CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

Part I

Theory

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

1
From Usage to Meaning
The Foundations of Distributional Semantics

Distributional semantics is the study of how distributional information can be used to model semantic facts. Its theoretical foundation has become known as the **Distributional Hypothesis**:

> Lexemes with similar linguistic contexts have similar meanings.

This chapter presents the epistemological principles of distributional semantics. In Section 1.1, we explore the historical roots of the Distributional Hypothesis, tracing them in several different theoretical traditions, including European and American structuralism, the later philosophy of Ludwig Wittgenstein, corpus linguistics, and psychology. Section 1.2 discusses the place of distributional semantics in theoretical and computational linguistics.

1.1 The Distributional Hypothesis

Distributional semantics was born in the early 1960s within the emerging field of computational linguistics. One of the first explicit mentions of this term is in Garvin (1962), who defines it as follows:

> Distributional semantics is predicated on the assumption that linguistic units with certain semantic similarities also share certain similarities in the relevant environments. [. . .] it may be possible to group automatically all those linguistic units which occur in similarly definable environments, and it is assumed that these automatically produced groupings will be of semantic interest. (Garvin, 1962, p. 388)

This definition already contains all the essential ingredients of distributional semantics, in particular its grounding assumption:

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

Distributional Hypothesis
The semantic similarity between two lexemes is a function of the similarity of their linguistic contexts.

DISTRIBUTIONAL HYPOTHESIS

The Distributional Hypothesis is a tree with many branches and multifarious roots. There are at least three different theoretical soils from which the Distributional Hypothesis has sprung: American structuralism (Section 1.1.1), the writings of later Wittgenstein (Section 1.1.2), and corpus linguistics (Section 1.1.3). These are very different theoretical traditions, but they share a descriptive perspective on language, and they all emphasize the importance of language use as the primary datum in linguistic theory. In Section 1.1.4, we illustrate the influence of distributionalism in psychology and the interpretation of the Distributional Hypothesis as a cognitive principle.

1.1.1 The Distributional Methodology in Structural Linguistics

DISTRIBUTIONALISM

The history of the Distributional Hypothesis predates computational linguistics, and originates outside the realm of traditional semantics. Its main root lies in the **distributionalism** advocated by American structuralists as the central method for making linguistics an empirical science. Prominent figures in this tradition include Bernard Bloch, Archibald Hill, Charles Hockett, Martin Joos, and George Trager, in addition to **ZELLIG S. HARRIS** (1909–1992), who is widely recognized as one of the fathers of mathematical approaches to linguistics, and the most influential theoretician of distributionalism. Harris' distributional program is delineated in his *Methods in Structural Linguistics* (1951), and is a consistent topic throughout works such as *Distributional Structure* (1954), *Mathematical Structures of Language* (1968), and *A Theory of Language and Information: A Mathematical Approach* (1991).

According to Harris, the basic elements of language can be identified in terms of their relative distributions:

> To be relevant these elements must be set up on a distributional basis: *x* and *y* are included in the same element *A* if the distribution of *x* relative to the other elements *B*, *C*, etc. is in some sense the same as the distribution of *y*. (Harris, 1951, p. 7)

The essence of the distributional methodology, as defined by Harris in the quote above, is thus quite clear and simple: The basic building blocks of language can be identified by their relative distributions in (samples of) language. This means that if we have two elements x and y with identical distributions, then they are functionally equivalent and should be regarded as the same

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

distributional element. Of course, everything depends on what we mean by "distribution." Harris clarifies his use of the term in the following words:

> The ENVIRONMENT or position of an element consists of the neighborhood, within an utterance, of elements which have been set up on the basis of the same fundamental procedures which were used in setting up the element in question. [...] The DISTRIBUTION of an element is the total of all environments in which it occurs, i.e. the sum of the (different) positions (or occurrences) of an element relative to the occurrence of other elements. (Harris, 1951, pp. 15–16)

The term "environment" refers here to the **linguistic context** of an element and is formed by its neighboring elements. Harris' statement above thus constitutes a very clear and concise formulation of the distributional methodology:

LINGUISTIC CONTEXT

DISTRIBUTION

> Linguistic elements are identified by their **distributions**, defined as the sum of the contexts in which they occur.

