A DEFENSE OF MEANING ELIMINATIVISM:
A CONNECTIONIST APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

TOLGAHAN TOY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF PHILOSOPHY

MARCH 2022

Approval of the thesis:

A DEFENSE OF MEANING ELIMINATIVISM:
A CONNECTIONIST APPROACH

submitted by **TOLGAHAN TOY** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Philosophy, the Graduate School of Social Sciences of Middle East Technical University** by,

Prof. Dr. Yaşar KONDAKÇI
Dean
Graduate School of Social Sciences _______________

Prof. Dr. Halil Ş. TURAN
Head of Department
Department of Philosophy _______________

Prof. Dr. David GRÜNBERG
Supervisor
Department of Philosophy _______________

Examining Committee Members:

Assoc. Prof. Dr. Aziz F. ZAMBAK (Head of the Examining Committee)
Middle East Technical University
Department of Philosophy _______________

Prof. Dr. David GRÜNBERG (Supervisor)
Middle East Technical University
Department of Philosophy _______________

Prof. Dr. Catherine Z. ELGIN
Harvard University
Department of Philosophy _______________

Assoc. Prof. Dr. Ingrid L FALKUM
Oslo University
Department of Philosophy, Classics, History of Art and Ideas _______________

Assist. Prof. Dr. Majid D. BENI
Middle East Technical University
Department of Philosophy _______________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Tolgahan TOY

Signature:

ABSTRACT

A DEFENSE OF MEANING ELIMINATIVISM:
A CONNECTIONIST APPROACH

TOY, Tolgahan
Ph.D., The Department of Philosophy
Supervisor: Prof. Dr. David GRÜNBERG

March 2022, 158 pages

The standard approach to model how human beings understand natural languages is the symbolic, compositional approach according to which the meaning of a complex expression is a function of the meanings of its constituents. In other words, *meaning* plays a fundamental role in the model. In this work, because of the polysemous, flexible, dynamic, and contextual structure of natural languages, this approach is rejected. Instead, a connectionist model which eliminates the concept of *meaning* is proposed.

Keywords: Natural language understanding, Connectionism, Polysemy, Compositionality, Context-dependency

ÖZ

ELEYİCİ ANLAM KURAMININ BİR SAVUNUSU:
BAĞLANTICI BİR YAKLAŞIM

TOY, Tolgahan
Doktora, Felsefe Bölümü
Tez Yöneticisi: Prof. Dr. David GRÜNBERG

Mart 2022, 158 sayfa

İnsanların doğal dilleri nasıl anladığını modellemenin bir yolu, karmaşık bir ifadenin anlamının, bileşenlerinin anlamlarının bir fonksiyonu olarak görüldüğü sembolik, bileşimsel yaklaşımdır. Bu çalışmada, anlam kavramı üzerine kurulu olan sembolik bileşimsel görüşün, doğal dillerin çokanlamlı, esnek, dinamik ve bağlamsal yapısını modelleyemeyeceği iddia edilmektedir. Alternatif olarak, anlam kavramı üzerine kurulu olmayan bir dil anlama modeli önerilmektedir. Bu modelin gerçekleştirilmesi noktasında ise bağlantıcılık görüşündeki gelişmelerden faydalanılması önerilmektedir.

Anahtar Kelimeler: Doğal dil anlama, Bağlantıcılık, Çokanlamlılık, Bileşimsellik, Bağlamsallık

To freedom and individuality

ACKNOWLEDGEMENT

First and foremost, I am extremely grateful to my supervisor, Prof. David Grünberg, for his invaluable advice, continuous support, and patience during my Ph.D. study.

I would like to express my gratitude to Prof. Catherine Elgin for her invaluable supervision and support during my visiting at Harvard University.

I would like to thank Prof. Øystein Linnebo for his invaluable guidance on studying Frege's context and compositionality principles. I would like to thank CSMN for providing me with generous funding to spend a semester at Oslo University.

I would like to thank Prof. Cem Bozşahin, Dr. Ceyhan Temürcü, Assist. Prof. Umut Özge, Prof. Deniz Zeyrek Bozşahin from the Department of Cognitive Science at METU; Dr. Özkan Kılıç from CISCO; Prof. Ferda Nur Alpaslan, Assoc. Prof. Sinan Kalkan from the Department of Computer Engineering at METU; Prof. Roger Levy from the Department of Brain and Cognitive Sciences at MIT; Prof. Kai von Fintel, Prof. Irene Heim, Prof. Danny Fox from the Department of Linguistics and Philosophy at MIT for allowing me to take their classes. These courses have advanced my knowledge of computer science, linguistics, and cognitive science. Without them, this research would not have been possible.

I also thank MIT Open Courseware and edX.

I am grateful to Assoc. Prof. Ingrid Falkum for her constructive criticism of my approach to language understanding.

I appreciate Prof. Agustin Rayo for our delightful conversations on Frege's context principle and his own language understanding model.

I would like to thank Assoc. Prof. Aziz Zambak for encouraging me to go further in computational philosophy.

I would like to thank Prof. László E. Szabó for inviting me to spend a summer in the Department of Logic at Eötvös University in Budapest.

I would like to thank Ph.D. candidates Berk Yaylım and Serdal Tümkaya for helpful discussions on neurophilosophy and connectionism.

Special thanks to my beloved parents, Necati Toy and Nezahat Toy from whom I learned the value of freedom and trust. I would also like to thank my brothers Alaaddin Toy and İsmail Toy for motivating discussions on science, religion, art, music, and mathematics.

Finally, I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for its generous grant, 2214-A International Research Fellowship Programme for Doctoral Students, which provided me with an opportunity to carry out my doctoral research at Harvard University.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The subject of natural language understanding is the relation between natural language expressions and language users' behaviors. Upon hearing a sequence of words (e.g., "dog," "dog chases the cat," "snow is white," etc.), a language user responds to it with certain behaviors.

Expressions in natural languages are in the part-whole structure. Every expression is either a primitive, basic, or complex one consisting of other expressions. In this work, since we will not go into morphology, by *word*, I will refer to the basic, primitive expressions. In addition to these primitive expressions, *words*, there are *complex phrases* (e.g., "the book I read," "on the mat," "gave Kripke a book"), *sentences* (e.g., "Snow is white," "Cat is on the mat," etc.) and so on. An essential part of semantics is the way we understand sentential expressions. How do we understand the sentence "John believes that Superman is a hero."? Do we grasp a proposition, a truth value, or confront a state of affairs?

Lexical semantics, in particular, is interested in the way we understand the basic constituents, *words*. How does a word contribute to our understanding of expressions? For example, how does "Aristotle" contribute to our understanding of the sentences "Aristotle is the teacher of Alexander the Great" and "Her Ph.D. dissertation is on Aristotle"? The standard approach is to postulate *meanings*. For this approach, each expression has a meaning. Meanings of smaller expressions contribute to the meaning of complex expressions recursively. It implies that primitive expressions' meanings are the base in this recursive model. Depending on the degree of contribution, there is a wide spectrum of theories. On the one edge, there is the ordinary concepts view according to which word meaning makes a very significant contribution to the meaning of complex expressions. This view is in tandem with the idea that the word meaning determines the sentence meaning. Once we know the meaning of words, there is nothing much left to understand the meanings of sentences that consist of these words. On the other edge of the universe, we have radical contextualism, which claims that word meaning makes little contribution. This idea is consistent with the context dependent features of words. For example, if a word does not make the same contribution in all sentences, then this view is a real alternative. There are hybrid models between these two views in which word meanings make a systematic contribution but only with a contextual adjustment. Therefore, which theory of word meaning we embrace depends on the two opposite directions below:

- The relation between the surface syntax of a complex expression (i.e., sentence) and its meaning.
- Context dependency of word meaning.

The first direction is compositionality, while the second represents context dependency. In this work, because of the polysemous, network-like, intricate, dynamic structure of natural languages, I will reject these meaning theories altogether. I will defend the idea that expressions do not have any meaning. In other words, I will propose a framework that explains language understanding without postulating meanings. I will appeal to connectionist, sub-symbolic approaches to show how this idea can be implemented.

In the rest of the work, first, I will show the place of the symbolic, compositional approach to language understanding in the history of thought. I will introduce both rule-based and pragmatics-oriented versions of the compositional approach. Then, I will argue that the polysemous, context dependent, dynamic sides of the natural language pose a vital problem for the compositional, symbolic approach. After identifying the problem with using meanings in the computation, I will adopt the most radical contextualist approach that eliminates meaning altogether. To maintain this position, I appeal to the sub-symbolic modeling. In the last chapter, I will respond to the general philosophical questions regarding the proposed approach.

CHAPTER 2

THE COMPOSITIONALITY PRINCIPLE

2.1 Compositionality of meaning

One of the question meaning theories ask is how to compute the meanings of sentences from words. One way to achieve this involves the compositionality principle, according to which the meaning of a complex expression is determined by the meanings of its parts plus syntax (Szabó (2012), p.75) (Szabó (2017)) (Partee (1984), pp. 281-282) (Gamut (1990), p. 140) (Pagin and Westerståhl (2010a), p. 250) (Hodges (2001), p. 7). Compositionality, which stems from Gottlob Frege's function-argument analysis of sentences, is formalized as a homomorphism from syntax to semantics (Frege (1967),$\S 9$) (Linnebo (2018), pp. 120-121) (Montague (1974b), p. 227). Given a set of expressions *A*, a set of meanings *B*, a semantic operation *G*, a syntactic operation *F*, and a set of indices to identify operations, "a semantic interpretation for a language is then defined as some homomorphism from $< A, F_\gamma >_{\gamma \in \Gamma} to < B, G_\gamma >_{\gamma \in \Gamma}$" (Dowty (2007), pp. 33-34).
In this account, a semantic interpretation of an expression $F_ß(\alpha_1 .... \alpha_n)$ is provided as below:

$$h(F_ß(\alpha_1 .... \alpha_n)) = G_ß(h(\alpha_1) .... h(\alpha_n)).$$

For example, compositionally, the meaning "Fido barks" is given below.

$$meaning - of(SYNTACTIC - COMBINATION_ß - of(\mathbf{F}ido, barks)) =$$
$$SEMANTIC - FUNCTION_ß - of(meaning - of(Fido), meaning - of(barks))$$

(Dowty (2007), p. 34) (Pagin (2003), p. 299) (Barbara H. Partee (1990), p. 334)

2.2 The history of the compositionality principle

2.2.1 Ancient and Medieval Approaches

Let's now go on with how the compositional, symbolic approach of language understanding is prevalent in the history of human thought. Aristotle starts *De Interpretatione* with the following remarks. “First we must define the terms ‘noun’ and ‘verb,’ then the terms ‘denial’ and ‘affirmation,’ then ‘proposition’ and ‘sentence’ " (McKeon (2001), 16a). Then he continues, “nouns and verbs, provided nothing is added, are like thoughts without combination or separation; ‘man’ and ‘white,’ as isolated terms, are not yet either true or false. In proof of this, consider the word ‘goat-stag.’ It has significance, but there is no truth or falsity about it, unless ‘is’ or ‘is not’ is added, either in the present or in some other tense" (McKeon (2001), 16a).

This passage shows that Aristotle makes a categorical distinction between sentence and sub-sentential parts in terms of semantics. He claims that a sentence is semantically decomposable into smaller parts while a name or a verb is not (McKeon (2001), 16a-16b). Similarly, complex sentences are built upon simple sentences in a logical way ((McKeon (2001), 17a-24b).

Aristotelian medieval philosopher Al-Farabi summarizes this point as follows. “A statement is composed of a noun and a verb connected, whenever their combination results in a statement... And that a predicate is connected with a subject means that the predicate holds of the subject" (Zimmermann (1981), p.1). Wilfrid Hodges reports that another Aristotelian philosopher Ibn Sina, in addition to formulating the semantic compositionality, mentions context dependency of word meaning (Hodges (2012), pp. 1-2).

Among the medievals, Peter Abelard's philosophy of language is more closer to the contemporary approaches. There are some similarities between his approach and Frege's approach. Abelard, rejecting the existence of universals and Roscelin's vocalism, which treats universal expressions as mere sounds, faces Frege's dilemma in “Über Sinn und Bedeutung." Like Frege or maybe the more contemporary two dimensionalists, he analyses meaning at two distinct levels. One component of his semantics is the ideas that expressions convey to our minds. The second component is the objects that expressions refer to (King and Arlig (2018)). Peter King reports that at the first level, we can talk about the compositionality:

> The expression ‘rational stone’ corresponds to a conjoining understanding, in this case | |rational| + |stone| |, as much as ‘red rose’ does, despite the fact that there are red roses and there are no rational stones. Whether an understanding accurately reflects the world is a separate question, namely whether it is “sound or empty/vain" (*sanum uel casus/uanus*), as Abelard puts it. Semantics is not metaphysics" (King (2007), p.179).

According to King, in Abelard’s semantics, syntactic structure plays a pivotal role in understanding a sentence (King (2007), pp. 175-176). To exemplify this point, King quotes Abelard: “Someone who hears [the sentence “Man walks"] proceeds by collecting the appropriate understandings from each of the words: first by understanding *man* when he hears “man" (which is instituted to signify it); thereafter by understanding *walking* when he hears “walks"; finally, connecting it to *man*."(King (2007), p. 176)

In modern philosophy, Descartes attracts our attention to the productive and systematic structure of meaning. We are able to understand an infinite number of expressions without having any difficulty. According to Descartes, this faculty is what distinguishes us from *merely physical beings*. In other words, a human-made machine can never understand the language.

> If there were machines which had a likeness to our bodies and imitated our actions, inasmuch as this were morally possible, we would still have two very certain means of recognizing that they were not, for all that, real men. Of these the first is, that they could never sue words or other signs, composing them as we do to declare our thoughts to others. For one can well conceive that a machine may be so made as to emit words, and even that it may emit some in relation to bodily actions which cause a change in its organs, as, for example, if one were to touch it in a particular place, it may ask what one wishes to say to it; if it is touched in another place, it may cry out that it is being hurt, and so on; but not that it may arrange words in various ways to reply to the sense of everything that is said in its presence, in the way that the most unintelligent of men can do (Descartes (1968), pp. 73-74).

The dualist position here is in contrast with the position, later taken by Alan Turing. For him, the imitation of human language is possible with a mechanical procedure (Turing (1950), p. 442). For the creativity objection, he says, “a variant of Lady Lovelace’s objection states that a machine can ‘never do anything really new.’ This may be parried for a moment with the saw, 'There is nothing new under the sun.' Who can be certain that 'original work' that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles." (Turing (1950), p. 442)

The precursor of this position is another philosopher, Thomas Hobbes, who is a contemporary of Descartes. Hobbes does not postulate immaterial, mysterious entities to explain the meaning of complex expressions. Instead, he thinks that what stands behind our understanding of language is the *ratiocination*.

> By ratiocination, I mean *computation*. Now to compute, is either to collect the sum of many things that are added together, or to know what remains that are added together, or to know what remain when one thing is taken out of another. *Ratiocination*, therefore, is the same with *addition* and *subtraction*; and if any man add *multiplication* and *division*, I will not be against it, seeing multiplication is nothing but addition of equals one to another, and division nothing but a subtraction of equals one from another, as often as is possible. So that all ratiocination is comprehended in these two operations of the mind, addition and subtraction" (Hobbes (1655), p. 3).

To explain how we compute the meaning of a sentence first, he introduces the meaning of different word types (Hobbes (1655), pp. 16-28). Then, he continues with how we combine meanings of the constituents to obtain the meaning of a sentence: Subjects and predicates are combined with copulas (Hobbes (1655), pp. 30-32).

2.2.2 Gottlob Frege

The compositionality principle in contemporary semantics is attributed to Frege (Szabó (2017)). Frege, in his paper, "Über Sinn und Bedeutung," introduces this basic principle for meaning studies: the meaning of an expression is a function of the meanings of its parts. In his theory, the semantic structure reflects the syntactic structure. He says that "I have here used the word 'part' in a special sense. I have in fact transferred the relation between the parts and the whole of the sentence to its referent, by calling the referent of a word part of the referent of the sentence, if the word itself is a part of the sentence." (Frege (1948), p. 217)

The compositionality principle is the reason behind proposing sophisticated solutions for the two famous puzzles in the philosophy of language literature: The identity puzzle and the propositional attitude puzzle. In his earliest work *Begriffsschrift*, he used *conceptual content* for the semantic value of expressions. However, this confronts him with the identity puzzle. The puzzle is both sameness and difference of the semantic value of sentences of the form:

$$a = a$$

$$a = b$$

It can be exemplified with the sentences below.

$$The\ morning\ star\ is\ the\ morning\ star$$

$$The\ morning\ star\ is\ evening\ star.$$

If identity is a relation between what is signified by the terms flanking the identity sign, then these two sentences have the same meaning. However, we know that there is a difference between these two sentences. For example, while the first one can be known a priori, the other can only be known a posteriori. In *Begriffsschrift*, Frege's solution is that "identity of content differs from conditionality and negation in that it applies to names and not contents. Whereas in other context signs are merely representatives of their content, so that every combination into which they enter expresses only a relation between their respective contents, they suddenly display their own selves when they are combined by means of the sign for identity of content; for it expresses the circumstance that two names have the same content" (Frege (1967), p. 20). However, in his "Über Sinn und Bedeutung," he argues that identity relation cannot be held between *arbitrarily chosen signs*. For this reason, Frege comes up with a new solution: Meaning has two components. These are *sense* and *reference*. Reference is the semantic value, and sense is the cognitive value of an expression. In other words, reference is the object denoted, the sense is the mode of presentation. He analyses the identity relation and the cognitive difference at two different levels. Identity is a relation holding between the object to itself denoted by the terms flanking identity sign. Consider

Frege's example from geometry. There are three different lines a, b, and c. The intersection point of lines a and b is identical with the intersection point of lines b and c. In this case, consider the following two identity sentences below.

- The intersection point of the lines a and b is the intersection point of the lines a and b.
- The intersection point of the lines a and b is the intersection point of the lines b and c.

The cognitive difference between the two identity sentences above arises due to the difference in the modes of presentation of the denoted objects. While, in the first sentence above, the denoted objects and the modes of presentations are the same; in the second sentence, modes of presentations are different. Therefore, the identity relation holds at the reference level, and the cognitive difference holds at the sense level (Frege (1948), pp. 209-210). Frege's distinction between sense and reference gives us two different understandings of compositionality. The compositionality principle can be applied differently to sense and reference. Composing the references of the parts of a sentence, we obtain the truth value as the reference of the sentence. Mathematics is a good example of this case. Composing the reference of the parts in a mathematical equation, we end up with either truth or falsity. Consider the sentence "7+3=10." "7+3" refers to an object, and the expression "=10" refers to a concept. By composing these two, we get a truth value. From a Fregean perspective, we can say that the sentence refers to the True. The composition of the senses of the constituents of a sentence results in the sense of the sentence, *the proposition, or the thought*.

In his later works, he emphasizes the psycholinguistic aspects of the principle of compositionality. He is interested in creativity in language use. For him, without the role of the principle of compositionality in our cognitive system, this would not be possible.

> It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a human being for the very first time can be put into a form of words which will be understood

by someone to whom the thought is entirely new. This would be impossible, were we not able to distinguish parts in the thought corresponding to the part of a sentence, so that the structure of the sentence serves as an image of the structure of the thought... If, then, we look upon thoughts as composed of simple parts, and take these, in turn, to correspond to the simple parts of sentences, we can understand how a few parts of sentences can go to make up a great multitude of sentence to which, in turn, these correspond a great multitude of thoughts. But the question now arises how the thought comes to be constructed, and how its parts are so combined together that the whole amounts to something more than the parts take separately. (Frege (1963), p. 1)

2.2.3 Bertrand Russell

Bertrand Russell is another philosopher who is interested in the logical, mathematical structure of language. For he too thinks that this structure is compositional. The quote below clearly shows that he is an advocate of the compositionality principle.

> You can understand a proposition when you understand the words of which it is composed even though you never heard the proposition before. That seems a very humble property, but it is a property which marks it as complex and distinguishes it from words whose meaning is simple. When you know the vocabulary, grammar, and syntax of language, you can understand a proposition in that language even though you never saw it before. In reading a newspaper, for example, you become aware of a number of statements which are new to you, and they are intelligible to you immediately, in spite of the fact that they are new, because you understand the words of which they are composed. This characteristic, that you can understand a proposition through the understanding of its component words, is absent from the component words when those words express something simple. Take the word “red," for example, and suppose - as one always has to do - that “red" stands for a particular shade of colour. You will pardon that assumption, but one never can get on otherwise. You cannot understand the meaning of the word “red" except through seeing red things." (Russell (2010), p.20)

Just as Frege, he believes that mathematical truths can be justified on logical grounds alone. For this reason, he uses Frege’s quantifier logic. Unlike Frege, he believes that quantifier logic can help us to get rid of the metaphysics that he thinks arises from the surface syntax (Russell (1905)). However, this approach deviates from the compositional semantics. It is because compositionality in its standard form is a relation between surface syntax and semantics. However, for Russell, the surface syntax may be deceptive. Some linguistic entities that we call proper names at the surface are *actually* disguised definite descriptions (Russell (1905)). There is an epistemological background here. For him, there are two kinds of knowledge: knowledge by acquaintance and knowledge by description (Russell (1910)). Semantically, we can only talk about the object we are directly acquainted with (e.g., a is red). On the other hand, we cannot talk *about* anything we are not acquainted with. Instead, we can only make descriptions (e.g., *There is an object that is red.*). According to this approach, the meaning of “the author of Metaphysics is wise" is not composed of meanings of its surface constituents, “the author of Metaphysics" and “is wise." In other words, the sentence does not mean that *wise* is predicated of *the author of Metaphysics*. Instead, it says something general: $\exists x((M(x) \wedge \forall y(M(x) \rightarrow y = x)) \wedge W(x))$, where M stands for the author of Metaphysics and W stands for being wise (For the compositional version of Russell’s theory of description see Neale (1990)).

2.2.4 Ludwig Wittgenstein

Ludwig Wittgenstein, in *Tractatus Logico-Philosophicus*, seeks to understand the meaning relation between language and the world. Wittgenstein’s aim is more metaphysical than the Frege - Russell project. He is neither interested in developing a representational tool for all sciences nor in providing a logical justification of mathematics. His only interest is the nature of the relation between language and the world.

To show the relation between language and the world, Wittgenstein starts with the metaphysical structure of the world. In his picture, building blocks of the world are not objects but facts. He says all possible reality is built on atomic facts. “An atomic fact is a combination of objects (entities, things)." (Wittgenstein (1922), §2.01). So, reality consists of atomic facts, which are combinations of objects. I think this is more or less the common sense picture of reality. On the other side of the meaning relation, we have language and thought. The meaning relation between language and the world in *Tractatus* is an isomorphic one as Wittgenstein says that “the logical picture of the facts is the thought." (Wittgenstein (1922), §3)

All complex statements can be decomposed into elementary propositions. “The simplest proposition, the elementary proposition, asserts the existence of an atomic fact" (Wittgenstein (1922), §4.21). Elementary propositions consist of names. Wittgenstein’s project is essentially compositional since he doesn’t admit anything other than atomic facts in his metaphysics. There are no molecular facts out there. Instead, all the descriptions can be decomposed into atomic facts. In several sections, he confirms the compositionality principle: “I conceive the proposition –like Frege and Russell– as a function of the expressions contained in it." (Wittgenstein (1922), §3.318)

“A proposition about a complex stands in internal relation to the proposition about its constituent part." (Wittgenstein (1922), §3.24)

“The names are the simple symbols, I indicate them by single letters (x, y, z). The elementary proposition I write as function of the names, in the form fx, $\phi(x, y)$, etc."(Wittgenstein (1922), §4.24)

2.2.5 Donald Davidson

Donald Davidson, in several articles, Davidson (2001c), Davidson (1967), Davidson (2001b), and Davidson (2001a), is interested in the learnability of languages. For him, the learnability of a language is based on its productivity and systematicity. Otherwise, we have to memorize every single expression. For example, since Frege’s treatment of oblique contexts is not productive and systematic, as we discussed above, his language theory fails to meet the learnability criteria (Davidson (2001c), pp. 14-16).

