polysemy in sentence comprehension: effects of meaning dominance." *Journal of Memory and Language* 67, no. 4: 407-425. https://doi.org/10.1016/j.jml.2012.07.010.

A study on how polysemous senses are represented and processed during sentence comprehension. The authors suggest that instead of a core meaning, readers select an individual sense when reading a polysemous word.

Frisson, Steven. 2015. "About bound and scary books: the processing of book polysemies." *Lingua* 157: 17–35. https://doi.org/10.1016/j.lingua.2014.07.017.

A set of two experiments to study the representation of polysemous words such as *book.* The results support the core meaning view.

Klein, Devorah E., and Gregory L. Murphy. 2001. "The representation of polysemous words." *Journal of Memory and Language* 45, no. 2: 259-282. https://doi.org/10.1006/jmla.2001.2779.

The results of five experiments suggest that each sense of a polysemous word has its own representation, and that there is only minimal semantic overlap between them.

Klepousniotou, Ekaterini. 2002. "The processing of lexical ambiguity: homonymy and polysemy in the mental lexicon." Brain and Language 81: 1-3.
https://doi.org/10.1006/brln.2001.2518.
A cross-modal sentence-priming lexical decision task that shows that the distinction between homonymy and polysemy is psychologically real.

Klepousniotou, Ekaterini, and Shari R. Baum. 2007. "Disambiguating the ambiguity advantage effect in word recognition: an advantage for polysemous but not homonymous words." *Journal of Neurolinguistics* 20, no. 1: 1-24.
The authors propose that polysemous words are processed faster than homonymous words. In addition, the processing of polysemous words depends on the type of sense extension mechanism since metonymies showed faster reaction times than metaphors.

Pylkkänen, Liina., Rodolfo Llinás, and Gregory L. Murphy. 2006. "The representation of polysemy: MEG evidence." *Journal of Cognitive Neuroscience*, 18, no. 1: 97-109.
https://doi.org/10.1162/089892906775250003.
The authors use behavioral and magnetoencephalographic measures in order to study the way polysemous senses are represented. The results corroborate the idea that the various senses are storage separately.

Rodd, Jennifer, Gareth Gaskell, and William Marslen-Wilson. 2002. "Making sense of semantic ambiguity: semantic competition in lexical access." *Journal of Memory and Language* 46, no. 2: 245-266.
A set of three lexical decision experiments that show that showed that highly polysemous words are recognized faster than polysemous words with few senses or homonyms.

Online Resources

WordNet is a database of lexical relations for English. For a description see Fellbaum 1998. **FrameNet**, as described in Baker, et al. 2003, is a lexical database of English that is based on Frame theory. CELEX2 is a database that includes morphological information (see Baayen, et al. 1995).

Baayen, Harald R., Richard Piepenbrock, and Leon Gulikers. 1995. *CELEX2 LDC96L14. Web Download.* Philadelphia: Linguistic Data Consortium.
Includes information on lemma frequencies and inflectional paradigms. Available *online[https://catalog.ldc.upenn.edu/ldc96l14]*.

Baker, Collin F., Charles J. Fillmore, and Beau Cronin. 2003. "The structure of the FrameNet database." *International Journal of Lexicography* 16, no. 3,1: 281-296.

    Describes the structure of FrameNet.

Fellbaum, Christiane, ed. 1998. *WordNet: an electronic lexical database*. Cambridge, MA: MIT Press.

    A description of WordNet.

FrameNet[https://framenet.icsi.berkeley.edu/fndrupal/]

    A lexical database of English that is based on Frame theory.

WordNet[https://wordnet.princeton.edu/]

    A lexical database of English in which items are grouped into sets of cognitive synonyms (synsets).
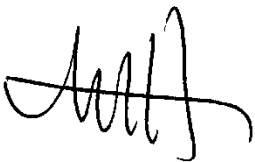
OXFORD
UNIVERSITY PRESS

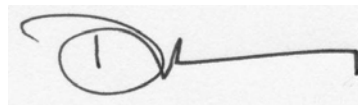November 19, 2018

Dear Dr. Andreou,

Congratulations on the acceptance of your article "Lexemes" for publication in *Oxford Bibliographies* in Linguistics. Your article has been peer reviewed by our editorial board and outside reviewers and final approval was given on November 16th, 2018. The articles within *Oxford Bibliographies* meet our highest standards of scholarship and your article will be a vital part of this important scholarly work.

Thank you very much for your contribution.

Sincerely,

Mark Aronoff
Editor in Chief
Oxford Bibliographies in Linguistics

Damon Zucca
Publisher, Reference
Oxford University Press, Inc.