In the same way, categories of elements can be identified by the distributional similarity of their constituent elements. For Harris (and other proponents of the distributional methodology), the entirety of language – phonology, morphology, grammar – could be described according to distributional criteria.

As explicitly acknowledged by Harris, distributionalism originates in the pioneering works of **EDWARD SAPIR**[1] (1884 – 1939) and **LEONARD BLOOMFIELD** (1887 – 1949), whose program for a **structural** and **descriptive** linguistics is founded on three main tenets:

STRUCTURAL LINGUISTICS

1. every language has a structure of its own, and there are no universal linguistic categories;
2. the study of language must be primarily synchronic;
3. linguistics must be autonomous with respect to other disciplines, especially psychology.

The distributional method is considered instrumental in achieving these goals, in particular to provide linguistics with an independent and methodologically sound foundation: The only data that are scientifically valid for Bloomfield are observable linguistic phenomena in the form of distributional patterns.

[1] Quoting a letter by Morris Swadesh, Nevin (1993) suggests that the origin of the use of the term distribution in linguistics was Sapir, who employed it as a geographical metaphor: "It was an application of the usage represented by 'geographic distribution', an expression which was much used by Sapir as by other anthropologists and linguists."

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

Distributionalism and Meaning

It is a common conception that semantics is in the periphery, if not completely ignored, in the structural linguistic tradition in general, and in the works of Zellig Harris in particular.[2] It is certainly true that Harris himself does not offer a distributional semantics and that his distributional project is primarily occupied with phonology and morphology, but the reason for this seemingly agnostic stance toward semantics is less commonly understood.

In order to fully appreciate Harris' perspective of the relationship between distributional properties and meaning, it is useful to first flesh out the ultimate conclusion of Harris' concerns about the scientific status of linguistic methodology. At the core of these concerns is the realization of the peculiar position of linguistics as a science, since it does not have recourse to a metalanguage that is external to its object of study. On the contrary, *language contains its own metalanguage*. We cannot describe language in something other than language, and any use of symbols needs to be defined ultimately in language:

> There is no way to define or describe the language and its occurrences except in such statements said in that same language or in another natural language. Even if the grammar of a language is stated largely in symbols, those symbols will have to be defined ultimately in a natural language. (Harris, 1991, p. 274)

This quandary has an important consequence: If the information in language can only be described in language, then it follows that this information cannot be an encoding of some prior representation of the information (e.g., a mental representation). This is a strong argument against mentalism, and for the scientific viability of the distributional approach. For Harris, this means that the science of language cannot deal with anything other than the elements of language, and their relationships to one another, that is, their distribution.

Following this line of reasoning, Harris argues – just like Bloomfield – that the form of a lexeme is not something different from the meaning conveyed by it. In the words of Bloomfield (1943): "in language, forms cannot be separated from their meanings." This is a point in the descriptive tradition that is often misconstrued. The insistence on the unity of form and meaning should not be understood as a denial of the existence of extralinguistic meaning. On the contrary, Harris – and even more explicitly Bloomfield – vigorously argues that meaning in all its social manifestations is most decidedly outside the scope of linguistic theory. As Bloomfield states: "the statement of meanings is therefore the weak point in language study, and will remain so until human knowledge advances far beyond its present state" (Bloomfield, 1933, p. 140). The best we

[2] Some commentators even label the whole American structuralist tradition "anti-semantic" (e.g., Murphy, 2003).