> These matters appear to be connected in the following informal way with the possibility of learning a language. When we can regard the meaning of each sentence as a function of a finite number of features of the sentence, we have an insight not only into what there is to be learned; we also understand how an infinite aptitude can be encompassed by finite accomplishments. For suppose that a language lacks this feature; then no matter how many sentences a would-be speaker learns to produce and understand, there will remain others whose meanings are not given by the rules already mastered. It is natural to say such a language is *unlearnable*. (Davidson (2001c), p. 8)

In his seminal paper, “Truth and Meaning" he asks “what it is for a theory to give an account" of the fact that “on mastering a finite vocabulary and a finitely stated set of rules, we are prepared to produce and to understand any of a potential infinitude of sentences" (Davidson (1967), p. 304) For him, “a theory of meaning for a language L shows “how the meanings of sentences depend upon the meanings of words" if it contains a (recursive) definition of truth-in-L" (Davidson (1967), p.310). To provide such a theory, Davidson appeals to Alfred Tarski's truth schema:

$$S\ is\ true - in - L\ if\ and\ only\ if\ p$$

where p is the translation of the sentence S in the meta-language. Tarski aims to provide a formal truth theory, not a meaning theory. For that reason, in Tarski's system, the meaning of a sentence is already given. On the contrary, Davidson aims to introduce a theory of meaning. While for Tarski, the truth cannot be defined for natural languages, Davidson takes the truth as the essential part of communication, natural language understanding and language acquisition.

> The clues used by a learner of a first language and the data consciously sought by the field linguist are just what is needed to confirm, inductively, that a language in use is correctly described by a theory of truth based on Tarski's methodology. The concept of truth palys the leading role throughout. Ostensive learning, broadly conceived, depends either on the attempt on the part of teacher or informant to say what is true, or on the ability of the learner to detect when a speaker is saying what he or she holds to be literally true. Naturally, what is held to be true is not necessarily true. But the learner must assume in the case of ostension that what is held to be true *is* true until enough of the relations among sentences are in place to justify treating some ostensions as false" (Davidson (2005b), pp. 162-163)

2.2.6 Richard Montague

Richard Montague is one of the pioneers in mathematical semantics. His work in the 1970s is one of the motivations behind our attempt to understand the mathematical structure of natural languages. His papers aim to provide a mechanical procedure for meaning in natural languages.

> There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory. (Montague (1974b), p. 222) (Montague (1974a), p. 188)

With this aim, Montague provides a mechanical procedure to compute the meaning of any expression in English. The procedure is conveyed compositionally. In Montague semantics, a fragment of English is divided into syntactic categories. For each of these categories, a semantic category is assigned. Semantic categories are of sets of possible denotation functions:

Table 2.1: Semantic and syntactic categories for a portion of English in Montague semantics

	Syntactic category	Semantic category
0	the set of basic name phrases (proper names with variables)	the set of possible individuals
1	the set of basic formulas	the set of functions with **domain**: possible worlds and **range**: truth values 0, 1
2	the set of basic intransitive verb phrases	the set of functions with **domain**: the set of possible individuals and **range**: the set of functions from possible worlds to truth values 0, 1
3	the set of basic transitive verb phrases	the set of functions with **domain**: the cartesian product of the set of possible individuals and **range**: the set of functions from the set of functions from possible worlds to truth values 0, 1
4	the set of basic common noun phrases	the set of functions with **domain**: the set of possible individuals and **range**: the set of functions from possible worlds to truth values 0, 1
5	the set of basic adformula phrases	the set of function with **domain**: the set of functions from possible words to truth value 0,1 and **range**: the set of functions from possible worlds to truth value 0,1
6	the set of basic ad-one-verb phrases	the set of function with **domain**: the set of functions with domain: the set of possible individuals and range: the set of functions from possible worlds to truth values 0, 1 and **range** : the set of functions with domain: the set of possible individuals and range: the set of functions from possible worlds to truth values 0, 1

Table 2.1: Continued

	Syntactic category	Semantic category
7	the set of basic ad-two-verb phrases	the set of function with **domain**: the set of functions with domain: the cartesian product of the set of possible individuals and range: the set of functions from the set of functions from possible worlds to truth values 0, 1 and **range**: the set of functions with domain: the cartesian product of the set of possible individuals and range: the set of functions from the set of functions from possible worlds to truth values 0, 1
8	the set of basic adjective phrases	the set of functions with **domain** : the set of functions with domain: the set of possible individuals and range: the set of functions from possible worlds to truth values 0, 1 and **range**: the set of functions with domain: the set of possible individuals and range: the set of functions from possible worlds to truth values 0, 1

After providing a semantic category for each syntactic category, Montague lists a semantic combination rule for each syntactic combination rule.

Table 2.2: A semantic rule is assigned for each syntactic rule in Montague semantics

	Syntactic operations	Semantic operations
S1	Basic syntactic categories are subsets of full syntactic categories	
S2	if δ is an intransitive verb phrase $\wedge$ α is a name phrase then $\ulcorner\delta\alpha\urcorner$ is a well formed formula	$F_2(d,a)$ such that $F(d,a)(x) = d(x)(a(x))$

Table 2.2: Continued

	Syntactic operations	Semantic operations
S3	δ is a transitive verb phrase $\wedge$ α, β are name phrases then $\ulcorner \delta\alpha\beta \urcorner$ is a well formed formula	$F_3(d,a,b)$ such that $F(d,a,b)(x) = d(x)(a(x), b(x))$
S4	if δ is an adformula phrase $\wedge$ ϕ is a formula then $\ulcorner \delta\phi \urcorner$ is a well formed formula	$F_4(d,p) = F_2(d,p)$
S5	if δ is a basic ad-one-verb phrase $\wedge\mu$ is an intransitive verb phrase then $\ulcorner \delta\mu \urcorner$ is an intransitive phrase.	$F_5(d,m) = F_2(d,m)$
S6	if δ is a ad-two-verb phrase $\wedge\mu$ is a transitive verb phase then $\ulcorner \delta\mu \urcorner$ is a transitive verb phrase	$F_6(d,m) = F_2(d,m)$
S7	if δ is an adjective phrase $\wedge\zeta$ is a common noun phase then $\ulcorner \delta\zeta \urcorner$ is a common noun phrase	$F_7(d,z) = F_2(d,z)$
S8	if ϕ is a well formed formula $\wedge\alpha, \beta$ are proper name phrases then ψ is a well formed formula such that ψ is obtained by replacing free occurrence of α in ϕ by β	$F_8(p,a,b)$ such that $F_8(p,a,b)(x) = p(x)^a_{b(x)}$
S9	if ϕ is a well formed formula $\wedge\alpha$ is a proper name phase $\wedge\zeta$ is a common noun phrase then ψ is a well formed formula such that ψ is obtained by replacing free occurrence of α in ϕ by $\ulcorner every\zeta \urcorner$	$F_9(p,a,z)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds $iF_9(p,a,z)(x)(i) = 1$ if and only if $p(x^a_t)(i) = 1$ "whenever t is a member of A for which $z(x)(t)(i) = 1$"
S10	if ϕ is a well formed formula $\wedge\alpha$ is a proper name phase $\wedge\zeta$ is a common noun phrase then ψ is a well formed formula such that ψ is obtained by replacing free occurrence of α in ϕ by $\ulcorner a\zeta \urcorner$ and all other occurrence of α by $\ulcorner that\eta \urcorner$ where η is a basic common noun phrase	$F_{10}(p,a,z)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds i $F_{10}(p,a,z)(x)(i) = 1$ if and only if $p(x^a_t)(i) = 1$ "for some $t \in A$ such that $z(x)(t)(i) = 1$"

Table 2.2: Continued

	Syntactic operations	Semantic operations
S11	if ϕ is a well formed formula $\wedge\alpha$ is a proper name phase $\wedge\zeta$ is a common noun phrase then ψ is a well formed formula such that ψ is obtained by replacing free occurrence of α in ϕ by $\ulcorner the\zeta \urcorner$ and all other occurrence of α by $\ulcorner that\eta \urcorner$ where η is a basic common noun phrase	$F_{11}(p, a, z)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds i $F_{11}(p, a, z)(x)(i) = 1$ if and only if there exists $t \in A$ such that t is the set of objects $u \in A$ for which $z(x)(u)(i) = 1$," and $f(x^a_t)(i) = 1$.
S12	if ζ is a common noun phrase $\wedge\alpha, \beta$ is a proper name phase then ψ is a common noun phrase such that ψ is obtained by replacing free occurrence of α in ϕ by β	$F_{12}(z, a, b)$ such that for all sets of infinite sequences of the members of the universe $F_{12}(z, a, b)(x) = z(x^a_{b(x)})$
S13	if $\zeta\eta$ is a common noun phrase $\wedge\alpha$ is a proper name phase then ψ is a common noun phrase such that ψ is obtained by replacing free occurrence of α in ϕ by $\ulcorner every\zeta \urcorner$	$F_{13}(e, a, x)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds i $F_{13}(e, a, x)(x)(i) = 1$ if and only if $e(x^a_t)(i) = 1$ "whenever t is a member of A for which $z(x)(t)(i) = 1$"
S14	if $\zeta\eta$ is a common noun phrase $\wedge\alpha$ is a proper name phase then ψ is a common noun phrase such that ψ is obtained by replacing free occurrence of α in ϕ by $\ulcorner a\zeta \urcorner$ and all other occurrence of α by $\ulcorner that\eta \urcorner$ where η is a basic common noun phrase	$F_{14}(e, a, z)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds $iF_{14}(e, a, z)(x)(i) = 1$ if and only if $e(x^a_t)(i) = 1$ "for some $t \in A$ such that z(x)(t)(i) = 1"
S15	if $\zeta\eta$ is a common noun phrase $\wedge\alpha$ is a proper name phase then ψ is a common noun phrase such that ψ is obtained by replacing free occurrence of α in ϕ by $\ulcorner the\zeta \urcorner$ and all other occurrence of α by $\ulcorner that\eta \urcorner$ where η is a basic common noun phrase	$F_{15}(e, a, z)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds i $F_{15}(e, a, z)(x)(i) = 1$ if and only if there exists $t \in A$ such that t is the set of objects $v \in A$ for which $z(x)(v)(i) = 1$, and $e(x^a_t)(u)(i) = 1$

Table 2.2: Continued.

	Syntactic operations	Semantic operations
S16	if ζ is a common noun phrase $\wedge\alpha$ is a formula then η is a common noun phrase such that η is $\ulcorner\zeta such that\psi$ where ψ is obtained by replacing all free occurrences of α in ϕ by $\ulcorner that\theta\urcorner$ where θ is a basic common noun that occurs first in ζ	$F_{16}(z, p, a)$ such that for all sets of infinite sequences of the members of the universe x, for all members of the universe t and for all possible worlds i $F_{16}(z, p, a)(x)(i) = 1$ if and only if $p(x_t^a)(i) = 1$
S17	if α is a proper name phrase $\wedge\zeta$ is an adjective phase then $\ulcorner\alpha is\zeta\urcorner$ is a well formed formula	$F_{17}(a, d)$ such that for all sets of infinite sequences of the members of the universe and for all possible worlds i $F_{10}(a, d)(x)(i) = 1$ "if and only if d(x)(z(x))(a(x)) (i) = 1, where z is that function such that, for all infinite object sequences x, all objects t, and all possible worlds i z(x)(t)(i) = 1"

The aim of this rule by rule approach is to compute the meaning of any given expression. As it is seen, semantic rules consist of function application and modal logic. (Montague (1974a)) (Montague (1974b))

2.2.7 Jerrold Katz and Jerry Fodor

Jerry A. Fodor and Jerrold J. Katz's seminal paper Katz and Fodor (1963) highlights the psycholinguistic aspect of the compositionality principle. Their aim is to provide a psycholinguistic mechanism behind the productivity and the systematicity of meaning in natural languages.

> A fluent speaker's mastery of his language exhibits itself in his ability to produce and understand the sentences of his language, including indefinitely many that are wholly novel to him (i.e., his ability to produce and understand any sentence of his language). The emphasis upon novel sentences is important. The most characteristic feature of language is its ability to make available an infinity of sentences from which the speaker can select appropriate and novel ones to use as the need arises (Katz and Fodor (1963), p. 171).

We do not memorize the meaning of every single sentence we hear. Instead, we capture the rules behind the meaning of the sentences we hear.

> That is to say, what qualifies one as a fluent speaker is not the ability to imitate previously heard sentences but rather the ability to produce and understand sentences never before encountered. The striking fact about the use of language is the absence of repetition: almost every sentence uttered is uttered for the first time. This can be substantiated by checking texts for the number of times a sentence is repeated. It is exceedingly unlikely that even a single repetition of a sentence of reasonable length will be encountered (Katz and Fodor (1963), p. 171).

The point here is the systematicity and the productivity of meaning in natural languages. These two properties retain their importance in the contemporary work in semantics. Productivity means that with a finite set of meanings, we can compute meanings of an infinite number of sentences (Szabó (2012), p. 75)(Pagin and Westerståhl (2010b), pp. 266-267). Anyone who knows English understands the sentence, "Two weeks ago in Pyongyang Albert Einstein and Britney Spears had dinner," which I believe nobody has ever heard before reading this paper. Similarly, a language is systematic if "a lexical item must make approximately the same semantic contribution to each expression in which it occurs. It is, for example, only in so far as 'the,' 'girl,' 'loves' and 'John' make the same semantic contribution to 'John loves the girl' that they make to 'the girl loves John' that understanding the one sentence implies understanding the other" (Fodor and Pylyshyn (1995), p. 124)(Pagin and Westerståhl (2010b), p. 267).

For Katz and Fodor, to explain the systematicity and productivity of the meaning in natural languages, our semantic theory must obey the principle of compositionality. Indeed, this is the prevalent view in formal semantics. (Szabó (2012), p. 72) (Szabó (2000), p. 477)

> Since the set of sentences is infinite and each sentence is a different concatenation of morphemes, the fact that a speaker can understand any sentence must mean that the way he understands sentences which he has never previously encountered is compositional: on the basis of his knowledge of the grammatical properties and the meaning of the morphemes of the language, the rules which the speaker knows enable him to determine the meaning of a novel sentence in terms of the manner in which the parts of the sentence are composed to form the whole. Correspondingly,

> we can expect that a system of rules which solves the projection problem must reflect the compositional character of the speaker's linguistic knowledge (Katz and Fodor (1963), pp. 171-172).

2.3 Different interpretations of the principle of compositionality

Collective (Fine 2007) versus individual (Szabó (2012), p. 70) interpretations of the constituent terms, correct understanding of the compositional function, and the nature of the combination cause different interpretations of the compositionality principle (Szabó (2012), pp. 67-70). Zoltán Gendler Szabó formulates the compositionality principle as a strong supervenience claim: "For all possible languages L, for any meaning property M and any complex expression e in L, if e has M in L, then there is a constitution property C such that e has C in L, and for any possible language L' if any complex expression e' in L' has C in L' then e' has M in L'" (Szabó (2000), p. 499). Similarly, based on the syntax-semantics relation, there are different versions of this principle. For the direct compositionality, the "interpretation is [not] 'postponed' until a later stage in the grammatical computation," while according to indirect versions, it is postponed due to hidden syntactic variables (Barker and Jacobson (2007), pp. 1-2).

The empirical significance of the principle gives rise to different epistemological evaluations of the principle (Szabó (2012), p. 189) (Dowty (2007), pp. 25-27). According to Wlodek Zadrozny, compositionality is a trivial principle. Any semantics can be made compositional with an additional semantic function, μ. Through the use of ZFA set theory, Zadrovzny proves that for any language S, meaning function m, any expressions s and t, an end of expression symbol \$ and a concatenation symbol ., there is a function μ such that

$$\mu(s.t) = \mu(s)(\mu(t))$$

$$\mu(s.\$) = m(s).$$

In other words, the meaning of any complex expression $s.t$ is reduced to the meanings of s and t (Zadrozny (1992), p. 261) (Zadrozny (1994), p. 333) (Zadrozny (1994), pp. 332-334) (Zadrozny (1992), pp. 261-262). For Zadrozny, what makes compositionality robust is the systematicity (Zadrozny (1994), p. 329) (Zadrozny (1992), pp. 263-264). For others, a realistic account of lexical meaning and synonymity saves compositionality from being trivial (Dever (1999), pp. 321-323) (Kazmi and Pelletier (1998), pp. 632-633).

Another controversial topic is the relationship between compositionality and the theories of meaning. For example, Davidson (1967) proposes the truth conditional theory of meaning as an entailment of the compositional approach. On the contrary, Paul Horwich argues that compositionality is neutral to the theories of meaning (Horwich (1997), p. 531) (See Heck (2013) for an argument against Horwich's deflationary position). In this work, I try to be neutral to these controversies on compositionality as long as possible.

2.4 Problems with compositionality

Some cases pose problems for the compositionality principle. These are idioms, intonation, context-dependency of quantifiers, ellipsis, deixis, structural ambiguity, lexical ambiguity, etc. (Goldberg (2016), pp. 420-427). Among these problems, I propose lexical ambiguity to be the most challenging. In lexical ambiguity, a word is associated with more than one meaning. The problem is that while lexical disambiguation is a top-down process (Goldberg (2016), p. 429), compositional semantics is a bottom-up one ((Szabó (2017)) (Goldberg (2016), pp.419-420) (Recanati (2004), p.2) (Davies (2003), p.92). Let us consider sentences "Mary walked along the bank of the river" and "HarborBank is the richest bank in the city" (Pustejovsky (1996), p. 27). According to the compositional approach, to compute the meaning of this expression, the meaning of "bank" is needed. However, there may be several entries for "bank" in a lexicon: *a financial institution* or *the shore* (Pustejovsky (1996), p. 34). Thus, the meaning of "bank" depends on the context in which it occurs. Interestingly, lexical ambiguity is a widespread phenomenon. As Thomas Wasow claims, "even function words are often ambiguous" (Wasow et al. (2005), pp. 268).

For this reason, the forerunners of the compositional semantics, Katz and Fodor, say, "the comparison between a fluent speaker and a machine reveals the respects in which a grammar and dictionary by themselves do not suffice to interpret sentences like a speaker of the language ... Thus, a semantic theory of a natural language must have such rules (which we shall call ‘projection rules’) as one of its components if it is to match the speaker’s interpretation of sentences" (Katz and Fodor (1963), p. 183). A contemporary defender of this approach, Emma Borg (Borg (2004), p.19), lists several solutions for the lexical ambiguity. “An ambiguous string either has all its possible readings computed by the language faculty, leaving a choice between rival interpretations to post-linguistic procedures (e.g., by the agent’s general intelligence), or the string has already been disambiguated before it reaches the language faculty" (Borg (2004), p. 142). Additionally, within the language faculty, expressions can be disambiguated through the interaction between different linguistic levels. For example, in “She saw him duck", the *animal* sense of “duck" is eliminated through syntactic restrictions. To summarize, according to the building block approach, at three different stages: pre-linguistic, post-linguistic, and within linguistic, expressions can be disambiguated (Borg (2004) pp. 142-144) (Gasparri (2014), pp.153-161).

2.5 The polysemy

Nevertheless, there is a subtype of lexical ambiguity that deserves better treatment than the simple two-stage disambiguation methods: *polysemy*. To begin with, there are two kinds of lexical ambiguities depending on the relationship between the different senses of a word. If the senses are not related, it is called homonymy; if they are *related*, it is called polysemy (Manning et al. (1999), p. 110) (Recanati (2017), p. 384) (Saeed (2009), pp.63-64). Uriel Weinreich classifies polysemy as complementary ambiguity, and homonymy as contrastive ambiguity. He claims that a dictionary should represent the distinction between two categories (Weinreich and Webster (1964), p.406) (Pustejovsky (1996), pp. 27-33). To distinguish these categories, several formal tests were proposed, i.e., logical, linguistic, and definitional tests (Ravin and Leacock (2000), pp. 3-4) (Quine (2013), p. 119). Even though these tests are not coherent with each other (Geeraerts and Cuyckens (2007), pp. 141-144), the difference is evident in language use. For example, “çay" in Turkish has two homonyms which mean in English a creek and tea respectively. Since there is no relation between a creek and tea, it is relatively easy to understand how language users decide whether *çay* means tea or a creek in a given context. However, when it comes to the related senses of *çay*, things get complicated. Let us consider one of the homonyms, which means tea in English. In this case, *çay* has many different but related senses: *a tea plant, a tea leaf, a hot beverage, the liquid itself, tea party*, etc. When somebody orders tea in a cafe, the waiter understands “tea" to mean *a hot beverage*. The context helps him to specify the meaning, but both speaker and the hearer would be aware of the other related senses (Recanati (2004), p. 30)(Frisson (2009), p. 115) (Pustejovsky (1996), p. 32). However, the other homonym, i.e., which means creek, “that is not selected quickly decays" (Vicente (2018), p. 953) (Falkum and Vicente (2015), p. 4).

Compositional semanticians' *pre-linguistic, post-linguistic*, and *linguistic* disambiguation methods come with different ways of designing a lexicon and combination rules that aim to be compatible with the rich structure of polysemy. One way is Katz and Fodor's sense enumeration lexicon (Katz and Fodor (1963), pp. 191) (Falkum and Vicente (2015), p. 3) (Pustejovsky (1996), p. 34). To deal with polysemy better, polysemous items are grouped together (Montague (1974b), pp. 209-210) (Dowty (1991), pp.60-62) (Pustejovsky (1996), pp. 37-38).

Yet, to list every nuance of a meaning in the lexicon is thought to be unrealistic (Pustejovsky (1996), p. 127). The solution is to abandon the undiscriminating approach and to adopt “the one representation hypothesis" according to which lexical entries played the role of initial gateways to reach the occurrent meaning (Falkum and Vicente (2015), p. 5). There are two different ways to design a lexicon in tandem with this view: the overspecified and the underspecified approach. According to the underspecified view, word types in the lexicon have more *schematic, underspecified* meanings (Frisson (2009), pp. 116-118) (Frisson (2015), p. 30) (Harris (1994), pp. 208-209). Robyn Carston's pragmatics oriented “meaning-relevant components" is a good example of an underspecified lexical entry (Carston (2012), p. 17). In the next two sections, I will critize both rule-based and the pragmatics-oriented approaches. Then, I will propose an alternative approach.

CHAPTER 3

THE RULE-BASED APPROACH

3.1 The rule-based polysemy resolution

According to the rule-based approach, word meanings are so powerful that with the minimal contextual involvement they allow a linguistic system to understand any expression. One version of this approach is *Semantic Minimalism.* "The idea motivating Semantic Minimalism is simple and obvious: The semantic content of a sentence S is the content that all utterances of S share... Semantic Minimalism recognizes a small subset of expressions that interact with contexts of utterance in privileged ways; we call these the *genuinely context sensitive expressions*" (Cappelen and Lepore (2005), p. 143). Genuinely context sensitive expressions consist of indexicals (e.g., personal pronouns, demonstratives, etc.). According to Indexicalism, even the context dependency of the meanings of content words (e.g., small, green,... etc.) is governed by the logical form of the expression (Stanley (2000)) (Szabó (2001)).

From the computational perspective, James Pustejovsky's generative lexicon theory exemplifies this approach. Generative lexicon theory assigns an argument structure, an event structure, and a qualia structure, which consists of four Aristotelian components *a constitutive, formal, telic, and agentive aspect* to each lexical item. The meaning of any word in context is determined by three linguistic operations, *type coercion, co-composition, and selective binding* (Pustejovsky (1996), pp. 61-62). For example, as show in Figure 1, in interpreting "a long record" in "John bought a long record" as a record whose playing time is long, the adjective "long" modifies "the event description in the TELIC quale of the noun," "record": $\lambda x[...Telic = \lambda e[play'(x)(e) \wedge long(e)]...]$ (Pustejovsky (1996), pp. 128-130).

$$\begin{bmatrix} \textbf{record} & \\ \text{ARGSTR} = & \begin{bmatrix}\text{ARG1} = \textbf{x:physobj}\\ \text{ARG2} = \textbf{y:info}\end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix}\textbf{info.physobj\_lcp}\\ \text{FORMAL} = \textbf{R(x,y)}\\ \text{TELIC} = \textbf{play(e,x.y)}\end{bmatrix}\end{bmatrix}$$

Figure 3.1: An overspecified lexical entry for the noun *record*

· operator is used to construct polysemous *dot objects*. For example, “book" is represented as a dot object **physobj.info**. (Pustejovsky (1998), pp. 298-299)

$$\begin{bmatrix} \textbf{book} & \\ \text{ARGSTR} = & \begin{bmatrix}\text{ARG1} = \textbf{y:information}\\ \text{ARG2} = \textbf{x:physobj}\end{bmatrix} \\ \text{QUALIA} = & \begin{bmatrix}\textbf{info.physobj\_lcp}\\ \text{FORMAL} = \textbf{hold(x,y)}\\ \text{TELIC} = \textbf{read(e, w, x.y)}\\ \text{AGENT} = \textbf{write(e', v, x.y)}\end{bmatrix}\end{bmatrix}$$

Variables e and e' quantifies over events. Since *reading* and *writing* are different events, different letters are used.

Drawing an analogy to the formal language theory, Pustejovsky distinguishes three different levels of semantic expressivity: Monomorphic languages (e.g., enumerating each sense separately), weakly polymorphic languages (overspecified lexicon approach) and unrestricted polymorphic languages (underspecified lexicon approach). For Pustejovsky, natural languages are weakly polymorphic languages. To Pustejovsky, the traditional *sense enumeration model* is weakly compositional since it requires an infinite list of senses. On the other hand, his generative lexicon theory is strongly compositional as it is a “much more adequate model for cognitive concerns and computational tractability, while still preserving the principle of compositionality" (Pustejovsky (1996), p. 60). In other words, in a natural language, “all lexical items are semantically active, and have a richer typed semantic representation than conventionally assumed" (Pustejovsky (1996), p. 58). On the other hand, in the unrestricted polymorphic languages, “there is nothing inherent in the language that constrains the meaning of the words in context" (Pustejovsky (1996), p. 56).

For the rule-based view, disambiguation is done through selecting a part of the overspecified representation using the linguistic context (Vicente (2018), pp. 952-954) (Elman (2009), pp. 567-568). On the contrary, as we will see in Chapter 4, for the pragmatics-oriented view, for which natural languages are more like unrestricted polymorphic languages, disambiguation is done through pragmatic meaning enrichment processes. In other words, the overspecified view is more linguistic oriented and rule-based, while the underspecified view is more pragmatic, thus contextual.

In the end, Fodor’s statement below summarizes the place of polysemy within the rule-based approach.

> the amount of context-induced variation of lexical meaning is often overestimated because other sorts of context sensitivity are misconstrued as violations of compositionality. For example, the difference between ‘feed the chicken’ and ‘chicken to eat’ must involve an *animal/food* ambiguity in ‘chicken’ rather than a violation of compositionality since if the context ‘feed the ...’ could *induce* (rather than select) the meaning *animal*, you would expect ‘feed the veal,’ ‘feed the pork’ and the like. Similarly, the difference between ‘good book,’ ‘good rest’ and ‘good fight’ is probably not meaning shift but syncategorematicity. ‘Good *NP*’ means something like *NP that answers to the relevant interest in NPs:* a good book is one that answers to our interest in books (viz. it’s good to read); a good rest is one that answers to our interest in rest (viz. it leaves one refreshed); a good fight is one that answers to our interest in fights (viz. it’s fun to watch or to be in, or it clears the air); and so on. It’s because the meaning of ‘good’ is syncategorematic and has a variable in it for relevant interests, that you can know a good flurg is a flurg that answers to the relevant interest in flurgs is. (Fodor and Pylyshyn (1995), pp. 124-125)

I believe that polysemy is not *overestimated*. Indeed, it is *underestimated*. On the contrary, the problem with the “rule-based approach is its lack of interpretive flexibility" (Falkum (2015), p. 87). I agree with Jonathan Cohen that “a semantics that ignores such polysemy thereby abandons any claim to be a semantics for natural language, because it ignores one of the main feature –a feature comparable with indexicality– that differentiates a natural language from a Tarskian artificial one" (Cohen (1985), p. 132). In the next section, I will use arguments from philosophers and cognitive scientists to show how polysemy is essential in our conceptual, cognitive, and communicative faculties.

3.2 Arguments against the rule based approach

In this section, I will lay out three arguments against the rule-based approach. The first one is the intractability of concepts that words allegedly refer to. The second one is the intertwinedness between language and conceptual space. The third argument is that intractability of conceptual space is not a defect but a desired feature of human language.

3.2.1 Concepts are intractable

Investigating concepts, in the first place, is philosophers' job. However, history of philosophy leads us to think that philosophers are not successful at this job. More than two and a half century philosophers are asking questions that targets concepts: What is knowledge? What is science? What is truth? What is goodness? What is concept?... etc.

As I said above, philosophy, beginning with Socrates, is interested in providing a general definition of various terms such as “knowledge, goodness, time" etc. (Wittgenstein (1965), pp. 20,26). Socrates' role in Plato's dialogues is to show that proposed definitions of concepts, such as courage (*Laches*), piety (*Euthyphro*), virtue (*Meno*), love (*Symposium*), fail. As part of *the Elenctic method*, Socrates, by providing contexts where the terms above are used with different meanings, leaves his opponent in *aporia* (Fine (1992), p. 205). For example, in *Laches*, Socrates moves from more prototypical uses of “bravery" (e.g., “fighting a human enemy while remaining at one's post without fleeing") to less prototypical ones (e.g., “endurance of distressful situations (illness, poverty), situations that are no doubt to be feared"). Then, more metaphorical examples follow. In different parts of the text, "bravery" is associated with *sophrosyne* (temperance) and *sophia* (wisdom) (Rademaker (2005), p. 307). In his analysis of the ancient virtue *sophrosyne*, Adriaan Rademaker concludes that "Plato uses the polysemy of *sophrosyne* in argumentative passages to establish links with several other virtues" (Rademaker (2005), p. 322). The complexity of the conceptual network engenders the difficulty of providing clear meanings of these words.

Plato's philosophy is a response to the Socratic crisis. Alongside *flux of appearance, ontological problems with concepts in general*, the polysemous nature of language plays a key role in postulating *Platonic, ideal objects* (*Phadeo*, 102a-b) (*Lysis*, 219c-220b).

Aristotle, who is aware of Plato's response to the Socratic crisis (*Metaphysics*, $A6$, 987b1), developed a philosophy in which scientific research is interested in the primary, core sense of *words.* For him, *being* is polysemous in the same way *healthy* is polysemous. One single investigation, science should work on the nature of the core sense of *being* (*Metaphysics, Book* $\Gamma 2$, 1003b).

A sociological impact of the crisis in finding definitions of concepts is seen with appearance of the ordinary language philosophy movement. Advocates of this movement, including Ludwig Wittgenstein, Gilbert Ryle, and John Langshaw Austin, thought that the source of the crisis is the polysemous nature of language which plays a key role in our conceptualizations, metaphysics, and ontologies. In the *Blue Book*, Wittgenstein's opening question is "What is the meaning of a word" (Wittgenstein (1965), p. 1). How does a sign get its meaning? Similarly, Austin asks "Why do we call different things by the same name?" (Austin (1961), p. 37) The rules that govern our use of a word is what Wittgenstein calls *grammar*. The philosophical puzzles arise from misunderstanding *grammar* (Wittgenstein (1965), p. 27). Philosophically essential words such as "wishing," "thinking," "time," "length," "knowledge," "exist," "fascist," "pleasure" etc. may not be defined by one unique set of features (Wittgenstein (1965), pp.19-20) (Austin (1961), pp. 40,41). "The phrase "expecting that B will come" may not be a value of the function "expecting that x will come" (Wittgenstein (1965), 21). It means that there may not be a unique λx(expecting that x will come). For Ryle, "exist" in "there exists a cathedral in Oxford" and "there exists a square number between

9 and 25" have different logical powers (Ryle (2009), pp. 215-216). (See also Carnap (1950), and Hirsch (2002)). However, we, human beings, are “inclined to think that" there is a unique set of features that corresponds to each term (Wittgenstein (1965), p. 17). As Austin remarks, simple definitions to represent polysemous words “make[s] hashes of things" (Austin (1961), p. 42) (Ryle (2009), p. 216) (Wittgenstein (1965), pp. 17, 27). (See also Quine (2013), pp. 118, 120-121).

For Wittgenstein, the naive objectivist view toward the relationship between a word and an object, to which he too contributed in his earlier period, results from the ignorance of the essentiality of polysemy in human thought and language. The relatedness of different senses of “game" is due to the form of life that language users share (Wittgenstein (1953), §66-67). The agreement among mathematicians (Wittgenstein (1953), §240-241) or the disagreement between humans and Wittgenstein’s “talking lion" is deeply related to the polysemous nature of natural languages (Wittgenstein (1953), p. 225). Similarly, for Ryle, without polysemy, an “original thought" is not possible (Ryle (2009), p. 216).

Wittgenstein states that the source of “our craving for generality" is “our preoccupation with the method of science" (Wittgenstein (1965), p. 18). He indeed makes a distinction of human intellectual activities based on spontaneity and implicitness. It distinguishes philosophical activities from the scientific ones. Similarly, it makes a distinction between natural language understanding and logical theory. In other words, Wittgenstein opposes a general theory of semantics. For this reason, G. P. Baker and P.M.S. Hacker claims that "... Wittgenstein builds no such theories. He does not contend that a language is monolithic structure run through with truth-conditions or assertion-conditions that give meanings to sentences and words. It is not a calculus of rules, either in the form of classical logic or in the form of intuitionistic logic. It is a motley of language games, an endlessly variegated form of human activity, interwoven with our lives at every level" (Baker and Hacker (1984), p. 441).

In this sense, as Michael Pelczar points out, natural language semantics has the property of *essential openness*, which is “what distinguishes common law from statutory law, improvisational jazz from closely scored musical styles, and conversation from litany" (Pelczar (2000), p. 500). Pelczar’s comments on jazz that “the rules that structure the improvisation now are not a score, but something more mysterious: unspoken rules of phrasing and harmony that the players would probably be hard pressed to articulate (but quick to detect violations of)" (Pelczar (2000), p. 499) are in line with Louis Armstrong's statement that "if you have to ask what jazz is, you'll never know" is in line with the picture of language understanding defended in this work. In short, natural language understanding, improvisational jazz or common law cannot be reduced to a few explicit rules. There is spontaneouity in them.

Thomas Kuhn goes one step further and argues that there is the property of "essential openness" in scientific practice as well (Kuhn (1996) pp.44-47, 101-103). "A network of similarities... plays an essential role in establishing links between scientific language and the world" (Kuhn (1979), p. 539). For this reason, "the magnitude denoted by 'mass' in Newtonian mechanics is not the one denoted by 'mass' in relativistic mechanics" (Elgin (1996), pp. 19, 78). This is, indeed, a demonstration of the incommensurability theory. However, as "mass" is a polysemous term, there must be a relation between different paradigms. That is why Kuhn says *incommensurability* does not mean *incomparability* (Kuhn (1979), pp. 539-540). For Catherine Elgin, Kuhn and later Wittgenstein are in the same camp (Elgin (1996), p. 16), *pure procedural epistemology*, a term she borrows from John Rawls (Elgin (1996), pp. 4, 5). For pure procedural epistemology, *procedure, reasoning, norms* are constitutive of knowledge. For Elgin, Wittgenteinian version is monopolistic as "conventions that underlie our form of life are constitutive of human rationality. Similarly, "Kuhn contends that monopoly is required for, and is imposed by, mature science" (Elgin (1996), p. 18). In other words, for these philosophers, concepts we use in science, art, politics etc. are not frozen and neat. Instead, they are dynamic and messy as they are governed by the dynamic forces of our lives and cultures.

The history of philosophy gives us enough reason to believe that it is not easy to handle our concepts. Otherwise, we would have a book of concepts where you pick up any concept you need. One reason behind this is I think concepts are dense if continuous in the mathematical sense. In set theory, density is defined as follows:

> In general, if a relation R linearly orders a set M, the ordering R is called dense

if between any two elements of M there is another element. Formally, R is dense if for all a and b in M, if $(a, b) \in R$, then there is a c such that $(a, c) \in R$ and $(c, b) \in R$ (Kossak (2018), p. 55).

I think concepts have this property since nothing prevents us from imagining a concept between any two concepts. For example, between any two related senses of “bank", we can imagine a third one. Besides, the concept similarity relation is a multi-dimensional one. Considering the multi-dimensionality and the density of the conceptual space, concepts that we use in life are too complicated to be traceable. Suppose we name a concept at a certain location mass_1. There is a huge number of very close concepts: $\text{mass}_{1.000...1}$, $\text{mass}_{1.000...2}$... This complexity is the reason behind ordinary language philosophers' pessimism that we mentioned. It is impractical to trace individual concepts if it is not impossible at all.

3.2.2 Human language and conceptual structure are intertwined

Even though our concepts are too complex to be individuated and handled easily, one may still insist that it is not a problem with language. In other words, one may attempt to isolate linguistic meaning from our full-fledged conceptual network. However, several works on the cognitive side of language show that human language and conceptual structure are intertwined due to the polysemous nature of human language.

Work in the cognitive semantics tradition is illuminating to see the deep relation between language and human conceptual network. Cognitive semantics introduces three cognitive mechanism to see the essential place of polysemy in language cognition: image schema transformation, metaphorical projection, and metonymy. In contemporary linguistics, the school of thought that centers its research around the cognitive importance of polysemy is cognitive semantics. Cognitive semantics was originally developed against the symbolic computation of meaning (Lakoff (1988), p. 149). In cognitive semantics, meaning is analyzed in relation to cognitive beings that process it. So, the faculty of understanding is crucial to meaning theories in cognitive semantics (Johnson (2013), p. 190). In this sense, cognitive semantics is a maximalist approach as opposed to the minimalist tendencies of formal semantics (Geeraerts (2013), p. 577) (Geeraerts (2010), p. 240).

Cognitive semantics is traced back to Eleanor Rosch's analysis of human categories (Lakoff (1987), p.15). According to Rosch, human categories are not as simple as Aristotelian categories, which are defined in terms of necessary and sufficient conditions. For Rosch, "the task of category systems is to provide maximum information with the least cognitive effort" (Rosch (1978), p. 28). Both horizontally and vertically, human categories are structured to achieve this task. Vertically, the category hierarchy is centered around basic categories even though they are neither the most encompassing nor the most specifying sets of objects. However, they are cognitively the most fundamental categories. At this level, cognitive agents "maximize within-category similarity relative to between-category similarity"(Mervis and Rosch (1981), p. 92). By several experiments that investigate "attributes in common, motor movements in common, objective similarity in shape, and identifiability of averaged shapes" (Rosch (1978), p. 31 (Rosch et al. (1976), pp. 390-405), Rosch adds that basic level categories also play the central role in imagination, perception, language acquisition, and language evolution (Rosch (1978), pp.34-35).

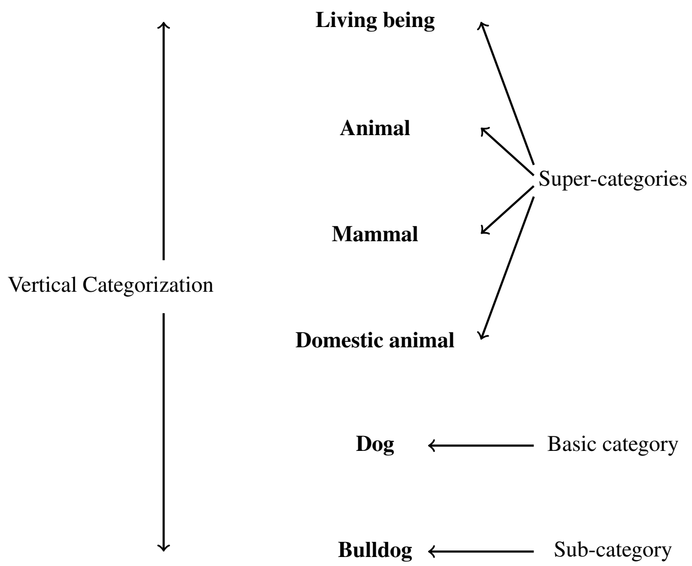


Figure 3.2: A vertical representation of the human categorization according to Rosch

For Rosch, at the horizontal level, most of the concepts are centered around prototypes. This means that unlike the classical categories, membership is given by degree (Rosch (1978), pp. 35-36). For example, a robin instantiates the category bird better than a penguin. (Rosch (1978), p. 39) By several experiments on *language processing time, speed of category learning, the logic of natural language using*, etc., Rosch shows the cognitive importance of the prototype effect (Rosch (1978), pp. 38-41).

For Rosch, the prototype effect is not the endpoint to define our categories (Rosch (1978), p. 40). Instead, we need a mechanism to explain the highly complex structure of human categories, the Wittgensteinian *forms of life* (Wittgenstein (1953), §23). The degree of membership evokes fuzzy set theory (Zadeh (1965), p. 339). Nevertheless, Rosch is not interested in modeling human categories with a handful of rules (e.g.,intersection, union, etc.) (Rosch (2013), pp. 592-593).

Cognitive semantics aims to describe this mechanism (Lakoff (1987), p.45). This mechanism involves a *background* against which word sequences are processed. In need of a background to process a linguistic input has been pointed out by several semanticians from different traditions (Schank and Abelson (2013), (Rumelhart (1975), Searle et al. (1983), Searle (1978)) (Miller (1999), pp. 13,17 (Clark (1983), pp.324-325). Cognitive semanticians' notion of *background* is rooted in Charles Fillmore's *Frame Semantics*. Frame means "any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits" (Fillmore (2006), p.373). These frames are *the way the world is* from our cultural and cognitive perspective (Fillmore (2006), p. 379). Rather than "a genuine body of assumptions about what the world is like," frames are prototypical, ideal models that we have about the world (Fillmore (2006), p. 379).

In Fillmore's example, the word "breakfast," illustrates this point. To understand what this word means, we use the frame that human beings eat three times a day. Breakfast is the one we have early in the day, and it consists of special foods. However, this is just an ideal model since, in reality, there are instances of breakfast that violate all the properties listed above (Fillmore (2006), p.380). George Lakoff calls these ideal models *Idealized Cognitive Models* (Lakoff (1987), p. 68). Language users use these idealized cognitive models to give meaning to linguistic expressions. When these idealized cognitive models conflict with our experiences, they are radially extended (Lakoff (1987), p. 111). A good example is the word "bachelor." With respect to an idealized cognitive model, "bachelor" is defined as *an unmarried man*. However, *the Pope, Tarzan, gays, Muslims with only three wives* are not considered in this idealized model of the world (Lakoff (1987), pp.85-86). These prototypical, idealized cognitive models are extended by experience (see also (Quine (1951), pp. 40-43) (Hylton (2019), p. 11)). For example, an idealized cognitive model in which *mother* is defined as a person "who is and always has been female, and who gave birth to the child, supplied her half of the child's genes, nurtured the child, is married to the father, is one generation older than the child, and is the child's legal guardian" does not take *foster mothers* into account. A revised model that takes foster mothers into the consideration ignores a transexual mother "who gave birth but has since had a sex-change operation" (Lakoff (1987), p.83). These step-by-step extended categories are called *radial categories* (Lakoff (1987), p. 84). Radial categories are formed through *image schema transformations, metaphorical projections*, and *metonymies* (Lakoff (1987), pp. 107-109). In a sense, radial categories are what Wittgenstein calls *family resemblance* categories (Wittgenstein (1965), pp. 17,20).

According to cognitive semanticians, image schemas are analog structures that provide us with structural knowledge about meanings. Several formal characteristics can be extracted from these schemas. For example, the container schema below (Figure 2) is helpful to understand prepositions like *in, into, out* (Lakoff and Johnson (2008), pp. 30-32). It shows that the object contained is isolated from the objects located outside of the container. Containers limit the behaviors of contained objects. Similarly, "if B is *in* A, then whatever is *in* B is also *in* A. If I am *in* bed, and my bed is *in* my room, then I am *in* my room" (Johnson (2013), p. 22).

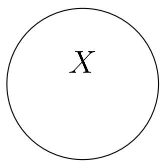


Figure 3.3: Container Schema
(Johnson (2013), p. 23)

- "There's a lot of land *in* Kansas."
- "The ship is coming *into* view."
- "He's *out* of sight now."
- "Halway *into* the race, I ran out of

energy."

- “He’s *out* of the race now.
- “He’s *in* love.
- “He fell *into* a depression."

Zenon Pylyshyn argues that image schemas, as opposed to propositions, are not “intrinsic properties of certain representational media or of certain mechanisms that are not alterable in nomologically arbitrary ways by tacit knowledge" but rather sets of propositions we use tacitly (Pylyshyn (1981), p. 17). So, according to Pylyshyn, image schemas are nothing but a set of discrete propositions. However, to cognitive semanticians, image schemas cannot be represented by finitary structures (i.e., finite sets of propositions), because they are continuous structures, “experiential gestalts" (Johnson (2013), pp 3, 27-28, 41). According to Mark Johnson, image schemas are neither propositional structures nor rich, concrete images. They are abstract, analog structures by which language users interpret words in different contexts (Johnson (2013), pp. 29-30).

One way polysemy arises is due to *image schema transformations*. In some contexts, an image schema transforms itself into another image schema. For example, the preposition *across* in the sentences below is polysemous. In one sentence, the trajectory in the image schema is a line; in the other one, it is the end-point of a line (Brugman and Lakoff (1988), p.502).