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

can do within the descriptive linguistic project is to describe the observable manifestations of meaning, and indeed, *any* meaning (regardless of what it is and where it comes from) that can be conveyed in language *must* have a formal manifestation, since otherwise it would not be expressible in language: "a language can convey only such meanings as are attached to some formal feature: the speakers can signal only by means of signals" (Bloomfield, 1933, p. 168). Therefore, Bloomfield concludes, "in all study of language we must start off from forms and not from meanings" (Bloomfield, 1943, p. 402). The proper interpretation of this claim is that semantic considerations cannot enter into the definition of linguistic elements (e.g., "nouns denote things"), which must instead be defined in distributional terms (Goldsmith and Huck, 1991).[3]

Likewise, according to Harris, linguistic analysis cannot be founded on "some independently discoverable structure of meaning" (Harris, 1954, p. 152). It is meaning that must be studied as a function of linguistic distributions:

MEANING AND DISTRIBUTION

> if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution. (Harris, 1954, p. 156)

Harris' words echo those of another structuralist, Martin Joos, who claims that "the linguist's meaning of a morpheme [. . .] is by definition the set of conditional probabilities of its occurrence in context with all other morphemes" (Joos, 1950, p. 708). The point is that a difference in meaning between two lexemes will be reflected by a difference in distribution, and this difference in distribution will be observable through distributional analysis:

> If A and B have almost identical environments except chiefly for sentences which contain both, we say they are synonyms: *oculist* and *eye-doctor*. If A and B have some environments in common and some not (e.g. *oculist* and *lawyer*) we say that they have different meanings, the amount of meaning difference corresponding roughly to the amount of difference in their environments. (Harris, 1954, p. 157)

In his later works, Harris characterizes linguistic environments in terms of syntactic dependencies involving relations between a word acting as **operator** and a word acting as its **argument**. The "selection" (i.e., the distribution) of a word

OPERATOR AND ARGUMENT

[3] Bloomfield's semantic skepticism concerns any approach to meaning, including the behaviorist and physicalist ones that he favored, since statements about meaning lie well beyond the limits of linguistic science: "There is nothing in the structure of morphemes like *wolf*, *fox*, and *dog* to tell us the relation between their meanings. This is problem for the zoölogist. The zoölogist's definition of these meanings is welcome to us as a practical help, but it cannot be confirmed or rejected on the basis of our science" (Bloomfield, 1933, p. 162).

© in this web service Cambridge University Press & Assessment www.cambridge.org

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

is the set of operators and arguments with which it co-occurs with a statistically significant frequency, and is strongly correlated with its meaning:

> It is thus that selection can be considered an indicator, and indeed a measure, of meaning. Its approximate conformity to meaning is seen in that we can expect that for any three words, if two of them are closer in meaning to each other than they are to the third, they will also be closer in their selection of operators and arguments. (Harris, 1991, p. 329)

Meaning "is a concept of no clear definition" (Harris, 1991, p. 321), but distributional analysis can turn it into a measurable and scientific notion:

> Selection is objectively investigable and explicitly statable and subdividable in a way that is not possible for meanings – whether as extension and referents or as sense and definition. (Harris, 1991, p. 329)

The goal of Harris' distributional program is therefore not to exclude meaning from the study of language (Harris, 1991, pp. 42–43), but rather to provide a scientific foundation for its investigation. Even if Harris has never explicitly formulated the Distributional Hypothesis, he argues that if we are to deal with meaning in language, *we can only do so through distributional analysis*. It is this idea of a correlation between meaning differences and distributional properties that lies at the heart of distributional semantics.

Syntagms and Paradigms: Distributionalism in Europe

Distributionalism is a direct product of American structuralism but is also strongly indebted to European structuralists like Ferdinand de Saussure, Louis Hjelmslev, and the Prague School, most notably represented by Nikolai Trubetzkoy and Roman Jakobson. According to Harris, the (semantic) relation between two words or morphemes is defined differentially, based on their distributional behavior within the language system, without recourse to an external world. This view recalls the words of the father of structuralism, **Ferdinand de Saussure** (1857–1913): "dans la langue il n'y a que des differences" ("in language there are only differences"; Saussure, 1916, p. 166).