- Harriet walked *across* the street. (path)
- Harriet lives *across* the street. (end-of-path)

Similarly, in the sentences below, the trajectory transforms from points to a line. (Brugman and Lakoff (1988), p. 504).

- Sam ran *through* the forest. (0-dimensional trajectory)
- There is a road *through* the forest. (1-dimensional trajectory)

For Lakoff, these transformations reflect our experiences. For example, “when we perceive a continuously moving object, we can mentally trace the path it is following" (Lakoff (1987), p.442). Therefore, polysemy is grounded in our experiences.

Another source of polysemy arises from metaphorical projections. For cognitive semantics, metaphors are not linguistic ornaments, but they reflect our conceptual structures. This approach is relatively new in metaphor theories. Starting with Plato's criticism of poetry (Plato, Republic Book X, 603b-607c), philosophers in Western tradition such as Aristotle, Thomas Hobbes, John Locke, George Berkeley, and John Stuart Mill, approached metaphors normatively. Rather than building theories on how we grasp the meaning of metaphorical expressions, they argued against the use of metaphors in reasoning (Aristotle, *Posterior Analytics*, 97b-35) (Hobbes, *Leviathan*, Pt. I, Ch. 5, para. 20) (Mill (1882), pp. 49-50, 773, 779) (Berkeley (2009), p. 244) (Locke (1975) Book 3. Ch. 10-11). On the other hand, in post-Kantian philosophy, interest in human cognition due to Kant's *Copernican revolution* which takes human understanding into the center, leads philosophers to understand the cognitive mechanism behind the figurative use of language. For example, Nietzsche claims that due to our neural interaction with the world, the meaning we express, "when we talk about trees, colors, snow and flowers" are "nothing but metaphors for things which do not correspond in the slightest to the original entities" (Nietzsche (2009), p.256).

In contemporary philosophy of language and linguistics, there are several proposals on to understand *how we process metaphorical expressions*. Even though metaphorical meanings may become literal through frequent use, there is a sharp, clear boundary between metaphorical and literal uses of an expression (Devitt (2013), p. 95). Searle proposes that "where the utterance is defective if taken literally, look for an utterance meaning that differs from sentence meaning"(Searle (1993), p. 103). Similarly, to Grice, if the literal interpretation of a sentence violates the maxim of quality, which means that the literal interpretation is *obviously false*, then the sentence must be interpreted metaphorically (Grice (1989), p. 34). For relevance theorists, what matters is not the *truth* but the *relevance* of the interpretation when we interpret any expression (Wilson and Sperber (2002), pp. 601-603), including metaphors (Wilson and Sperber (2002), pp. 617-618). David Lewis analyses metaphors as if they are translations from another language to the literal language (Lewis (1975), p. 28). For Davidson, metaphorical expressions do not convey any meaning other than expressions' literal meanings. Instead, they attract our attention to a point. In other words, they have purely *use* value (Davidson (2001d), p. 259). Their point is not to provide some special meaning but only to attract language users' attention (Davidson (2001d), pp. 246, 263).

Unlike the approaches above, Nelson Goodman attracts our attention to the difference between metaphors and mere ambiguity (Goodman (1968), p. 70). To him, the difference is that there is a systematic relation between literal and metaphorical meanings:

> How, then, do metaphor and ambiguity differ? Chiefly, I think, in that the several uses of a merely ambiguous term are coeval and independent; not either springs from or is guided by another. In metaphor, on the other hand, a term with an extension established by habit is applied elsewhere under the influence of that habit; there is both departure from and deference to precedent. When one use of a term precedes and informs another, the second is the metaphorical one

(Goodman (1968), p. 71).

In other words, understanding them requires a conceptual schema (Goodman (1968), pp. 71-85). Cognitive semantics, from this perspective, goes deep into the rich conceptual structure behind metaphors. They claim that in metaphors, one domain is projected onto another domain systematically (Lakoff and Johnson (2008), pp. 7-9) (Johnson (2013), p. 130) (For a similar view, see Rumelhart (1979), p. 81)).

In these metaphors, above (Figure 3), *argument, love*, and *theories* are target domains that are understood in terms of a source domain, i.e., *war, journey*, and *buildings*, respectively (Lakoff and Johnson (2008), pp. 4, 44-45, 46). Language users systematically map items from the source domain to the target domain. The systematicity here shows that metaphors are located at the conceptual level. The metaphorical expression, "He attacked every weak point in my argument," shows that our concepts about a physical phenomenon, *war*, are extended to an intellectual, relatively abstract phenomenon, the *argumentation*. So, in cognitive semantics, "metaphor is the main mechanism through which we comprehend abstract concepts." Our everyday, bodily experiences are the source of metaphorical projections (Lakoff (1993), p. 244-245). For example, as bipedal animals, our physical and cultural experiences enable us to understand *orientational metaphors* in sentences such as "The discussion *fell to the*

Love is Journey

- Look *how far we've come.*
- We're *at a crossroads.*
- We'll just have to *go our separate ways.* We can't *turn hack now.*
- I don't think this relationship is *going anywhere.*
- *Where* are we?
- We're *stuck.*
- It's been *a long, humpy road.*
- This relationship is *a dead-end street.* We're just *spinning our wheels.* Our marriage is *on the rocks.*
- We've gotten *off the track.*
- This relationship is *foundering.*

Argument is War

- He attacked every weak point in my argument. His criticisms were right on target.
- I demolished his argument.
- I've never won an argument with him.
- You disagree? Okay, *shoot!*
- If *you* use that *strategy*, he'll *wipe you out*. He *shot down* all *of my* arguments.

Theories are Buildings

- Is that the *foundation* for your theory?
- The theory needs more *support.*
- The argument is *shaky.*
- We need some more facts or the argument will *fall apart.*
- We need to *construct strong argument for that.*
- The argument *collapsed.*

Figure 3.4: Conceptual metaphor theory

emotional level, but I *raised* it back *up to the rational* plane." or "He couldn't *rise above* his *emotions*" (Lakoff and Johnson (2008), pp. 15-17). In a nutshell, to cognitive semantics, our cultural and physical experience as a whole shapes our understanding of metaphorical expressions (Lakoff and Johnson (2008), p. 19). Marina Rakova, with whom I agree on this topic, goes one step further and through using psychological and neuroscientific work on double-function adjectives (e.g., "sharp knife" vs. "sharp person") (Rakova (2003), ch. 5) and synaesthetic adjectives ("bright sounds," "loud colors," "sharp tastes," etc.) (Rakova (2003), ch. 4) argues that there is no conceptual primacy of concrete meanings to abstract ones (Rakova (2003), pp. 2-3, 139-140). Instead, there is a psychological primitive concept behind both literal and so-called metaphorical meanings (Rakova (2003), p. 166). “Sharp" in both “sharp knife" and “sharp" sound has a literal meanings. What makes this case not homonymy is the presence of the primitive concept $SHARP$, which is being deployed to detect sharpness in different domains (Rakova (2003), p. 142).

Metonymy is another way to extend our categories. While metaphorical transfers are associated with *irregular polysemy*, metonymy is associated with *regular polysemy*. “The polysemy of a word A with the meanings a_i and a_j is called regular if, in the given language, there exists at least one other word B with the meanings b_i and b_j, which are semantically distinguished from each other in exactly the same way as a_i and a_j and if a_i and b_i and a_j and b_j are nonsynonymous" (Apresjan (1974), p. 16).

In the sentence “We need a couple of *strong bodies* for our team," the metonymical expression “strong bodies" refer to *strong people*. Similarly, “good heads" in “There are a lot of *good heads* in the university," refer to *intelligent people*. Nevertheless, metonymy is more than using one expression to refer to something in a special way. Rather, it is about how language users conceptualize the referent. For example, the expression, “good heads," emphasizes certain properties of those people (Lakoff and Johnson (2008), p.36).

Similarly, to understand the exclamation “Get *your butt* over here! We don't hire *longhairs*," or frequently used metonymical expressions like place names for institutions (*the White House*, producer names for products (*Ford, Heidegger*), one needs a conceptual background ((Lakoff and Johnson (2008), 38-40).

In sum, by image schema transformations, metaphors, and metonymies, cognitive semantics, at least, provide us with useful data to show the vital role of polysemy in our conceptual system.

Cognitive semantics tradition has two components. The first is their intricate description of the linguistic data. The second one is the theory by which they explain this data. Above, we are not interested in the second one. Rather, we are interested in the first one. I think that cognitive semantics is successful enough to attract our attention to the relation between language and the conceptual strcuture. Their analysis of the polysemous nature of language above provides enough findings to show that language and conceptual structure are too intertwined to be separated from each other.

3.2.3 Intractability is a not a mere acciendetal property of human conceptual structure

In the last two part, I try to show that human concepts are intractable and cannot be isolated from human language by referring to philosophical and scientific works. Is this merely contingent property? In other words, is it a defect? In this part, I will try to persuade the reader to believe that the massive complexity of the conceptual structure is not a defect but a beneficial property of efficient communication and cognitive system. Firstly, we will see that the ambiguous structure of language is an advantage in terms of communication. Secondly, the polysemous structure of language provides us with a well-connected small network that is beneficial for both cognitive and communicational systems.

Let us begin with the ambiguity part first. Lexical ambiguity, in general, is thought to be a problem in communication (Wasow et al. (2005), pp.269-270). For this reason, Noam Chomsky thought that language is not designed for communication (Chomsky (2002), pp. 106-107). For a sentence of $n-$word with m possible meanings for each word, the number of possible meanings is n^m. Depending on the size of n and m, the computation might be very costly (Falkum and Vicente (2015), p. 3). If these meanings are non-denumerable, as Herbert H. Clark claims (Clark (1983), p. 301), then the computation is impractical (see also Cohen (1985), pp. 132,134). This is the reason behind Yehoshua Bar-Hillel's pessimistic historical claim, "no existing or imaginable program will enable an electronic computer to determine that the word pen in the given sentence ["The box was in the pen"] within the given context has the second of the above meanings ["an enclosure where small children can play"], whereas every reader with sufficient knowledge of English will do this "automatically" (Bar-Hillel (1964)). For him, the only solution is that a machine needs the inference mechanism identical to our inferential mechanism to make the disambiguation (Bar-Hillel (1964)).

The *puzzle* or the *paradox* here is that what seems very challenging from a computational perspective is an everyday task for human language users (Taylor (2003), p.647) (Wasow et al. (2005), p. 272) (Wasow (2015), pp. 30,32,37-42). Polysemy, in particular, is not an exception but a norm in natural languages (Cohen (1985), pp.

131-132) (Clark (1983), pp. 299-300, 305). According to a view that dissolves this puzzle, lexical ambiguity, in general, is not a problem but rather a *desired* property of languages (Zipf (1949), p. 27) (Cohen (1985) p. 135) (Harris (1994), pp. 222-223). George Kingsley Zipf explains the desire for ambiguity through his theory of cognitive economy, which he considers one of the mechanisms behind human behavior (Zipf (1949), p. 19). For Zipf, in both individual and group behaviors, there are two conflicting economies. The balance that resolves the conflict determines the behavior of the group or the individual. In linguistic communication, speakers and hearers have conflicting interests. Speakers want to articulate multiple meanings with a single word, while hearers prefer the least ambiguity because both speakers and hearers would like to spend less effort (Zipf (1949), pp. 20-22). The conflict is resolved with a reasonable amount of ambiguity, which satisfies both the speakers' and the hearers' needs (Zipf (1949), pp. 27-28). For Zipf, this explains why more frequent words are more ambiguous (Zipf (1949), pp. 30-31).

From the information-theoretical perspective, an "optimally efficient communication system should look ambiguous, as long as context is informative about meaning" (Piantadosi et al. (2012), p. 282). In Claude Shannon's proposal, when information flows from a sender to a receiver, the degree of uncertainty is measured by the entropy (Shannon (1948), p.394). Information theoretically, "the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design" (Shannon (1948), p.379). Below is the entropy formula in the case of an ambiguous word.

$$H[M] = -\Sigma_{m \in M} P(m) log P(m)$$

However, what we need is the conditional entropy because hearers process words *given a particular context.*

$$H[M|Context] = -\Sigma_{c \in Context} P(c) \Sigma_{m \in M} P(m|c) log P(m|c)$$

Since conditional entropy cannot be greater than unconditional entropy, a language with no ambiguity is not economical (Piantadosi et al. (2012), p. 283).

(Ferrer i Cancho and Solé (2003)) modeled the overall energy spent in a conversation (see also Solé and Seoane (2015))

Speaker's effort

$$H_n(S) = -\Sigma_{i=1}^{n} p(word_i) log_n(p(word_i))$$

Hearer's effort

$$H_m(R|S) = -\Sigma_{i=1}^{n} p(word_i) \Sigma_{j=1}^{m} p(meaning_j|word_i) log_m p(meaning_j|word_i).$$

Overall energy consumption in a conversation

$$\Omega(\lambda) = \lambda H_m(R|S) + (1-\lambda) H_n(S).$$

If there is no ambiguity in the language, $H_m(R|S)$ is 0. This means that hearers would not need to spend any effort in disambiguation. This situation would be more difficult for speakers. On the other hand, if a speaker is able to mean anything by using only one word, then $H_n(S)$ would be 0 (Solé and Seoane (2015), pp. 21-22). Depending on the λ in the Ω formula, either $H_n(S)$ or $H_m(R|S)$ is more important. The algorithm designed by (Ferrer i Cancho and Solé (2003)) aims to minimize Ω results $\lambda = 0.5$, which confirms Zipf's view that the speaker's and the hearer's effort must be considered equally (Ferrer i Cancho and Solé (2003), p.789) (Solé and Seoane (2015), p. 25). When $\lambda < 1 - \lambda$, there is no communication; while $\lambda > 0.5$, the situation is more appropriate for artificial languages (Ferrer i Cancho and Solé (2003), pp. 789-790). At the human level, “a moderate level of ambiguity" is needed (Solé and Seoane (2015), p. 25) (Ferrer i Cancho and Solé (2003), pp. 790-791).

Empirical work on English, Dutch and German shows that both polysemy and homonymy are associated with phonological forms that are easier for language users to process (Piantadosi et al. (2012), pp. 285-287). Catherine L. Caldwell-Harris adds that “words that are highly frequent are those that will be easiest to access, and thus are likely to be extended into new semantic territory, either to fill a new semantic niche that has appeared due to technological or cultural innovation, or to supplant existing words that may be harder to access because of their lower frequency" (Harris (1994), p.223).

Nelson Goodman says, “when a system is semantically but not syntactically dense, the result may be inadequacy or ambiguity" (Goodman (1968), p. 162). In other words, if the syntax of L falls short of semantics, ambiguity arises. To counterbalance the semantic density, language users take advantage of a natural source, the context. Even though in some situations, the context is relatively uninformative, there is always a context (Falkum (2015), p. 90) (Rayo (2013), pp. 655-656). Sperber and Wilson states that “people do not come to the processing of new information with ‘blank mind’; they have some kind of short-term memory store (or several such stores, or devices functionally equivalent to short-term memory stores) whose contents are never simply erase, at least not when the individual is awake (Sperber and Wilson (1995), p. 139).

Now, let us go on with how polysemy is crucial for an efficient cognitive and communciation system. Solé and Seoane (2015) remarks, “the efficient character of the semantic network is associated to an important, universal, and yet apparently undesirable property of language: polysemy" (Solé and Seoane (2015), p. 16) (Solé et al. (2010), p. 22). The small world phenomenon in network theory shows that any two nodes in a network that consists of a large number of nodes (e.g., people, neurons, atoms, nations, etc.) can be connected with a few intermediary nodes (de Sola Pool and Kochen (1978)) (Milgram (1967)). de Sola Pool and Kochen (1978) “conjecture that, despite the effects of structure, the modal number of intermediaries in the minimum chain between pairs of Americans chosen at random is 2. We noted above that in an unstructured population with $n \approx 1000$, it is practically certain that any two individuals can contact one another by means of at least two intermediaries. In a structured population, it is less likely but still seems probable. And perhaps for the whole world’s population probably only one more bridging individual should be needed" (de Sola Pool and Kochen (1978), p.42). Milgram (1967) reports that “chains varied from two to 10 intermediate acquaintances, with the median at five" (Milgram (1967), p. 65).

Due to the polysemy, language provides us with a well-connected semantic network with a small-world structure (Solé et al. (2010), p. 22). For this reason, Peter

Norvig and George Lakoff state that “the difference between a network structure and a list is critical here. There are not just some random similarities and differences among the senses. Rather, the differences are minimal and of a restricted number of types. Only a network structure permits the statement of the minimal differences" (Norvig and Lakoff (1987), p. 205) (Brugman and Lakoff (1988), p.477) (Fellbaum (2013), p. 3). Additionally, children benefit from this property of semantic networks when they are learning how to use a word (Srinivasan and Rabagliati (2015), pp. 146-147) (Fellbaum (2013), p. 5).

I would like to illustrate the small-world phenomena in our conceptual structure with an example. Consider concepts like *education* and *tree leaves*. At first sight they do not seem related. Just as I and David Lynch is not related. However, since the human network is a well-structured one due to globalist politics, there must be a few people between me and David Lynch. Similarly, due to the polysemous structure of language, human concepts are well-stuctured as well. At least in English (also in Turkish), there is a poylsemous word, *book* which is used to refer to a physical entity and an abstract entity. It's physical meaning is an entity that is made from trees. It's abstract meaning is an entity which is an essential item of education.

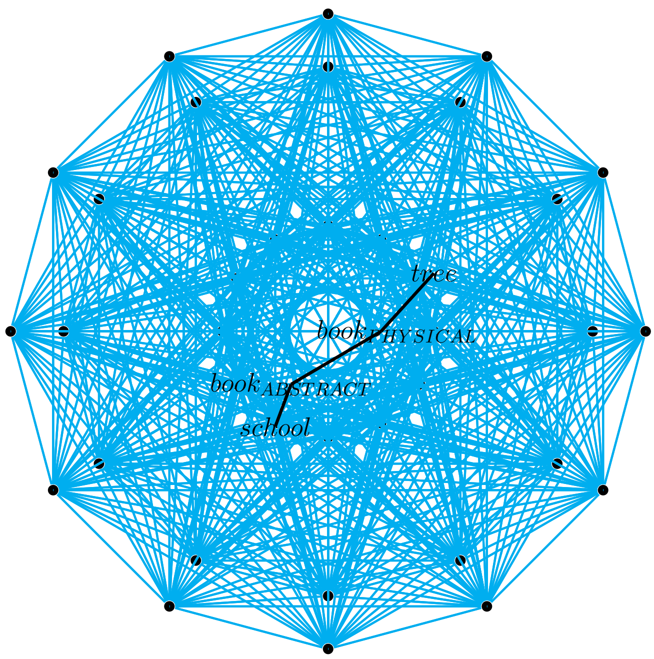


Figure 3.5: A polysemous word, "book", relates *tree* to *school*

CHAPTER 4

THE PRAGMATICS-ORIENTED APPROACH

4.1 Radical Contextualism

We have seen that polysemy is essential to our linguistic cognition. Modeling our excessive use of polysemy is a set of linguistic rules is not a solution. As Wasow claims, “saving memory at the expense of extra processing looks like a false economy" (Wasow et al. (2005), p 276). Taking the continuity from one meaning to another (Recanati (2004), pp. 134-135, 152) (Recanati (2010), p. 18) and the novel senses (Pustejovsky (1996), pp. 42-46) into consideration, the problem gets more complicated. Similarly, to explain semantic phenomena like “in the right contexts, *do a Napoleon* could be used to mean “pose with one’s hand tucked in one’s vest" as in *do a Napoleon for the camera*, “conquer by overrunning" as in *Hitler did a Napoleon to Poland in 1939*, or “go into exile" as in *The Shah of Iran did a Napoleon to an island off Panama in 1980*" (Clark and Gerrig (1983), p. 599), is a big problem.

For John R. Taylor, the source of the problem is that “we tend to think of word meaning as objects which can be contemplated independently of the linguistic means of their expression and which, when combined, allow the meaning of a complex expression to be computed from the meanings of its parts" (Taylor (2003), p.648). Similarly, George Miller claims that “the ambiguity appears only when we, quite arbitrarily, call isolated words the units of meaning" (Miller (1951), p. 112) (Miller (1999), p. 12). Word meaning is almost intractable. It is not even determinate which word is the source of ambiguity in a sentence. For example, in the sentence “I finished the book" according to one analysis, “finish" has the sense extension (Langacker (1984), pp. 181-182). "Finish" means *finish reading*. According to another analysis, “book" has that role (Pustejovsky (1996), pp. 199-200). (Recanati (2004), p. 34). In other words, " the book" means *reading the book*.

Contextualism is an approach to solving this problem. The view that words in isolation do not have any meaning is found in the works of several semanticians like Frege, Quine, and Davidson. (Frege (1953), §60) (Davidson (1967), p. 308) (Quine (1951), p. 39). Both context and compositionality principles are traced back to Frege's writings. There is substantial literature on the apparent tension between these two principles (see Pelletier (2001), pp. 87-94)(Janssen (2001), pp. 115-118) (Linnebo (2018), pp. 120-122).

In the literature, this approach is called radical contextualism, according to which word types may not be assigned full-fledged concepts. Instead, they may be abstract schemas, semantically underspecified entries. In some cases, they may be *collections of memory traces of previous uses and bundles of contingent encyclopedic knowledge* (Recanati (2004), pp. 146-151). Rumelhart, as a defender of this view, "reject[s] the traditional program of semantics and try to formulate a new account of both literal and conveyed meanings" (Rumelhart (1979), p. 74). For him, "the traditional program of semantic analysis (cf. Katz & Fodor, 1963) provides a set of meanings for the individual lexemes of the language and then provides a set of rules of composition whereby the individual meanings of the lexemes are combined to form the meaning of the sentence" (Rumelhart (1979), p. 74). Agustin Rayo's model below is an example of radical contextualism:

> With each expression of the basic lexicon, the subject associates a 'grab bag' of mental items: memories, mental images, pieces of encyclopedic information, pieces of anecdotal information, mental maps, and so forth... Different speakers might associate different grab bags with the same lexical item... In some cases - such as the logical connectives - a grab bag might contain something akin to a semantic rule, which is enough on its own to settle a range of application. And in some cases - such as explicitly defined technical terms - the grab-bag might contain the mental analogue of a piece of paper on which an object-language definition of the relevant lexical item has been written. (Rayo (2013), p. 648)

The position I defend is a radical contextualist one too. However, my version is not a representationalist one. I will clarify my position in the next chapter. The subject of this chapter is the representationalist ones. To note, in this view, there are two different approaches to meaning: *Truth Conditional Pragmatics* and the *Wrong Format* *view*. According to Truth Conditional Pragmatics, lexical entries need primary pragmatic processes to enter the utterance meaning. These processes are saturation and modulation. Saturation is a bottom-up contextual process that “takes place whenever the meaning of the sentence includes something like a ‘slot’ requiring completion or a ‘free variable’ requiring contextual instantiation" (Recanati (2004), p.7). Assigning values to indexical expressions is a well-known example of saturation as per David Kaplan (Kaplan (1989), p.505). According to the *indexicalist* version of compositional semantics, polysemous adjectives behave like indexicals ((Fodor and Pylyshyn (1995), pp. 124-125) (Szabó (2001), pp. 133-139) (Ludlow (1989), pp. 520-524). For example, “good" in “good philosopher" and “good thief" has the same value with an open slot, which is filled with different values. Modulation, on the other hand, is a top-down process that involves sense enrichment, sense loosening, and semantic transfer (metonymy) (Recanati (2004), pp. 23-25)(Recanati (2017), p. 379). “Neither enrichment, nor loosening, nor transfer, nor any other of the mechanism at work in modulation seems to require, on the side of the input, a ‘slot’ or gap in semantic structure demanding to be filled and triggering the search for an appropriate filler" (Recanati (2004), p.136). According to both the Wrong Format view and Truth Conditional Pragmatics, polysemous expressions need modulations in addition to saturations. Nevertheless, in Truth Conditional Pragmatics, modulation is optional. In other words, zero-modulation is allowed (Recanati (2017), p. 380). According to the Wrong Format view, which is more radical, modulation is not optional but mandatory. At least, “the distinction between mandatory and optional pragmatic processes is somewhat blurred" (Recanati (2004), p. 97). In this view, lexical entries are in the wrong format to enter the compositional process; they need to be modulated (Recanati (2004), p. 97) (Recanati (2017), p. 394). Designing lexicons with overspecified, underspecified lexical entries that we mentioned above are in tandem with these views.

In general, word types do not/may not have meaning at all (Recanati (2004), p. 141). Only in the context of a speech act does a sentence express a determinate content (Recanati (2004), pp. 3, 154). Szabó, an opponent of the radical contextualist approach, claims that “not only is this ‘the lazy man’s approach to philosophy,’ it undermines systematic theorizing about language" (Szabó (2006), p. 32). To him, “the more we believe context can influence semantic content, the more we will find ourselves at a loss when it comes to explaining how ordinary communication (let alone the transmission of knowledge through written texts) is possible" (Szabó (2006), p. 32). To Fodor and Katz, a realistic context representation is impossible (Katz and Fodor (1963), pp. 178-179). On the contrary, as a radical contextualist, I believe that instead of running away from the difficulties, we should aim at approximating the context. I think this is possible due to the progress in computer science, cognitive sciences, and robotics. As Salmon notes, this approach and the one according to which sentence types have meaning in isolation are "two radically opposing conceptions of semantics." A speech-act centered or a use-based approach does not postulate sentence type meanings (Salmon (2005), p. 321). (Borg (2004), p. 4). The way words contribute to the meaning of an utterance is described by Rumelhart:

> This approach is, I believe, quite different from the "standard" approach. The standard view emphasizes the "bottom-up" processes of constructing meaning from smaller component meanings. Nonlinguistic knowledge comes into play only after the set of possible meanings has been selected. My approach suggests that comprehension, like perception, should be likened to Hebb's paleontologist (Hebb, 1949), who uses his beliefs and knowledge about dinosaurs in conjunction with the clues provided by the bone fragments available to construct a full-fledged model of the original. In this case, the words spoken and the actions taken by the speaker are likened to the clues of the paleontologist, and the dinosaur, to the meaning conveyed through those clues. On this view, the processing is much more "top down" in that internal hypotheses are actively imposed on the observed utterances (Rumelhart (1979), p. 78).

So, instead of the rule-based approach, a new approach where language users are likened to Hebb's paleontologist or a crime scene investigator is adopted. The rule-based approach is strictly modular (Borg (2004), pp. 9-10). Lexical entries plus a set of combinatory rules do the first part of the job. Then, an inference mechanism does the context-dependent part (Grice (1989), pp. 31,34-37). For this view, the compositionality of sentence meaning and the context-dependency of what is communicated are "the results of radically different processes" (Borg (2004), p. 9). Nevertheless, according to the contextualist approach, context and words work in parallel to produce utterance meaning (Recanati (2004), pp. 27-29). In the next part, I will analyze two main models in line with this view:

- Relevance theory's inferential model
- Recanati's associationist model

4.2 Relevance theory

4.2.1 Gricean pragmatics

In his 1967 William James Lectures, subsequently published as "Logic and Conversation," Paul Grice was interested in the logical and ordinary uses of connectives such as not, and, or, if, all, some. Their uses in natural languages are different from their logical uses (Grice (1991a), p. 41). For example, *and* as a logical connective does not give us any temporal information. Rather, it states the truth of both conjuncts. In this sense, there is no distinction between the truth conditions of the following sentences.

- Grass is green, and snow is white.
- Snow is white, and grass is green.

However, in natural languages, "and" has another use that informs us about the temporal order as well. In this sense, the following two sentences do not have the same truth condition.

- They were married and they had children.
- They had children and they were married.

Similarly, logical and the ordinary use of "most" provide us with different information. Most animals, in logical use, does not exclude all animals. However, in English, there is a use of "most," which is an alternative to all. For Peter F. Strawson, these words have more than one meaning (Strawson (1952), p. 82). Grice rejects this view based on what he calls Modified Occam's Razor: "Senses are not to be multiplied beyond necessity" (Grice (1991a), p. 47). Instead of adding distinct entries for these words, he provides us with a mechanism to derive so-called secondary uses from the primary uses. In other words, his process begins with conventional meanings then moves to what is communicated.

To move from the sentence meaning to what is communicated, he appeals to the rational, cooperative behavior of natural language users.

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. One might label this the Cooperative Principle." (Grice (1991b), p. 26)

In relation to this cooperative behavior, language users follow several maxims. These maxims are grouped under the four categories below:

QUANTITY:

1. Make your contribution as informative as is required.
2. Do not make your contribution more informative than is required.

QUALITY:

1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

RELATION:

1. Be relevant.

MANNER:

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly. (Grice (1991b), pp. 26-27)

The hearer first computes the conventional meaning of an expression. Then if there is a violation of any of these maxims, she infers what else the speaker means. Here is the famous letter of reference example in which the first maxim of quantity is flouted. In the reference letter, a philosophy professor writes that "Dear Sir, Mr. X's command of English is excellent, and his attendance at tutorials has been regular. Yours, etc." The first maxim of quantity says make your contribution as informative as required. However, the letter is not informative about the philosophical skills of the student. Based on the context, background knowledge as well as the conventional meaning of the letter (Grice (1991b), p. 31), the hearer infers that the student is not good at philosophy (Grice (1991b), p. 33).

According to Grice, ironies and metaphors are produced by this inferential mechanism. Both ironies and metaphors flout the first maxim of Quality: do not say what you believe to be false. For example, if we think that the speaker does not believe that X is a fine friend, then her utterance, "X is a fine friend," is ironic. For the same reason, the sentence "You are the cream in my coffee" is used metaphorically (Grice (1991b), p. 34).

4.2.2 Relevance based inferential semantics

Relevance theory deviates from the code model of communication through using the Gricean framework above in its broadest sense (Wilson and Sperber (2006), p. 607). According to the code model, speakers encode their messages and send them to their audiences. In order to get the message, hearers decode received packages. The Shannon-Weaver communication model below illustrates this view (Weaver (1953), p. 264) (Shannon (1948), p. 381). Figure 4.1 is its human communication version (Wilson and Sperber (2006), p. 5)

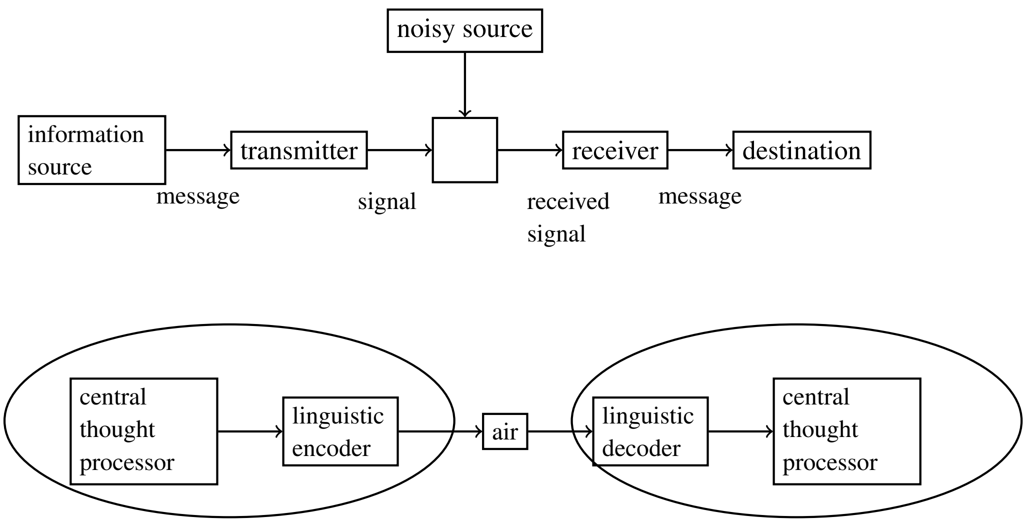


Figure 4.1: Code model of communication

The rule-based approach that we discussed in the previous chapter is an example of the code model of communication (Falkum (2011), pp. 77-82). In other words, it is the dominant model in semantics. "From Aristotle through to modern semiotics, all theories of communication were based on a single model, which we will call the code model" (Sperber and Wilson (1995), p. 2).

Even though Relevance theory uses Gricean elements, there are crucial differences between the Gricean approach and Relevance theory. First of all, Grice's approach above is not a deviation from the rule-based one. The meaning of an expression is computed through lexical meanings and a set of compositional rules. The Gricean approach is a two-step, sequential one. The first step is to compute literal meanings. The second step is to compute non-literal meanings if literal meanings, that are initially computed, violate Grice's maxims. In sum, Grice does not eliminate the code model. For each expression, its literal meaning is computed through the rule-based tools. In Relevance theory, instead of the literal-first strategy, the most relevant interpretation is computed first. Secondly, inferential processes in Relevance Theory are fundamentally different from the one in Grice's approach. It is not based on Gricecan rational, cooperative principles. The fundamental drive is the relevance of the interpretation. (Wilson and Sperber (2006), pp. 614-615, 619-620).

The motivation behind the relevance-based approach is the evolutional benefits: to spend less effort to process the expressions and increase the positive cognitive effects. "A positive cognitive effect is a worthwhile difference to the individual's representation of the world" (Wilson and Sperber (2006), p. 608). Similarly, "in relevance-theoretic terms, other things being equal, the greater the PROCESSING EFFORT required, the less relevant the input will be" (Wilson and Sperber (2006), p. 609). For this reason, the inferential process in Relevance theory must obey only one super-rule: maximize the relevance (Falkum (2011), pp. 98-106).

There are two inter-related principles of the Relevance theory. These are

1. Cognitive Principle of Relevance

2. Communicative Principle of Relevance

The first one is language users' strategy to maximize the relevance (Wilson and Sperber (2006), p. 610). The second one is the so-called Ostensive-Inferential communication model, which consists of speakers' informative and communicative intention

(Wilson and Sperber (2006), pp. 611-612). The second principle is related to Gricean account of natural meaning. In his 1957 paper, "Meaning," Grice distinguishes natural meaning from non-natural one by emphasizing speakers' intentions. For example, measles on somebody's face means that the person is sick. This is an example of natural meaning. However, in linguistic meaning, speakers have intention to inform their audience and make their intentions known by their audience (Grice (1957)). In the same way, according to Relevance theory, speakers, in addition to informing their audience, intend to inform their audience about their informative intentions. They do this by making their intentions manifest to both sides of the communication. In other words, there is the concept of mutual manifestness, which is an alternative to common ground or common knowledge/belief models (Clark (1996), pp. 92-121) (Lewis (2002), pp. 52-82) (Stalnaker (2002)). By uttering words, the communicator contributes to what is mutually manifest to both participants of the communication (Wilson and Sperber (2006), 1996, pp.58-63) (Falkum (2011), pp. 91-94).

In addition to the immediate perceptual input to which both participants are already exposed, the communicator adds ostensive stimulus to achieve his communicative goal. According to the Ostensive-Inferential communication model, "every ostensive stimulus conveys a presumption of its own optimal relevance" (Wilson and Sperber (2006), p. 256). In linguistic communication, each word is considered an ostensive stimulus. What makes a word relevant is the meanings associated with it. Word meanings, along with contextual ingredients, are regarded as clues to the hearer to make a relevance-based inference. Therefore, there is both a lexicon with underspecified lexical entries and compositionality rules. They produce logical forms of expressions. A logical form is an incomplete meaning which is an input to the inferential process. The relevance-based inferential process takes logical forms and contextual ingredients. Then, it outputs implicatures and complete meanings, which Relevance theorists call explicatures (Sperber and Wilson (1995), pp. 172-183) (Carston (2015), p. 198).

Let me quote Carston's illustration of a simplified version of the relevance-based computation.

Ann: I expected Jane to be here by now.

Bob: She missed her coach.

a Output of linguistic decoding of Bob's utterance:
SHEX MISSED HER COACH1 COACH2
where: 'SHEX' indicates the requirement to assign a referent
'COACH1' = instructor, 'COACH2'=bus

b Input to the pragmatic system:
Bob has said SHEX MISSED HER COACH_1 COACH_2
[Decoded logical form is embedded in a description of the ostensive act]

c Ann expects Bob's utterance to be optimally relevant to her
[General expectation of relevance triggered in the addressees of utterances]

d Bob's utterance will achieve relevance by explaining why Jane hasn't arrived yet.
[Specific expectation of the kind of cognitive effects the utterance will have]

e MISSING A DESIGNATED COACH2 IS A REASON FOR A PERSON ARRIVING WHEN EXPECTED
[Highly accessible assumption in the context which, together with other appropriate premises, might satisfy expectation (d)]

f JANE MISSED HER COACH2
[Highly accessible development of the logical form of Bob's utterance which can combine with (e) to lead to the satisfaction of (d)]

g JANE ISN'T HERE YET BECAUSE SHE MISSED HER COACH2 [Inferred from (e) and (f), satisfying (d)]

h JANE MAY STILL ARRIVE AT A LATER TIME [Inferred from (g) plus background knowledge about Jane and transport possibilities, etc. One of several further cognitive implications which, together with (g), satisfy expectation (c)] (Carston (2007), pp. 20-21) (see also (Wilson and Sperber (2006), p. 616)).

In step f, the system infers the explicature. Even though it is an over-simplified version of the inference in Relevance theory, at least we see how the system uses contextual clues (e.g. Ann's previous utterance, background information, etc.) and the linguistic clues (i.e. the logical form of Mary's utterance, SHE MISSED HER COACH_1 COACH_2). The same model continues to compute what is implicated.

4.2.3 Ad hoc concept construction

Unlike the simplified example above, in a relevance-theoretic computation, lexical entries are modulated. These lexical entries are the encoded concepts that contribute to logical forms. Once the inferential process turns logical forms into explicatures, these encoded concepts turn out to be ad hoc concepts. For example, in the sentence "Philosophy is the mother of all sciences," the encoded concept, MOTHER, contains encyclopedic information. At the end of the inferential process, the concept MOTHER turns to the ad hoc concept MOTHER*. Even though in the encyclopedic entry we have information about sexuality, in the ad hoc concept MOTHER* we don't have anything about sexuality (Carston (2015), p. 202) (Falkum (2011), pp. 116-124).

There are two versions of the ad hoc concept construction. These are concept broadening and concept narrowing. In the concept narrowing case, the communicated concept is narrower than the encoded concept. For example, in the sentence "Tom has a brain," the extension of the communicated concept BRAIN* is a subset of the encoded concept BRAIN (Carston (2002), pp. 324-325) (Wilson and Sperber (2006), pp. 617-618) (Carston (2021), pp. 17-18). The other ad hoc concept construction method is the concept broadening in which the encoded concepts are broadened to be applied to more objects. There is an important distinction between Sperber and Wilson's original approach and Carston's novel approach concerning the concept broadening. In the original approach, the meaning of a metaphor or a loose talk is hidden in the implicatures (Sperber and Wilson (1995), pp. 231-237). For this reason, Carston thinks that this would be to return to the Gricean approach (Carston (2002), pp. 330-332). Rejecting this approach, Carston analyses concept broadening and concept narrowing at the same level. They turn encoded concepts into ad concepts by modulating their extensions (Carston (2002), p. 336). The diagram below shows the relation between concept broadening and concept narrowing. The first one is a concept narrowing example where ad hoc concept C* is a subset of the encoded concept L. The rest illustrates concept broadening/loosening. In the first of these examples, encoded concept L is a subset of the communicated concept C*. In the second one, they are distinct sets with a nonempty intersection. In the third one, extensions of the encoded concept and the constructed ad hoc concept are two disjoint sets (Carston (2002), pp. 325-326, 353-354).

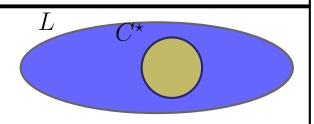
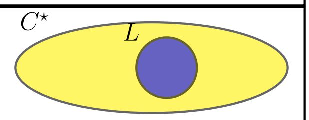
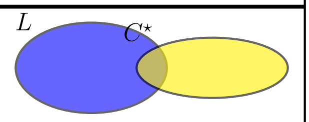
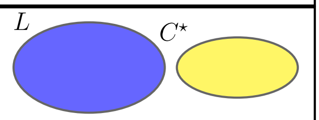
Modulation type	Example	Venn diagram
Concept narrowing	Boris is a *man* (Carston (2015), p. 202)	
Concept broadening	Boris is a *child*(Carston (2015), p. 202)	
Concept broadening (where literal use is contingently false)	Susannah is a *princess*(Carston (2002), p. 351)	
Concept broadening (where literal use is a category mistake)	Sally is a *block of ice*(Carston (2002), p. 345)	

Figure 4.2: Concept broadening and narrowing

In the original version of the Relevance theory, lexical entries are full-fledged concepts. A concept consists of three components.

i Logical entry (e.g., CAT → ANIMAL)

ii Encyclopedic entry

iii Lexical entry (i.e., phonological entry, etc.) (Carston, Thoughts and Utterances The Pragmatics of Explicit Communication, 2002, p. 312)

As the table above shows, these lexically encoded concepts turn to ad hoc concepts which may apply to a disjoint set of objects. Similarly, upon hearing a word, there may be several ad hoc concepts that a hearer may construct. For example, from the encoded concept HAPPY, several ad hoc concepts may be constructed: HAPPY*, HAPPY**... etc. At this point, Carston deviates from the original version of the Relevance theory and considers that so-called encoded concepts are among the ad hoc concepts (Carston, 2002, p. 362). For this reason, Carston offers that encoded meanings are not full-fledged concepts but underspecified entities that point "to a conceptual region, or maps to an address (or node, or gateway, or whatever) in memory" (Carston (2002), p. 360) (Carston (2015), p. 207).

In her late work, Carston considers two different types of lexicons, I-lexicon and C- lexicon. The first one is the agent-internal lexicon that generative linguists are interested. Lexical entries in the I-lexicon do not contain rich semantic information. Instead, they contain information about the form of meanings. The communicational lexicon consists of communicated concepts (Carston (2019), pp. 157-161) (Carston (2021), p. 12). For example, for the word "back," in the I-lexicon, there is a root item $\sqrt{back}$ that does not even provide the syntactic category of the word "back" (Carston (2019), p. 160) (Carston (2021), p. 21). On the other hand, there are several entries for back in C-lexicon, each of which roots back to the same entry in I-lexicon, $\sqrt{back}$. This is how polysemy represented with two distinct lexicons. The reason behind introducing a lexicon with ad hoc concepts rather than constructing ad hoc concepts inferentially is the presence of conventionalized, established senses of words. For Carston, even though all polysemy originates from pragmatic inferences, there is a distinction between creative, novel senses and established ones. The former is called pragmatic polysemy, while the second one is called semantic polysemy (Carston (2021), p. 12). C-lexicon is introduced to deal with semantic polysemy.

In this sense, C-lexicon is frequency sensitive while I-lexicon is compositionality sensitive. After entering into the syntactic compositionality, an entry in the I-lexicon turns to be an atom for the semantic compositionality. The output of the semantic compositionality is an input for further pragmatic processes.

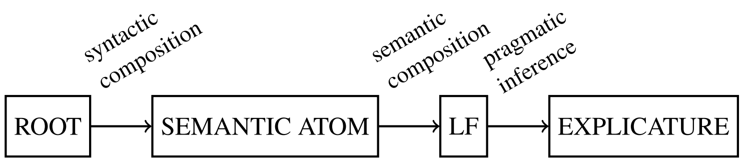


Figure 4.3: Carston's pragmatic model

Considering I-lexicon as the initial step in the compositional process, words, at this level, "have at most an underspecified schematic meaning (an possibly no meaning at all)" (Carston (2019), p. 162). Carston's point here is similar to the position I defend, Meaning Eliminativism. However, compositionality in Relevance theory which works with meaningful, representational entities, is an important gap between these two positions.

4.3 Recanati's associationist alternative

Recanati is another leading figure within the contextualist approach according to whom lexical meaning goes into a pragmatic process before entering into the compositional process. The compositional meaning of a complex expression is formulated as below:

$$I(\alpha^\wedge\beta) = f(g_1(I(\alpha)), g_2(I(\beta)))$$

In the formula above, I is the interpretation function. Functional variables f, g_1, and g_2 quantify over compositionality and pragmatic modulation functions, respectively. This approach is what Recanati calls pragmatic compositionality (Recanati (2010), s. 128-129) (Recanati (2004), pp. 138-140). The modulations in the formula is called primary pragmatic processes as opposed to secondary pragmatic processes. Primary pragmatic processes end up with what is said, while secondary pragmatic processes end with what is implied. As it is in the Relevance theory, primary pragmatic processes consist of concept broadening/narrowing, concept transfer (Recanati (2017), p. 379) (Recanati (2004), p. 23). To note, modulation is optional since a pragmatic function, g_n, may be the identity function. In other words, a pragmatic function may take the concept $MOTHER$ and give the same concept $MOTHER$.

In his late works, Recanati defends the view that lexical meanings cannot go into the composition process. In other words, pragmatic processes are not optional but mandatory. According to this view, a lexical meaning is a network of sense extensions. Even though each new sense is pragmatically introduced, they may be conventionalized. Conventionalized senses take place in the network structure of lexical meanings. In this way, there are two different ways of finding the true sense of a word in a context: The first one is to choose one of the conventionalized senses, which are originally pragmatic (see Falkum (2011), pp. 160-199) for the pragmatic origin of the systematic polysemy). The second one is to construct a new sense based on the lexical information and the contextual associations. At this point, we should not confuse three different entities: lexical meanings, literal meanings, and the contextual senses (Recanati (2019), pp. 220-222) (Recanati (2017), pp. 390-397). Below is an illustration of the lexical entity for the word "mother."

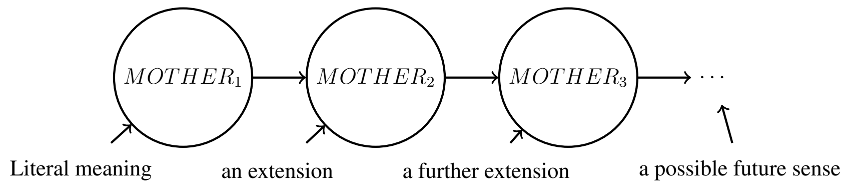


Figure 4.4: Meaning network that is used in Recanati's semantic and pragmatic polysemy resolution models

In terms of polysemy, sense selection and modulation correspond to two different polysemy resolution methods. In this sense, similar to Carston, Recanati distinguishes semantic and pragmatic polysemy examples.