VALUE AND SYSTEM

In structuralist theory, as it emanates from Saussure's posthumously published seminal work *Cours de linguistique générale* (1916), the term **valeur** "value" is used to define the function of a lexeme within the **language system**. A lexeme has a value only by virtue of being *different* with respect to the other lexemes. Such a differential view on the functional distinctiveness of linguistic elements highlights the importance of the system as a whole, since differences cannot exist in isolation from the system itself. A single isolated lexeme cannot enter into difference relations, since there are no other lexemes to differ from (and no system to define it functionally). In this view, the language system

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

Table 1.1 Syntagmatic and paradigmatic relations

	Paradigmatic relations Selections: “x or y”			
Syntagmatic relations Combinations: “x and y”	she he they	adores likes love	green blue red	paint dye colour

becomes an interplay of functional differences, which can be divided into two kinds: **syntagmatic** and **paradigmatic relations**.[4]

SYNTAGMATIC RELATIONS

> Syntagmatic relations hold between lexemes that co-occur in sequential combinations. A **syntagm** is such an ordered combination of lexemes.

PARADIGMATIC RELATIONS

> Paradigmatic relations hold between lexemes that do *not* themselves co-occur, but that co-occur with the same *other* lexemes. Paradigmatically related lexemes can be substituted for one another in the same context. Such a set of substitutable lexemes constitutes a **paradigm**.

The term “syntagm” corresponds to what we have called “context.” The only difference is that the former term implies an ordered set of neighboring lexemes, while the latter term does not. Syntagmatic and paradigmatic relations can be depicted as orthogonal axes in a two-dimensional grid. In the example in Table 1.1, the paradigms correspond to morphosyntactic classes, like adjectives and verbs, but they also define semantic categories, such as color terms.

In his essay *On linguistic aspects of translation*, **Roman Jakobson** (1896–1982) argues that meaning is a linguistic phenomenon:

> The meaning of the words “cheese,” “apple,” “nectar,” “acquaintance,” “but,” “mere,” and of any word or phrase whatsoever is definitely a linguistic – or to be more precise and less narrow – a semiotic fact [. . .]. There is no *signatum* without *signum*. The meaning of the word “cheese” cannot be inferred from a nonlinguistic acquaintance with cheddar or with camembert without the assistance of the verbal code. (Jakobson, 1959, p. 232)

Like for De Saussure, words have meaning only within a linguistic system, in which they are used and entertain various relations with other expressions. It is

[4] Saussure uses the term *associative* relation rather than paradigmatic relation. It was Hjelmslev who introduced the term “paradigmatic” relation.

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

LINGUISTIC AND DIRECT ACQUAINTANCE

the knowledge of such relations that Jakobson calls **linguistic acquaintance**, whose importance supersedes the role of the **direct acquaintance** with the entities words refer to. The latter may lack (e.g., we can use *ambrosia* correctly even without direct experience of its referent), while linguistic acquaintance is essential to understand the meaning of any lexeme (cf. Sections 8.7–8.8).

STRUCTURAL SEMANTICS

Compared to its American counterpart, European structuralism attributes considerable importance to word meaning analysis. **Structural semantics**, represented by Jost Trier, Adrienne Lerher, Eugenio Coseriu, Algirdas Greimas, John Lyons, Alan Cruse, among many others, is a family of theories focusing on the paradigmatic organization of the lexicon (Murphy, 2003; Geeraerts, 2010). The theoretical apparatus of structural semantics includes:

SEMANTIC FIELDS

SEMANTIC FEATURES

SEMANTIC RELATIONS

1. **lexical** or **semantic fields**, sets of mutually related lexemes defining the conceptual structure of a certain domain, such as the color domain;
2. **semantic components** or **features**, inspired by structuralist phonology and used to describe meaning in terms of basic oppositions (e.g., +/– ANIMATE); and
3. **paradigmatic semantic relations** between lexemes, such as **synonymy** (sameness in meaning; *sofa* – *couch*), **antonymy** (opposition in meaning, *good* – *bad*), **hypernymy** (a taxonomic relation where a **hypernym** is a more general term than its **hyponym**, *animal* – *dog*), and **co-hyponymy** (lexemes that share the same hypernym; *dog* – *cat*).