The difference between Relevance theory and Recanati's contextualism is the role of inference in pragmatic processes. In Relevance theory, both what is said (explicature) and what is implied is an outcome of the inference process. In other words, the same mechanism, inference, governs both pragmatic processes. On the other hand, Recanati claims that primary pragmatic processes that result in what is said are not inferential. The reason is that primary pragmatic processes for Recanati are not conscious processes.

Even though Relevance theorists claim that pragmatic inferences are unconscious processes in their theory, Recanati does not think so. For him, Relevance theorists' inferential steps are tacit, spontaneous, but still conscious ones. In other words, when p implies q, even though inference steps are not explicitly stated, they are consciously available (Recanati (2004), pp. 39-41). For Recanati, this is not the case in deriving what is said. In that case, pragmatic steps are not consciously available. Availability starts after we reach what is said. "On this picture, there are only two basic levels: the bottom level at which we find both the meaning of the sentence and the contextual factors which combine with it to yield what is said; and the top level at which we find both what is said and what is implied, both being consciously accessible (and accessible as distinct)" (Recanati (2004), p. 13).

4.4 Problems with pragmatics-oriented approaches

In the end, the approaches above are radically contextualist. When natural language users understand a natural language expression, context does most of the job. Background information consisting of encyclopedic knowledge about word meanings, perceptual ingredients, and the general theory of the world play a significant role in language processing.

I agree with the contextualist, holistic approach to natural language understanding. However, there is a crucial distinction between my approach and linguistic pragmatism. In their approach, meaningful entities enter the computational process. In other words, representational entities play roles in the syntactic steps. Consider following words from William Faulkner's, *Sound and Fury*.

> ...Because Father said clocks slay time. He said time is dead as long as it is being clicked off by little wheels; only when the clock stops does time come to life.

As I argued in the previous chapter, it does not seem possible to process an expression like the one above with a lexicon and a set of linguistic rules. I agree with linguistic pragmatism that in order to process it, we need to know how to deal with the context. However, to process the passage above, linguistic pragmatists integrate context into lexical meanings. In doing this, they deal with context through representing it. It means that there are intermediary entities, representations. What we need is not only representations of the key words in the passage (e.g. time, life, wheel, stopping, etc.,) but also every single item that is associated with them. My objection to linguistic pragmatism is the difficulty of this job. In the rest of this section, I will lay out my pessimism about the context representation.

Let us begin with the concept of context in the Relevance theory. A context for the Relevance theorist is not the direct physical inputs that the linguistic agent is exposed to. Rather it is the agent's construction:

> A context is a psychological construct, a subset of the hearer's assumptions about the world, that affect the interpretation of an utterance. A context in this sense is not limited to information about the immediate physical environment or the immediately preceding utterance: expectations about the future, scientific hypotheses or

> religious beliefs, anecdotal memories, general cultural assumptions, beliefs about the mental state of the speaker, may all play a role in interpretation. (Sperber and Wilson (1995), pp. 15-16)

For linguistic pragmatism, in processing linguistic expressions, a relatively small subset of immediate perceptual input and the agent's background is actively used. In other words, all the information is not used in comprehending a linguistic expression. To pick up the correct subset, the cognitive expectation of relevance or local associations are used. Otherwise, it would not be efficient. However, even though we do not need to use all the information in the online processing, the bigger set must be represented somewhere in the system. The context representation consists of two parts: the background beliefs, in other words, "expectations about the future, scientific hypotheses or religious beliefs, anectodal memories, general cultural assumptions, beliefs about the mental state of the speaker," etc. (Sperber and Wilson (1995), pp. 15-16) and the information about the immediate physical environment.

To be used in language comprehension mechanically, there must be a way to represent these items. If the information is not propositional, we don't know how to represent and use it algorithmically. If the information is propositional, one way is to enumerate them: $\{p1, p2, p3, \ldots, pn\}$. The first problem with this approach is that even if there are only a finite number of propositions to be represented, there must be too many of them to be stored. To see the complexity, consider the sentence Carston quote from Carl Sandberg. "The fog comes on little cat feet." Carston continues that "the question, again, is how we go from information plausibly stored in memory about cat feet to thoughts about fog, or from our knowledge about lighthouses to thoughts about the nature of love" (Carston (2002), p. 352). Concerning such examples, there must be a vast number of propositions to be represented. To pick up the relevant or locally associated propositions seems computationally impractical.

Even if we solve this problem, we confront another one. To apply inferential rules, we need to represent relations between these propositions. If there are no relations between these propositions, we cannot use them in inferential processes. Suppose the context consists of three propositions, $\{p_1, p_2, p_3\}$ and the linguistic information is $\{l_1, l_2 \ldots\}$. If there are no relations between these propositions, we can only infer trivial conclusions like $p_1 \wedge p_2$, $p_2 \wedge p_3$, $p_1 \vee p_3$, etc. In order to infer novel information like $n_1, n_2, \ldots$ we need to list rules in the form below:

$$\begin{array}{c} p_i \\ \vdots \\ l_j \\ \vdots \\ \hline \therefore n_z \end{array}$$

It would be problematic in two ways. The first one is the complexity of listing all these rules. The second one, we need to store all those possible consequences $n_1, n_2, \ldots$ beforehand.

The solution is to use a tool richer than propositional logic. As Russell and Norvig say, "propositional logic, as a factored representation, lacks the expressive power to concisely describe an environment with many objects" (Russell and Norvig (2020), p. 252). Using predicates, quantifiers, modal operators, variables serves better in representing the background information.

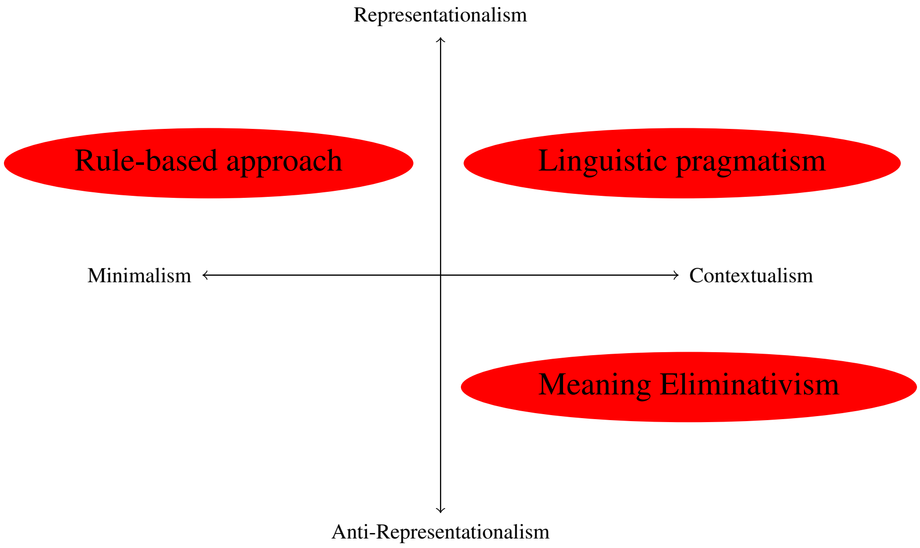
Now the problem is to find predicates to represent every fine shade of meaning. Plus, we need to represent the relations between predicates. The relation between the concept bank$_1$ the institution and bank$_2$ the building must be represented. An alternative way is to find primitive predicates on which other predicates can be built. However, even though there are such proposals in the literature (Wierzbicka (1996)) (Jackendoff (2002), pp. 333-377) (Jackendoff (1995), pp. 25-67) (for a criticism of this position see Fodor (2008), pp. 25-49), we don't have any clue for the reality of a primitive set. Indeed, the problem is not the lacking of a primitive set of concepts. Instead, it is the problem we discussed in the previous chapter when we argue against the rule-based approach. Human concepts are too complex and dynamic to be represented with a handful of rules. As we argue in the previous chapter, there is a possible meaning between any two concepts/meanings. This is the continuity of meaning claim that we discussed in the previous chapter. The reason we end up with the same problem is using meanings/representations at the syntactic level. A machine that is capable of understanding linguistic expressions is already subject to physical, meaningless inputs.

The point here is to add an intermediary layer of representation. In other words, this intermediary level which turns inputs into input representations, makes things more complicated.

As a conclusion, I am not even sure that context representation is possible. The trial above shows that possible tools to represent the context do not seem promising. This claim is in line with Fodor's pessimism about the global, contextual approach.

> In fact, I should like to propose a generalization; one which I fondly hope will some day come to be known as 'Fodor's First Law of the Nonexistence of Cognitive Science'. It goes like this: the more global (e.g., the more isotropic) a cognitive process, like analogical reasoning, aren't understood at all. . . .
>
> I am suggesting that, as soon as we begin to look at cognitive processes other than input analysis – in particular, at central processes of nondemonstrative fixation of belief – we run into problems that have a quite characteristic property. They seem to involve isotropic and Quineian computations; computations that are, in one or other respect, sensitive to the whole belief system. . .
>
> If we assume that central processes are Quineian and isotropic, then we ought to predict that certain kinds of problems will emerge when we try to construct psychological theories which simulate such processes or otherwise explain them; specifically, we should predict problems that involve the characterization of nonlocal computational mechanisms. . . Whereas, when we turn to the fixation of belief, we get a complex of problems that appear to be intractable precisely because they concern mental processes that aren't local. Of these, the frame problem is, as we have seen, a microcosm. (Fodor (1983), pp. 107-117)

To get rid of the problem, Fodor goes in the opposite direction from the Quinean holistic, contextual approach. Linguistic pragmatists and I agree on going the holistic and contextual direction. The distinction between linguistic pragmatists and me is their representationalism. The problem I see is not the imprecision of context representations. The problem is to use these entities in the syntactic processes. In that case, relatively ignorable misrepresentation of a single contextual ingredient may turn out to be giant problems at the end of the syntactic process. Therefore, the problem is not the imperfectness of the context representation, rather using these non-precise entities in syntactic processes. To illustrate the problem, consider modeling self-driving cars with context representations. Small mistakes in representing roads, buildings, trees, air, etc., may end up with a giant mistake. In the next chapter, to get rid of these problems, I will introduce a mechanistic approach: Meaning Eliminativism.


Representationalism
Rule-based approach
Linguistic pragmatism
Minimalism
Contextualism
Meaning Eliminativism
Anti-Representationalism

CHAPTER 5

A SOLUTION: MEANING ELIMINATIVISM

5.1 Introduction

According to the position I propose, expressions not have any meaning or meaning-like property at all. In other words, I do not assign any representational entity to expressions at all. In the last two sections, I try to show that those entities make hashes of things. Meanings, context representaitons, semantic combination rules, etc. are relatively simple tools.

Then, how do I explain language understanding? First of all, let me make theory - observations distinction. Word type meanings, sentence type meanings, grammar are theoretical tools to explain the observed phenomena, *understanding an utterance*. On the one side, there are full-fledged lexical entries, combination rules etc. On the other hand, for Jeffrey Elman “in this scheme of things there is no data structure that corresponds to a lexicon. There are no lexical entries" (Elman (2009), p. 556). Similarly, Miller says, “what people know when they know the meaning of a word is more a skill than a fact, the skill of incorporating that word appropriately into meaningful linguistic contexts" (Miller (1999), p. 5). A sequence of words is nothing but, as Fillmore says, “a record of the tools that somebody used in carrying out a particular activity" (Fillmore (2006), p. 374).

All these theoretical frames are designed to explain our observations. Let’s now go on with what we observe in people’s language use. I think language understanding in this respect is a set consists of pairs. Each pair consist of an observed linguistic expressions (i.e., phonological items), and the corresponding behavior of an average language users.

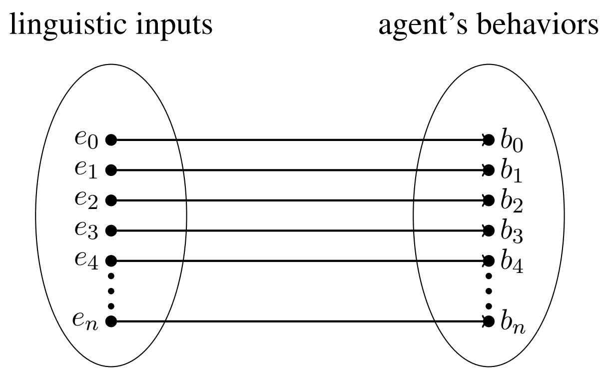


Figure 5.1: Language understanding is function from the set of linguistic inputs to the set of behaviors of a language user

Meaning is a tool to understand this set. In other words, meanings are hidden intermediary entities that relates expressions to language users behaviors. Even when a toddler or an adult using a word referentially, meaning is not directly available. What we observe is a linguistic agent, a sound wave and a set of behaviors. To make sense of this observation, researchers postulate meanings. There are two main reasons for doing this. The first one is that meanings are part of our cultures. In other words, since it is a pre-scientific notion, we are accustomed to use it. The second one is that any approach that is based on meanings is simple. For example, the rule-based approach we discussed is simpler than the linguistic pragmatism because linguistic pragmatism has other ingredients as well. However, the world is not so simple. For this reason, linguistic pragmatism relies less on meanings. For the reasons that I discussed in the last two chapters, I reject to postulate meanings to explain language understanding.

The contexualist approach I propose does not say that “context does everything, while the words themselves contribute nothing." In that case, “it does not matter any more whether the word one utters is ‘red’ or ‘rectangle’ " (Recanati (2004), p. 151) (See also (Gasparri (2013), p. 1021) for an argument against Meaning Eliminativism).

According to our approach, not *words' meanings* but *words themselves* contribute to understanding expressions. In other words, the relation is causal. Just as an electrical signal may cause a mental state, words themselves cause language users to move from one semantic state to another. It means that language has only a causal role in language understanding. Suppose, with a medical tool, I send electrity to someone's brain to make her feel that she is in Mars or to make her realize that there is a danger. In that case, I don’t think that there is representational relation between the medical tool and the mental state. Rather, the medical tool causes her to move from one state to another. This is a crucial distinction between my approach and the approaches I criticize. In those approaches, linguistic inputs build the semantic space. In my approach, the agent is already in a semantic space, no matter we send her linguistic input or not.

Let me illustrate this distinction. In a language understanding scenario, there are three main ingredients:

- linguistic input (e.g., sequence of words)
- the agent (e.g., a human being, a robot)
- the environment.

In the compositional approach, parts of the linguistic input correspond to what we call "meanings". These meanings construct the semantic state that the agent enters in. In other words, the agent is not in a semantic state without getting any input. In the rule-based version, it is mostly the combinatorial relations between meanings construct the semantic state. In the pragmatics-oriented approach, the state construction work is shared by environmental, agentive and linguistic elements. These elements are all representational.

On the other hand, I do not postulate building blocks of the semantic state. In other words, the state is not built upon with some meaningful, representational entities. Instead, the agent is already in a semantic state. What linguistic and environmental elements do is to move the agent from one state to another. The process is a causal one. It does not built a mental state out of simpler units, it only causes agent to wander in a mental space.

The mental state governs the behavior of the agent. The mental state is richer than both the linguistic input and the output behavior. Neither linguistic inputs, not the output behavior can be identified with the mental state. A mental state is the state that govern not only human behavior but also the overall animal behavior. To note, further questions concerning its nature is not the subject of this work. In this work, we are only interested in its role in language understanding.

The question remains is how to build such a system. In other words, “how to account for the move from a non-conceptual, non-semantic entity to a conceptual, contentful one?" (Carston (2012), p. 622) The answer is to use a purely connectionist system. In the next chapter, I explain how connectionism help us to build such a model.

5.2 The sub-symbolic approach

There are two approaches to design a mechanical procedure to imiatate language understanding: the symbolic and sub-symbolic models (Eliasmith and Bechtel (2006), p. 1). The symbolic approach is based on the idea that there is a structural similarity between the expressions' surface syntax and their meanings. It means that word meanings play a fundamental role in the computation. On the contrary, according to the sub-symbolic approach, fundamental units that go into the computation are not meaningful items.

The symbolic approach works best at imitating regular, precise phenomena. However, in the previous chapters, I argue that language understanding cannot be reduced to a few rules. At this point, Davidson's pessimism on language understanding models is noteworthy.

> For we have discovered no learnable common core of consistent behaviour, no shared grammar or rules, no portable interpreting machine set to grind out the meaning of an arbitrary utterance... We should realize that we have abandoned not only the ordinary notion of a language, but we have erased the boundary between knowing a language and knowing our way around in the world generally. For there are no rules for arriving at passing theories, no rules in any strict sense, as opposed to rough maxims and methodological generalities" (Davidson (2005a), p. 107).

I agree that there is no categorical difference between “knowing a language" and “knowing our way around in the world." However, I believe there are millions of implicit rules behind both of them. The sub-symbolic paradigm is an attempt to provide these rules.

In symbolic computation, units entering the syntactic process have semantic values; they are *symbolic* units. In other words, units with context-independent semantic values are causally efficacious (Smolensky (1995a), p. 191). On the other hand, according to the sub-symbolic view, “cognitive descriptions built up of entities that correspond to *constituents* of the symbols used in the symbolic paradigm; these fine-grained constituents could be called *sub-symbols*, and they are the activities of individual processing units in connectionist networks" (Smolensky (1995b), pp. 33-34). In language cognition, the building blocks toward understanding expressions are not the meanings of the words but “fine grained sub-symbols" (Eliasmith and Bechtel (2006), p. 3) (Chalmers (1993), p. 309). Sub-symbolic computation consists of two levels:

1. Processing (syntactic) level: Units that play a role in formal/syntactic/mechanical operations do not have any semantic value.

2. Interpretation (semantic) level: The outcome of the syntactic process is interpreted.

(Smolensky (1995a), p. 167). Before going through this processing/interpretation distinction, I will briefly give the background of connectionism. Firstly, in section McCulloch and Pitts model, we see the emergence of connectionist, neural models withing the symbolic approach. In section, Perceptron, we move to the divergence of connectionism from the symbolic approach. In the following section, the backpropagation algorithm which is used to learn huge numbers of parameters is introduced. Similarly, the idea of recurrent neural networks which I use in my model is outlined. In the last part, connectionist approaches language processing is summarized.

The history of connectionism demonstrates the processing / interpretation distinction. Adding more layers, neurons to neural nets makes processing part less interpretable. Similarly, increasing the dimensionality of inputs results in less interpretability of input components.

5.2.1 What is connectionism?

Connectionism and its so-called “opponent," classical architecture, has a very dense history. Classical architecture models human behavior through a handful of rules. These are meaningful, symbolic rules that are designed by experts. On the other hand, connectionists appeal to a vast number of rules that are not meaningful or symbolic. In other words, according to connectionism, the rules behind our actions are not the ones we use in our rational explanations. Let's consider a person's political behavior. A symbolic system goes with if-then rules like *if she is liberal, then yes*. Alternatively, the system may give weights to each option (see Figure 5.2). If the system does not provide good results, the expert may add new features.

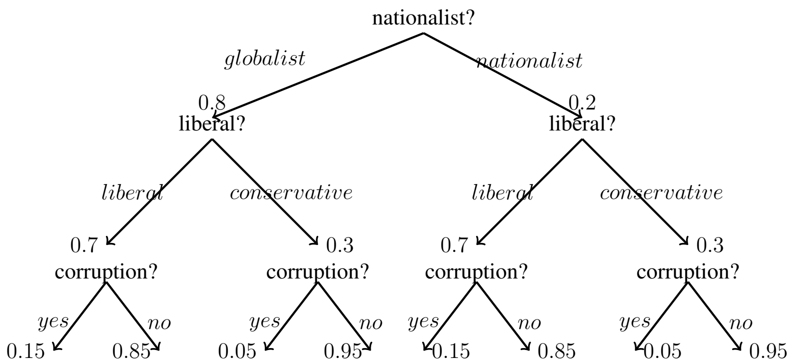


Figure 5.2: An oversimplified version of the rule-based account of the voting behavior of a person. Value of each leaf is the product of each choice in sequence

On the other hand, a connectionist system goes with implicit, micro rules that are not encapsulated by the rules in the symbolic system. I mean, in this case the system is not really interested in whether she is a liberal or not. Instead, there are a huge number of criteria that we are not aware of when we make choices. In this sense, Lakoff's point here can be considered as a background of connectionism:

> ...most of our thought is unconscious, not in the Freudian sense of being repressed, but in the sense that it operates beneath the level of cognitive awareness, inaccessible to consciousness and operating too quickly to be focused on... It means that we can have no direct conscious awareness of most of what goes in our minds. The idea that pure philosophical reflection can plumb the depths of human understanding is an illusion
>
> Conscious thought is the tip of an enormous iceberg. It is the rule of thumb among cognitive scientists that unconscious thought is 95 percent of all thought–and that may be a serious underestimate. Moreover, the 95 percent below the surface of conscious awareness shapes and structures all conscious thought.(Lakoff and Johnson (1999), pp. 10-12)

In my interpretation, connectionism goes one step further and claims that even the modeller does not have conscious access to every single determinant. Instead of being consciously aware of every single parameter, she lists them as a bunch of numbers: $x_1, x_2, ..., x_n$. Now I will continue with the history of connectionism.

5.2.2 McCulloch and Pitts model

Warren S. McCulloch and Walter Pitts' paper "A Logical Calculus of the Ideas Immanent in Nervous Activity" is often considered as the grandfather of connectionism (Goodfellow et al. (2016), p. 15). However, the paper is, indeed, the grandfather of both classical and connectionist architectures (Piccinini (2004), p. 204). The main idea in the paper is that human brain is a logic machine. Authors are originally psychiatrists who are interested in neural mechanism behind human psychology. Additionally, they are both interested in philosophy, logic and metamathematics (McCulloch and Pitts (1943), pp. 177-179, 183-184). For this reason, to model brain-mind relation, they appeal to the philosophical logic of the first half of the twentieth century. "To present the theory, the most appropriate symbolism is that of Language II of R. Carnap (1938), augmented with various notations drawn from B. Russell and A. N. Whitehead (1927), including the *Principia* conventions for dots." (McCulloch and Pitts (1943), p. 118)

Their aim is to match neural activity with the logical relations between propositions. For them, ultimate units of mental states, *psychons*, are related to each other logically. To reduce mental states to "all-or-none" behaviors of neurons is crucial since the standard logic is two-valued. "To psychology, however defined, specification of the net would contribute all that could be achieved in that field – even if the analysis were pushed to ultimate psychic units or "psychons," for a psychon can be no less than the activity of a single neuron. Since that activity is inherently propositional, all psychic events have an intentional, or "semiotic," character. The "all-or-none" law of these activities, and the conformity of their relations to those of the logic of propositions, insure that the relations of psychons are those of the two-valued logic of propositions. Thus in psychology, introspective, behavioristic or physiological, the fundamental relations are those of two-valued logic." (McCulloch and Pitts (1943), p. 131)

Let me briefly illustrate the neural network model they sketched. First of all, activation of any single neuron can be represented with the logical expression that authors call *temporal propositional expression* (TPE).

$N_i(z_1) \cdot \equiv \cdot S\{\prod_{m=1}^{q} \sim N_{jm}(z_1) \cdot \sum_{a \in K_i} \prod_{s \in a} N_{is}(z_1)\}$ (*Principia Mathematica* notation)

$N_i(z_1) \leftrightarrow S\{(\bigwedge_{m=1}^{q} \sim N_{jm}(z_1)) \bigvee_{a \in K_i} \bigwedge_{s \in a} N_{is}(z_1)\}$ (in Modern notation)

Above, $N_i(z_1)$ means that neuron c_i fires at time z_i. Neurons $c_{j1}, c_{j2}, c_{j3}, ...c_{jm}$ have inhibitory as neurons $c_{i1}, c_{i2}, c_{i3}, ...c_{im}$ have excitatory synapses upon c_i. K_i is the set of the subsets of neurons with excitatory synapses whose sum exceeds the threshold for c_i to fire. (McCulloch and Pitts (1943), p. 120)

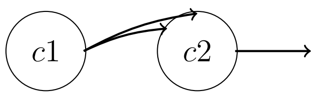
Not only every neural activity is represented with a temporal propositional expression, but also every temporal propositional expression can be expressed by a neural activity. (McCulloch and Pitts (1943), p. 120) Since any complex sentence can be reduced to elementary sentences and logical relations (i.e., conjunction, disjunction, conditionality and biconditionality), any expression can be associated with a neural activity (McCulloch and Pitts (1943), pp. 121-122).

> The “all-or-none" law of nervous activity is sufficient to insure that the activity of any neuron may be represented as a proposition. Physiological relations existing among nervous activities correspond, of course, to relations among the propositions; and the utility of the representation depends upon the identity of these relations with those of the logic of propositions. To each reaction of any neuron there is a corresponding assertion of a simple proposition. This, in turn, implies either some other simple proposition or the disjunction or the conjunction, with or without negation, of similar propositions, according to the configuration of the synapses upon and the threshold of the neuron in question" (McCulloch and Pitts (1943), p. 117).

Below is the representation of basic logical relations and their neural counterparts.

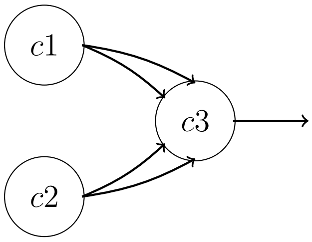
- a- Single proposition

$$N_2(t) \cdot \equiv \cdot N_1(t-1)$$



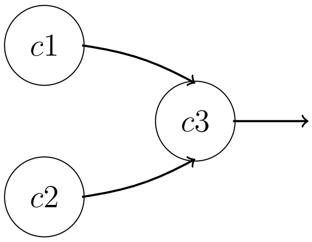
- b- Disjunction

$$N_3(t) \cdot \equiv \cdot N_1(t-1) \wedge N_2(t-1)$$



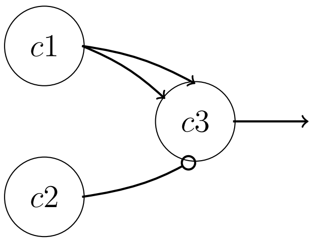
- c- Conjunction

$$N_3(t) \cdot \equiv \cdot N_1(t-1) \cdot N_2(t-1)$$



- d-

$$N_3(t) \cdot \equiv \cdot N_1(t-1) \cdot \sim N_2(t-1)$$



Theorem X below summarizes McCulloch and Pitts' point in the paper.

THEOREM X. Let us define a set K of S by the following recursion:

1. Any TPE and any TPE whose arguments have been replaced by members of K belong to K;
2. If $Pr_1(z_1)$ is a member of K, them $(z_2)z_1 \cdot Pr_1(z_2)$, $(Ez_2)z_1 \cdot Pr_1(z_2)$, and $C_{mn}(z_1) \cdot s$ belong to it, where C_{mn} denotes the property of being congruent to$m$ modulo, n, $m < n$.
3. The set K has no further members

Then every member of K is realizable.
For, if $Pr_1(z_1)$ is realizable, nervous nets for which

$$N_i(z_1) \cdot \equiv \cdot Pr_1(z_1) \cdot SN_i(z_1)$$

$$N_i(z_1) \cdot \equiv \cdot Pr_1(z_1) \vee SN_i(z_1)$$

> are the expressions of equation (4), realize $(z_2)z_1 \cdot Pr_1(z_2)$ and $(Ez_2)z_1 \cdot Pr_1(z_2)$ respectively; and a simple circuit, $c_1, c_2, ..., c_n$, of n links, each sufficient to excite the next, give an expression $N_m(z_1) \cdot \equiv \cdot N_1(0) \cdot C_{mn}$ for the last form." (McCulloch and Pitts (1943), pp. 128-129)

This neural network model is indeed compatible with the compositional approach as well. For example, using this model, we can relate the derivation steps in Montague semantics to neural activity. What is important here is that this human brain model is thought to be a Turing machine without a tape with infinite size.

> First, that every net, if furnished with a tape, scanners connected to afferents, and suitable efferents to perform the necessary motor-operations, can compute only such numbers as can a Turing machine; second, that each of the latter numbers can be computed by such a net; and that nets with circles [nets that inputs themselves] can compute, without scanners and a tape, some of the numbers the machine can, but no others, and not all of them. This is of interest as affording a psychological justification of the Turing definition of computability and its equivalents, Church's $\lambda-$ definability and Kleene's primitive recursiveness: If any number can be computed by an organism, it is computable by these definitions, and conversely (McCulloch and Pitts (1943), p. 129).

To add, the model is not actually equal to a Turing machine. Stephen Cole Kleene coined the introduce the concept, Finite State Automata, which he thinks exemplified by McCulloch-Pitts neural net (Kleene (1956), p.4, 32-34). He proved that finite state automata are only capable of representing *regular events*: "In any finite automaton (in particular, in a McCulloch-Pitts nerve net), started at time $t = 1$ in a given internet state b_1, the event represented by a given state existing at time p is regular" (Kleene (1956), p.34).

McCulloch-Pitts machine is not equivalent to a Turing machine because a Turing machine can represent *recursively enumerable languages* as it is seen in Figure 5.3.

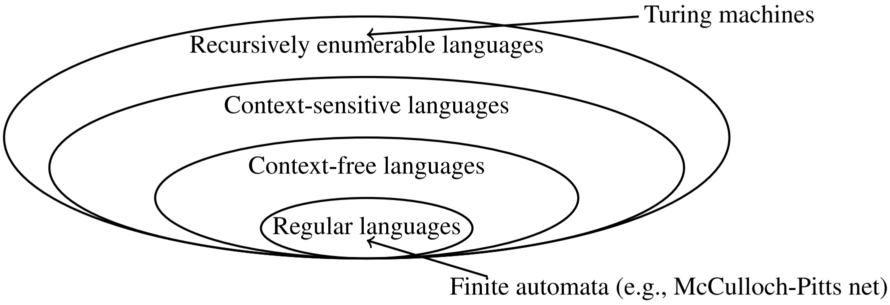


Figure 5.3: Chomsky hierarchy. See Sudkamp (1997), p. 310

5.2.3 Perceptron

Frank Rosenblatt, in his seminal paper, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain" takes this project further by asking two questions below:

- In what form is information stored, or remembered?
- How does information contained in storage, or in memory, influence recognition and behavior? (Rosenblatt (1958), p. 386)

To answer these questions he follows the connectionist approach "which stems from the tradition of British empiricism" (Rosenblatt (1958), p. 386). According to the connectionist approach, information is processed through the network of neural activity and connections. It is not recorded in the memory. For example, in our case different senses of a word is not stored anywhere in the system. This connectionist machine is called *perceptron* (Rosenblatt (1958), p. 387). A perceptron is a neural network model. Unlike the one designed in McCulloch and Pitts (1943), it rejects tools of symbolic logic.

> Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations. The need for a suitable language for the mathematical analysis of events in systems where only gross organization can be characterized, and the precise structure is unknown, has led the author to formulate the current model in terms of probability theory rather than symbolic logic (Rosenblatt (1958), pp. 387-388).

Lets briefly sketch the model.

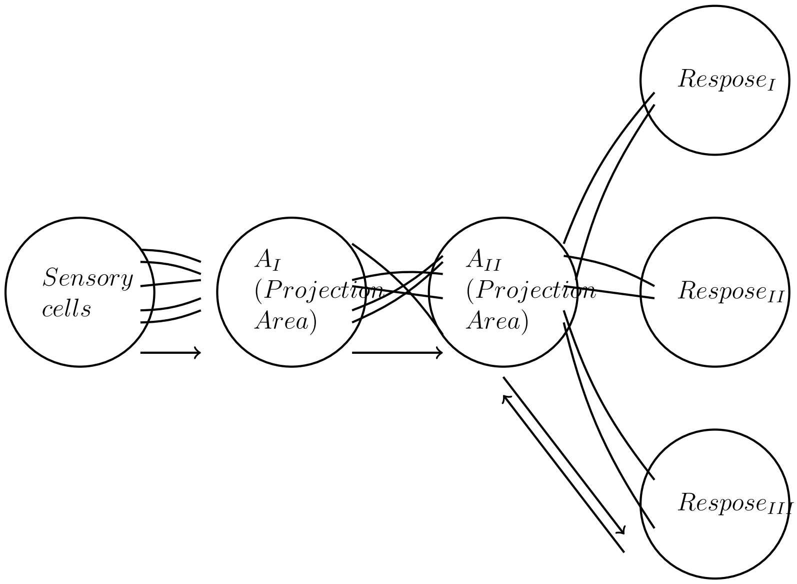


Figure 5.4: Frank Rosenblatt's artificial neural network

Two values are important to understand the system. The first one is P_a, the activation value in A by a stimulus. The second one is, P_c, "the conditional probability that an A-unit which responds to a given, stimulus, S_1, will also respond to another given stimulus, S_2." (Rosenblatt (1958), p. 392)

$$P_a = \sum_{e=\theta}^{x} \sum_{i=\theta}^{min(y,e-\theta)} P(e,i)$$

$$P_c = 1/P_a \sum_{e=\theta}^{x} \sum_{i=e-\theta}^{y} \sum_{l_e=0}^{e} \sum_{l_i=0}^{i} \sum_{g_e=0}^{x-0} \sum_{g_i=0}^{y-i} P(e,i)$$

(Rosenblatt (1958), p. 393)

In standard version a perceptron is a machine that first transforms a vector into a scalar through weighted summation. Then, it evaluates whether the value exceeds a certain threshold or not. Below is the standard version of perceptron defined by Marvin Minsky and Seymour Papert.

> Let $\Phi = \{\phi_1, \phi_2, ..., \phi_n\}$ be a family of predicates. We will say that Ψ is linear with respect to Φ fi there exists a number θ and a set of numbers $\{\alpha_{\phi_1}, \alpha_{\phi_2}, ..., \alpha_{\phi_n}\}$ such

that $\Psi(X) = 1$ if and only if $\alpha_{\phi_1}\phi_1(X) + \cdots + \alpha_{\phi_n}\phi_n(X) > \theta$. The number θ is called threshold and the α's are called the coefficients or weights.... $\Psi(X) = 1$ if and only if $\sum_{\phi\in\Phi} \alpha_\phi \phi(X) > \theta$ (Minsky and Papert (1969), p. 10).

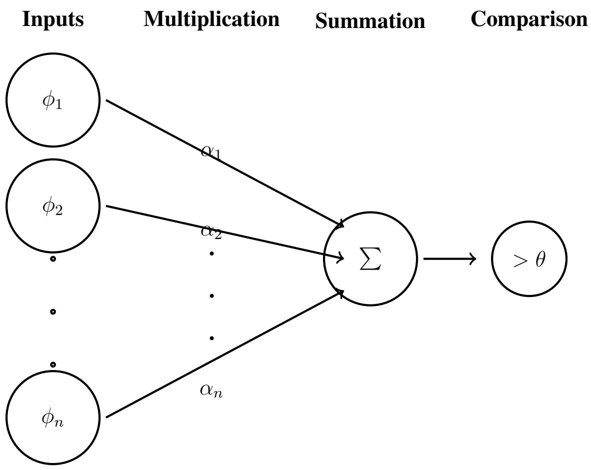


Figure 5.5: A perceptron

Linear algebra representation of these calculations is below.

$$f(x) = \begin{cases} 1 & \left(\begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_n \end{bmatrix} \bullet \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \right) > \theta \\ 0 & \text{ELSE} \end{cases}$$

An example is that we can represent the boolean function $\wedge(p, q)$ and $\vee(p, q)$ in this way. Suppose we use 1 for True and 0 for False. Suppose $\alpha_1 = 1, \alpha_2 = 1$.

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

In this case, when $\theta = 1$, the function is $\wedge(p, q)$. Similarly, if θ is set to 0, then the function is $\vee$.

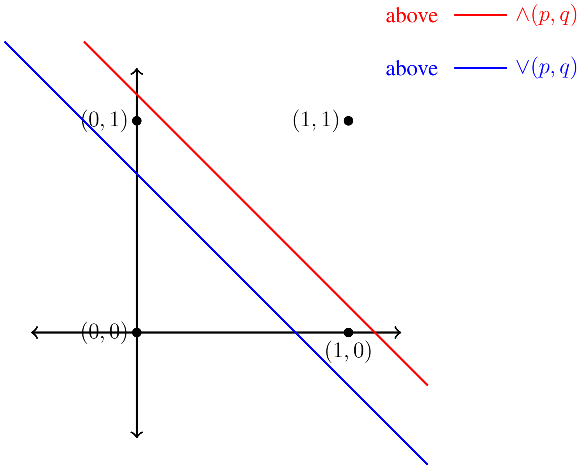


Figure 5.6: Two dimensional representation of the XOR function

By finding a relevant set of parameters it seems possible to imitate many useful functions. For this reason, Rosenblatt claims that “the theory reported here clearly demonstrates the feasibility and fruitfulness of a quantitative statistical approach to the organization of cognitive systems. By the study of systems such as the perceptron, it is hoped that those fundamental laws of organization which are common to all information handling systems, machines and men included, may eventually be understood" (Rosenblatt (1958), pp. 407-408).

However, Minsky and Papert, critics of connectionism, proved that there exist functions ϕ and ψ which can be imitated with a single layer perceptron while their combination with $\wedge$ or $\vee$ cannot be imitated with a single layer perceptron. (Minsky and Papert (1969), p. 62) Exclusive or, XOR, is an example of this theorem.

Table 5.1: Truth table for the XOR connective

p	q	$p \vee q$	$\neg(p \wedge q)$	$p\ XOR\ q\ :(p \vee q) \wedge \neg(p \wedge q)$
1	1	1	0	0
1	0	1	1	1
0	1	1	1	1
0	0	0	1	0

Below is the weighted sum of the possible inputs for the function XOR.

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \bullet \begin{bmatrix} \alpha_1 \\ \alpha 2 \end{bmatrix} = \begin{bmatrix} \alpha_1 + \alpha_2 \\ \alpha_1 \\ \alpha_2 \\ 0 \end{bmatrix}$$

There is no threshold value, θ to discriminate α_1 and α_2 from $\alpha_1 + \alpha_2$ and 0. In other words, we cannot linearly separate these two sets of values.

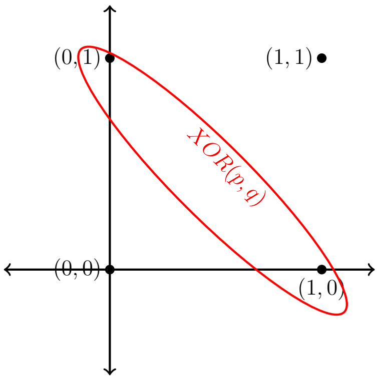


Figure 5.7: Non-linear separatin of the XOR connective

The solution is to add more layers. First we transform points to space, where second and third points can be linearly separated from the first and the last points.

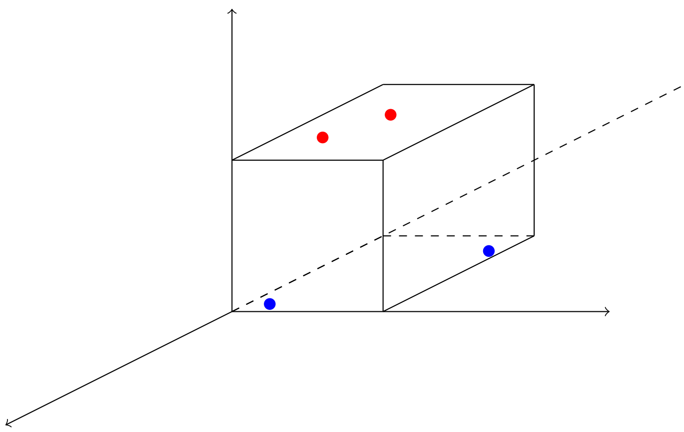


Figure 5.8: While in the 2-d space red points cannot be linearly separated from the blue ones; this can be done in a 3-d space

Minsky and Papert was aware of this solution:

Although the limitation as stated could be circumvented by adding another layer of logic to the machine scheme to permit “and" -ing two perceptrons together, this would certainly miss the point of the phenomenon. To be sure, the new machine will realize some predicates that the simpler machines could not. But if the and/or phenomenon is understood, then the student will quickly ask: Is the new machine itself subject to a similar closure limitation? We expect that no moderate extension of the machine-schema in such a direction would really make much difference to its ability to handle context-dependence" (Minsky and Papert (1969), p.228).

However, the problem is to find true parameters to solve XOR-like problems. In other words, the problem is the learning algorithm for multilayer perceptrons.

5.2.4 Backpropagation

David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams introduced a systematic, simple learning algorithm for multilayer networks. Lets briefly introduce this algorithm. Given random weights, first the output is computed. For the non-linearity part, instead of a threshold comparison, σ function, $1/(1 + e^{-x})$ is used. Then, the value of an error function that takes actual output and the desired output is as arguments is computed. By taking the derivation of the error function with respect to the weights, closest local minimum in the error graph is computed.

Forward propagation in two-layered network is computed as below.

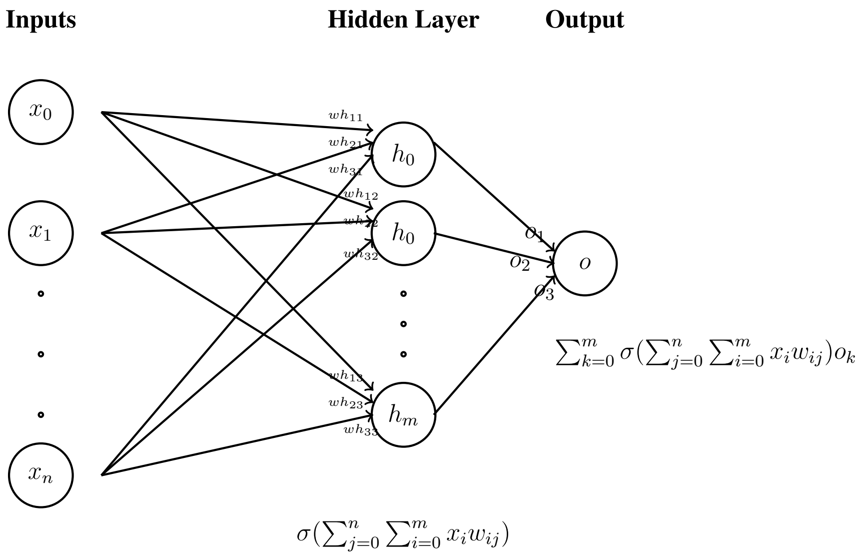


Figure 5.9: Forward propagation

The error for this example is computed with the formula

$$1/2(y - \sum_{k=0}^{m} \sigma(\sum_{j=0}^{n} \sum_{i=0}^{m} x_i w_{ij}) o_k)^2$$

where is y is the target. For more than one example, mean value of the errors is computed: $Error = 1/2\sigma_c(y_c - \sum_{k=0}^{m} \sigma(\sum_{j=0}^{n} \sum_{i=0}^{m} x_{ci} w_{ij}) o_k)^2$ “where c is an index over cases (input-output pairs)."(Rumelhart et al. (1986a), p. 534) At this point, the derivative of the error with respect to a single parameter, shows us how to change the value of the parameter in small steps, ϵ to minimize the error as it is shown in the graph below:

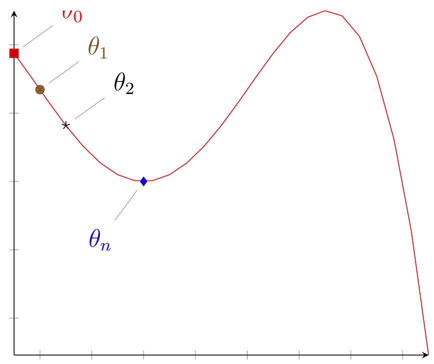


Figure 5.10: Backpropagation

In the graph above, for the initial value for the parameter θ_0, $\frac{\partial Error}{\partial \theta}$ less than 0. When we update θ_0 as $\theta_1 = \theta_0 - \epsilon \frac{\partial Error}{\partial \theta}$ we move toward a local minimum. At θ_n, since $\frac{\partial Error}{\partial \theta} = 0$ it will not be updated anymore.

It means that we are not guaranteed to find the perfect set of parameters. However, this is not a reason for being a pessimist about connectionism. “The most obvious drawback of the learning procedure is that the error-surface may contain local minima so that gradient descent is not guaranteed to find a global minimum. However, experience with many tasks show that the network very rarely gets stuck in poor local minima that are significantly worse than the global minimum." (Rumelhart et al. (1986a), p. 535)

The model sketched in Rumelhart et al. (1986a) is the basics of the most contemporary version of connectionism. However, it “is not a plausible model of learning in brains" (Rumelhart et al. (1986a), p. 536). It is rather a computational model to imitate natural systems. In contemporary works, the number of hidden layer and nodes, the choice of error and non-linearity functions, weight sharing, recurrence, algorithms to find better local minima etc., are studies empirically and mathematically.

Before finishing this part, I would like to briefly mention recurrent neural networks since they are the most relevant ones to the language processing. In Figure 5.11, x_i's are the input sequence. Each input in the sequence is processed with output from the previous step. For example, at step t_1, the input is x_1 and h_0. The output of this process is h_1. At step, t_2, the input is x_2 and the output of the previous step, h_1. To train the model, output set is compared to the expected output set. If the model is used in forecasting then the outpus may be the weather reports for each day in sequence (Elman (1990)) (Elman (1991)).

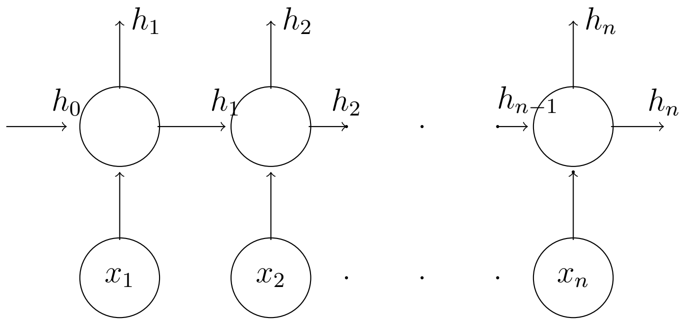


Figure 5.11: A recurrent neural network

5.2.5 Word Embeddings

Connectionism in semantic research is mainly associated with learning word vectors. Word vectors are n-dimensional real valued space representations of words. This allows us to represent the similarities between words. This roots back to the Latent Semantic Analysis, which is not a connectionist model. It is a statistical model. Latent Semantic Analysis aims to reveal hidden features of words through their distribution in a large corpus . To compute word vectors, first a term-document matrix is computed. Columns of matrix are words in the vocabulary and its rows are documents. The frequency of a word i in the document j is the value of the element w_{ij}.