Structural semantics is autonomous from distributionalism, but the latter is often adopted as a method to define semantic paradigms in terms of syntagmatic relations. The Distributional Hypothesis can indeed be reformulated in structuralist terms (Sahlgren, 2006):

> Words sharing syntagmatic contexts have similar paradigmatic properties.

For instance, Apresjan (1966) refers to Harris' distributional methodology as a way to provide more objectivity to the investigation of semantic fields by grounding it on linguistic evidence. Apresjan carries out a distributional analysis of adjectives in terms of their frequency of co-occurrence with various syntactic contexts. The interplay between syntagmatic and paradigmatic dimensions is also central for Cruse (1986): The greater the paradigmatic "affinity" of lexical items, the more congruent their patterns of syntagmatic relations.

1.1.2 Meaning as Use: The Echoes of Wittgenstein

The central principle of structuralism and the distributional methodology – that we should let data decide what our models of language encompass – echoes in

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-00429-0 — Distributional Semantics
Alessandro Lenci , Magnus Sahlgren
Excerpt
More Information

Wittgenstein's insistence that we should "look and see" (Wittgenstein, 1953, §66) rather than presume. The intellectual work of **LUDWIG WITTGENSTEIN** (1889–1951) can be divided into two distinct periods: the early period of the *Tractatus Logico-Philosophicus* (1922), where he professes a logic-centered view on language, and the later period of *Philosophical Investigations* (1953), in which he explicitly rejects his earlier ideas about the nature of language. In this work, Wittgenstein is openly polemic against the view that we need a logical representation to obliterate the vagueness and incompleteness of natural language. Wittgenstein urges us not to *assume* a general and fixed meaning of words. Instead, we should look at *how* the words are being used:

> For a large class of cases – though not for all – in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language. (Wittgenstein, 1953, §43)

MEANING AS USE

This has sometimes been called a **usage-based theory** of meaning, but Wittgenstein is not so much offering a theory of meaning in his later works as pointing out a misconception regarding the nature of meaning. The misconception, according to him, consists in construing meaning as primarily a naming relation, in such a way that meaning is something (like a mental or physical object) that a word (or phrase or sentence) names. Such a *nomenclaturist* view on meaning has been both widespread and withstanding in the history of linguistics and the philosophy of language (although Wittgenstein's aim was primarily to attack his own earlier views), and the *anti-nomenclaturist* stance is a position the later Wittgenstein shares with the structuralist movement, and in particular with the contemporary Saussure (Harris, 1988).

This is not the only point of contact between the ideas of Wittgenstein and structuralist linguistics, and more specifically with the ideas of Saussure.[5] The former's view on meaning as founded in the use of language has striking similarities to the latter's concept of *valeur*: It is the *role* of the lexeme in language that constitutes its meaning. Wittgenstein even expresses himself in terms that could as well have been Saussure's: "the sign (the sentence) gets its significance from the system of signs, from the language to which it belongs" (Wittgenstein, 1958, p. 5). For both Wittgenstein and Saussure, meaning can be likened with the role or function of a word within language; indeed, they make heavy use of the game-metaphor – and in particular chess – for describing the holistic functional character of the language system.

Wittgenstein also stresses the importance of the social aspect of meaning and language use, just as the prominent figures of the structuralist tradition

[5] Despite the similarities between the ideas of Wittgenstein and Saussure, there is no evidence that they were influenced by (or even aware of) each other's works (Harris, 1988).

© in this web service Cambridge University Press & Assessment www.cambridge.org