Suppose our vocabulary consists of seven words, *truth, fact, meaning, time, space,* *art* and the documents are *Tractatus, Grundlagen* and *Being and Time*. Then the term-document matrix is as below.

	Tractatus	Grundlagen	Being and Time
truth	44	28	78
fact	57	39	93
meaning	31	32	91
time	15	28	95
space	13	10	47
art	5	3	8

Then, the dimensionality of the matrix is reduced to capture the core information. The linear algebra technique done in Latent Semantic Analysis is the singular value decomposition, SVD. A matrix X can be represented as $U\Sigma V^T$ where U and V are orthogonal and Σ is diagonal matrices. U is related to the words, while V is related to the features of the documents. In the diagonal matrix, numbers are ordered from greater to lesser. Greater numbers provide more information. By using first n columns of both U and V, n-dimensional matrices of words and documents are computed respectively. These matrices provide us with the core information (Dumais (2004), pp. 192-195).

$$\begin{bmatrix} 44 & 28 & 78 \\ 57 & 39 & 93 \\ 31 & 32 & 91 \\ 15 & 28 & 95 \\ 13 & 10 & 47 \\ 5 & 3 & 8 \end{bmatrix}$$

$$=$$

$$\begin{bmatrix} -0.44 & 0.34 & 0.40 & -0.31 & -0.64 & -0.06 \\ -0.54 & 0.56 & -0.26 & 0.49 & 0.27 & -0.04 \\ -0.48 & -0.17 & -0.30 & -0.70 & 0.38 & 0.017 \\ -0.46 & -0.70 & -0.16 & 0.38 & -0.34 & 0.00 \\ -0.23 & -0.20 & 0.80 & 0.13 & 0.48 & -0.02 \\ -0.04 & 0.04 & 0.04 & 0.011 & -0.02 & 0.99 \end{bmatrix} \begin{bmatrix} 209.90 & 0.00 & 0.00 \\ 0.00 & 31.76 & 0.00 \\ 0.00 & 0.00 & 5.63 \end{bmatrix} \begin{bmatrix} -0.36 & -0.31 & -0.88 \\ 0.90 & 0.13 & -0.42 \\ 0.24 & -0.94 & 0.23 \end{bmatrix}$$

$$WordMatrix = \begin{bmatrix} -0.44 & 0.34 \\ -0.54 & 0.56 \\ -0.48 & -0.17 \\ -0.46 & -0.70 \\ -0.23 & -0.20 \\ -0.04 & 0.04 \end{bmatrix} \begin{bmatrix} 209.90 & 0.00 \\ 0.00 & 31.76 \end{bmatrix} = \begin{bmatrix} -93.17 & 10.89 \\ -114.45 & 17.81 \\ -101.15 & -5.68 \\ -97.66 & -22.27 \\ -49.14 & -6.51 \\ -9.77 & 1.57 \end{bmatrix}$$

$$DocMatrix = \begin{bmatrix} -0.36 & 0.90 \\ -0.31 & 0.13 \\ -0.88 & -0.42 \end{bmatrix} \begin{bmatrix} 209.90 & 0.00 \\ 0.00 & 31.76 \end{bmatrix} = \begin{bmatrix} -75.80 & 28.58 \\ -64.62 & 4.22 \\ -184.76 & -13.20 \end{bmatrix}$$

Figure 5.12: Singular Value Decomposition technique for the semantic analysis

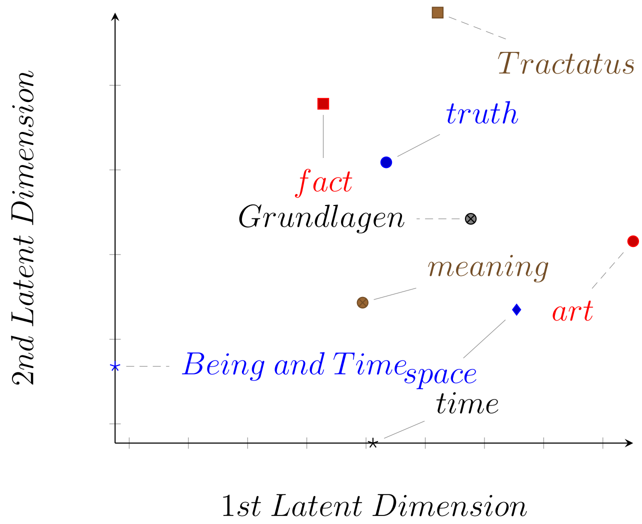


Figure 5.13: Latent semantic analysis through Singular Value Decomposition technique

Even though this approach represents some statistical features of the data, it does not go deep into the semantic relations in the data. Neural networks are used to extract more detailed semantic information from the data. To achieve this goal, two popular techniques, Continuous Skip-gram and Continuous Bag-of-Words models, are introduced in Mikolov et al. (2013a) and Mikolov et al. (2013b). The aim is to predict surrounding words from a given center word or reversely to predict center words from the surrounding words.

Suppose the data is the expression, “Tarski does not believe that snow is not white" and the surrounding words consists of two words on the left and right of center words. Then in the Continuous Skip-Gram model, given the word $snow$, the aim is to predict the surrounding words, { believe, that, is, not}. On the other hand, Continuous Bag-of-Words model aims to predict the center word $snow$ given the surrounding words, { believe, that, is, not}. Vectors that gives us true predictions are assigned to words. For example in the Skip-gram model the aim with vectors is to maximize the probability of a surrounding word given a center word.

$$\frac{1}{T}\prod_{t=1}^{T}\prod_{-c\leq j\leq c, j\neq 0} p(w_{t+j}|w_t)$$

where the conditional probability function is computed with the softmax function:

$$p(w_{context}|w_{center}) = \frac{e^{({v`_{w_{context}}}^T v_{w_{center}})}}{\sum_{w=1}^{W} e^{({v`_{w}}^T v_{w_{center}})}}$$

where W is the vocabulary size and v and $v`$ are center and context word vectors respectively.

Resulting word vectors at the end of the backpropagation process are able to represent “multiple degrees of similarity" (Mikolov et al. (2013a), p. 2). For example, $vector(“King”) - vector(“Man”) + vector(“Woman”)$ results in a vector that is closest to the vector representation of the word $Queen$. (Mikolov et al. (2013a), p. 2) Similarly, $vector(“China”) - vector(Bejing) \approx vector(“Poland”) - vector(“Warsaw”)$. (Mikolov et al. (2013b), p. 3115)

In another work, Pennington et al. (2014), a combination of Latent Semantic Analysis and the Skip-gram model combined to get better results efficiently (Pennington et al. (2014), pp. 1532, 1533). The paper aims the shed light on “how meaning is generated from these statistics, and how the resulting word vectors might represent that meaning." (Pennington et al. (2014), p. 1533) In sum the aim is to minimize the cost function, J, below:

$$J = \sum_{i,j=1}^{V} f(X_{ij})({v_i}^T v`_j + b_i + b`_j - logX_{ij})^2$$

where x_{ij} is co-occurrence of ith and jth words, b is the bias and

$$f(x) = \begin{Bmatrix} (\frac{x}{x_{max}})^{3/4} & if x < x_{max} \\ 1 & \text{otherwise} \end{Bmatrix}$$

“However, these approaches for learning word vectors only allow a single context-independent representation for each word" (Peters et al. (2018)) (Peters et al. (2017)) (Arora et al. (2018), p. 483) (Iacobacci et al. (2015), p.95). For example, even though the word “bank" has a single vector representation, it has different senses. For this reason, several contextual word representations are introduced. In bidirectional recurrent neural language models, (Peters et al. (2018), Peters et al. (2017)), in addition to context independent word representations, context dependent vector representations of the words are introduced. Given a set of parameters θ a bidirectional recurrent neural network propagates forwardly to predict the next word from both left to right and right to left order. Therefore, given an $N-$word sentence, and the set of parameters, θ, which contains context independent word vectors, the aim is to maximize the value of the likelihood below:

$$\prod_{k=1}^{N}(p(word_1|word_1, \cdots, word_{k-1}; \theta) * p(word_k|word_{k+1}, \cdots, word_N; \theta))$$

(Peters et al. (2018), p. 3) (Peters et al. (2017), p. 3)

At the end, each word is represented with a context independent vector, x_k and both left and right context dependent representations at each layers of the deep neural network. In short, an $L - layered$ bidirectional recurrent neural network represents kth word, R_k as $\{x_k, \overrightarrow{h_i^j}, \overleftarrow{h_i^j} | j = 1, ..., L\}$ Contextual representation of the word is the concatenation of $\overrightarrow{h_i^j}$ and $\overleftarrow{h_i^j}$. It is simply written as h_i^j. (Peters et al. (2018), p. 3) (Peters et al. (2017), pp. 3-4) For example, in Figure 5.14 below, h_3 is the contextual representation of the word “shaky."

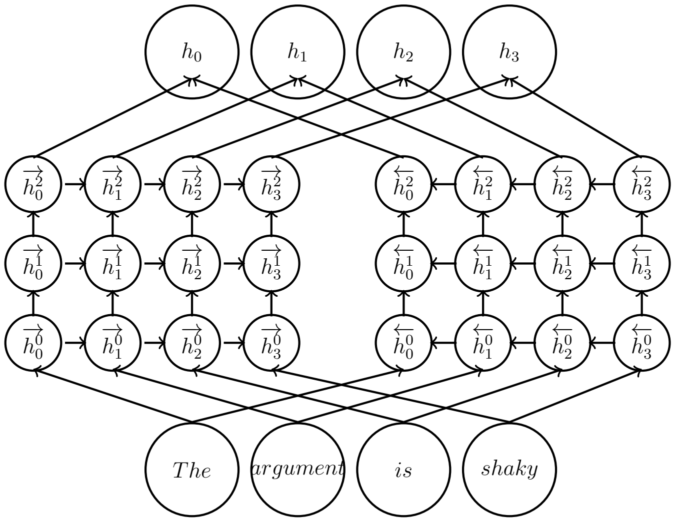


Figure 5.14: Contextual word representation with a bidirectional recurrent neural network

Another research, Arora et al. (2018), takes context independent word vectors as superpositions over contextual vectors. For example, context independent vector representation of the word “bank," v_{bank} is a superposition over different but related senses of the word *bank*.

$$v_{bank} \approx \alpha_1 sense1_{bank} + \alpha_2 sense2_{bank} + \alpha_3 sense3_{bank} + \cdots$$

where α_i's are nonnegative coefficients (Arora et al. (2018), p. 483). Contextual word vectors, $sense_i's$ are obtained as weighted sums of the “atoms of the discourse." Atoms of the discourse, $A_1, A_2, ..., A_m$, are the fundamental components of the semantic space. Linear combination of them takes us to different points on the space. That is how we arrive at the *institution* sense of the word “bank."

$$v_{bank} = \sum_{j=1}^{m} \alpha_{bank,j} A_j + \eta_{bank}$$

where η_{bank} is the error in specifying the exact position of the senses of the word "bank." Parameters (i.e., word vectors, α_i's) are optimized by minimizing the error. (Arora et al. (2018), pp. 487-488)

In the rest of this work, I will explain why connectionism is the best tool to model Meaning Eliminativism. To do this, I will highlight the syntax-semantic distinction in connectionist, sub-symbolic, approach to meaning in natural languages. The distinction involves in the philosophical interpretation of the sub-symbolic approach to the meaning.

5.3 Processing (syntactic) level

Jeffrey L. Elman states that for the sub-symbolic approach, "The effect that a given word produces is a function of two things: the prior state of the network, which encodes the context in which word input occurs; and the network's dynamical structure or grammar, which is encoded in its weights" (Elman (2009), pp. 555-556). Thus, we need *a background, a context*, and *a word representation*. The context is continuously updated upon processing a word in sequence.

Each word and the context is represented with a high dimensional vector. Their contribution to the semantic state of the system is seen after they are processed. So, they are not points in a logical or conceptual space with phenomenal dimensions *color, temperature, etc.* see (van Fraassen (1967), p. 172) (Stalnaker (1979), pp. 347-348) (Gärdenfors (1999), p. 25). This is because there is a sharp separation between the processing and the interpretation levels.

The cultural and biological background of the hearer is represented by a set of parameters by which context plus words are processed. Even though the context is continuously updated, it is not an easy task to update these parameters. When they are updated, it is called "learning." Understanding any utterance in Turkish requires a relevant set of parameters to process a context vector and a sequence of word vectors. This confirms Wittgenstein's dictum: "Understanding a sentence means understanding a language" (Wittgenstein (1965), p. 5).

"It is the numerical values comprising the vector... that really drive the machine"

(Smolensky (1995a), p. 190). For Smolensky, since these are real-valued vectors, “many of the concepts used to understand cognition in the sub-symbolic paradigm come from the category of continuous mathematics, while those used in the symbolic paradigm come nearly exclusively from discrete mathematics" (Smolensky (1995b), p.71) (Smolensky (1995a), p. 165) (Smolensky (1995a), p.187). As Bernard Victorri states, “continuity or discreteness are not properties of phenomena, they are characterizations of theories upon phenomena [and] one can use discrete models to represent continuous concepts, and the other way round" (Victorri (1994), p. 241). In both sub-symbolic and symbolic computations, operations are done through discrete steps in Turing/von Neumann fashion (Smolensky (1995b), pp. 43-44). However, since the end product is a real-valued vector, we have a continuous vector space to represent *meaning*(Victorri (1994), p. 248) (Touretzky (1994), p. 237). “A continuous space is the natural frame in which qualitative properties of dynamical systems can be handled, and nothing more" (Victorri (1994), p. 243). In the mathematical sense, we do not need to reach every single point in the continuous space. Instead, any computer, including the human mind, do discrete operations to reach *some points* in this continuous space. In other words, the so-called *analog digital* dichotomy (see Churchland (1995), pp. 242-243) is not the case here. Both symbolic and subsymbolic approaches are conveyed in a digital fashion (Eliasmith and Bechtel (2006), p. 3). What makes a computer (e.g., a human mind) better is its capability to do more fine-grained discrete operations. Continuous space must be the target because it is the only way to deal with the growing complexity of meaning. The history of *the ratio of a circle's circumference to its diameter* shows that better approximations of the *target*, the transcendental number π, helps us represent the world better. Turing remarks that the discrete system here approximates the value π probabilistically (Turing (1950), pp. 451-452). However, it does mean that either that ratio or the perfect circle exists in the physical world. Similarly, continuous semantic space, I believe, is nothing but a guide for researchers to deal with the highly complex nature of meaning. The parameters to processing language lending amenability to computation are highlighted in Figure 5.15.

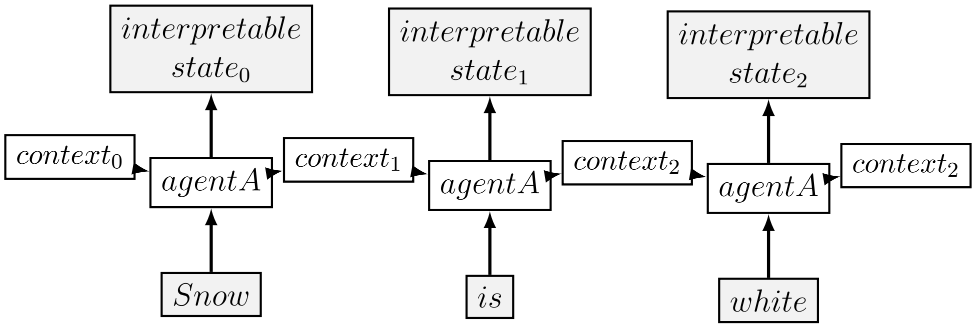


Figure 5.15: Processing level

To note, what we are seeking is not word vector models that locate the meaning of a word in the semantic/conceptual space through word co-occurrence frequency (Mikolov et al. (2013a)) (Pennington et al. (2014)) (Hofmann (2013)) (Dumais (2004)). Neither are we interested in sentence vector models that bridge denotational aspects of compositional, formal semantics with contextual aspects of distributional semantics (Ettinger et al. (2018), pp. 1790-1792) through using categorial grammar (Lewis (2019)), lambda calculus (Sadrzadeh and Muskens (2019)), and Fregean function application (Baroni et al. (2014)). In these models, vectors working at the lower level are encapsulations of meanings working at the higher level. This is what Smolensky and Fodor dispute over (Fodor and Pylyshyn (1995), pp. 146-148) (Smolensky (1995a), pp. 166-169). In Smolensky's account, "there is no account of the architecture in which the same elements carry both the syntax and the semantics. Thus, we have a fundamentally new candidate for the cognitive architecture which is simply *not* an implementation of the classical one"(Smolensky (1995a), pp. 168-169). On the other hand, implementing word meanings with vectors would be "to admit that subsymbols are themselves irrelevant for modeling cognition, and that nothing is gained by turning to subsymbols for a new and different understanding of cognition" (Eliasmith and Bechtel (2006), p. 5).

For connectionists, the difference between the symbolic model and the sub-symbolic model is "like that between quantum and classical mechanics (Smolensky (1995b), p. 55) or that "between the Ptolemaic and the Copernican/Keplerian accounts of planetary motion" (Churchland (2007) p. 136). In Kuhnian terms, sub-symbolic computation is a different paradigm than the symbolic one (Schneider (1987)). The opposition between atomism/corpuscularianism and Aristotelian hylomorphism according to which any change is reduced to four elements (*earth, air, fire, and water*) with *various forms* analogous to the distinction between sub-symbolic approach and the symbolic, formal compositional one. In Aristotle's physics, just as in compositional semantics, natural phenomena are explained with these macro-level objects (Ainsworth (2020)). Aristotle claims that the atomist philosophy of Democritus and Leucippus did not "give any account of combination; and they neglected almost every single one of the remaining problems, offering no explanation, e.g., of 'action' or 'passion'– how in physical actions one thing acts and the other undergoes action" (McKeon (2001), *On Generation*, 315a34-315b15). Just as a formal semantician's sophisticated tools that she uses to explain language understanding with word meanings, Aristotle's theory of *change* postulates a sophisticated set of formal tools (e.g., four causes). However, through seventeeth century corpuscularian philosophy and the developments in atomic theory in nineteeth and twentieth centuries, Aristotelian ontology, with its macro-objects (i.e., *water, earth, fire, air*), macro-qualities, *e.g., hot, cold*, etc., and its rational explanation (four causes) disappeared. Aristotle's macro-level explanations do not encapsulate the micro-level explanations we have today. Rather, Aristotle's macro-level approach is just *pointless*. It is in vain to explain the whole complexity of the universe with Aristotle's macro-ontology. Nevertheless, as Smolensky associates the symbolic approach with Aristotelian logic (Smolensky (1995a), p. 165), Aristotelianism is still alive in human sciences. Up to now, Aristotelianism might be 'the only game in town' (Fodor (1998), p. 23), but as Smolensky adds, "we'll need to be playing other games before too long" (Smolensky (1995a), p. 191).

In this respect, Fodor says, "there is a scientific story to be told about quarks; and a scientific story to be told about atoms; ... ditto rocks and stones and rivers" (Fodor and Pylyshyn (1995), p. 95). I agree that there is a scientific story to be told about rocks, but I do not think that the complexity of the earth can be precisely explained by an ontology that consists of *rocks* and *stones*. Just as *rocks* and *stones*, word meanings are too coarse-grained to explain the complex nature of language understanding. Analogously, Daniel Dennett's *intentional strategy* of "first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose" (Dennett (1997), p. 61) is not a substitute for the *physical strategy* according to which “if you want to predict the behavior of a system, determine its physical constitution (perhaps all the way down to the microphysical level) and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the outcome for any input" (Dennett (1997), p.60). Of course, the intentional strategy can be used when the physical strategy is “practically inaccessible" (Dennett (1997), p. 61). However, they do not have the same precision. In the same way, formal, compositional semantics with a handful of rules do not have the same precision as the subsymbolic approach in representing the metaphorical, polysemous, flexible, and dynamic nature of natural languages. As Smolensky says, “sub-symbolic models accurately describe the microstructure of cognition, whereas symbolic models provide an approximate description of the macrostructure" (Smolensky (1995b), p. 55).

Fuzzy, vague, indeterminate semantic phenomena emerge at the surface level because our macro-ontology *objects, predicates, etc.* that we use to describe "meaning" does not carve up the semantic space. Just as Aristotelian ontology with *earth, fire, water, air*, and forms (e.g., *chairs, tables, knives*) does not encapsulate the micro-objects *atoms, electrons, quarks* ontology; in semantics, macro decisions do not encapsulate but *only approximate* the micro-decisions (Rumelhart et al. (1986b), p. 56) (Smolensky (1995b), pp. 54, 75-76) (Rumelhart et al. (1986b), p. 56) (Smolensky (1995a), pp. 184-187). My description of a building’s collapse is only a weak approximation of what is going on. Similarly, the compositional account of the sentences below is only a very bad approximation of the way that an average language user understands it:

> [The land of pure understanding] is the land of truth (a charming name), surrounded by a broad and stormy ocean, the true seat of illusion, where many a fog bank and rapidly melting iceberg pretend to be new lands..." (Kant (1998), B295-A236).

> Ideas too sometimes fall from the tree before they are ripe (Wittgenstein (1977), p. 27e).

Robert Boyle, one of the leading critics of Aristotelian physics, asks, For if with the same bricks, differently put together and ranged, several bridges, vaults, houses, and other structures may be raised merely by a various contrivance of parts of the same kind, what a great variety of ingredients may be produced by nature from the various coalitions and contextures of corpuscles that need not be supposed, like bricks, all of the same size and shape, but to have, both in the one and the other, as great a variety as could be wished for?" (Boyle (2009), p. 314)

Similarly, a syntactic system working with corpuscles may account for rich semantic states. Using vectors at the syntactic level is one way to solve this problem.

5.4 Interpretation level

At the processing level, the system outputs a vector, a point in a high dimensional space. Instead of identifying this output with propositional content, I propose a more holistic model. Davidson says,

when we try to say what a metaphor 'means,' we soon realize there is no end to what we want to mention... How many facts are conveyed by a photograph? None, and infinity, or one great unstatable fact? Bad question. A picture is not worth a thousand words, or any other number. Words are the wrong currency to exchange for a picture" (Davidson (2001d), p. 263).

As we have seen above, because of the polysemous nature of natural languages, we should treat the whole language the way Davidson treats metaphors. Hence, rather than representing all semantic value with propositional forms (i.e., with relationships, predicates, quantifiers, etc.), it is better to extract relevant semantic information, just as extracting information from a photograph. The extraction mechanism is in tandem with the dispositionalist theories of *belief*.

Quine notes, "for all the liveliness of fluctuation of beliefs, believing is not an activity... Rather, believing is a disposition that can linger latent and unobserved. To believe that Hannibal crossed the Alps is to be disposed, among other things, to 'Yes' when asked" (Quine and Ullian (1970), pp. 9-10). In our case, understanding/processing a linguistic expression is to be disposed, among other things, to provide semantic information when it is asked.

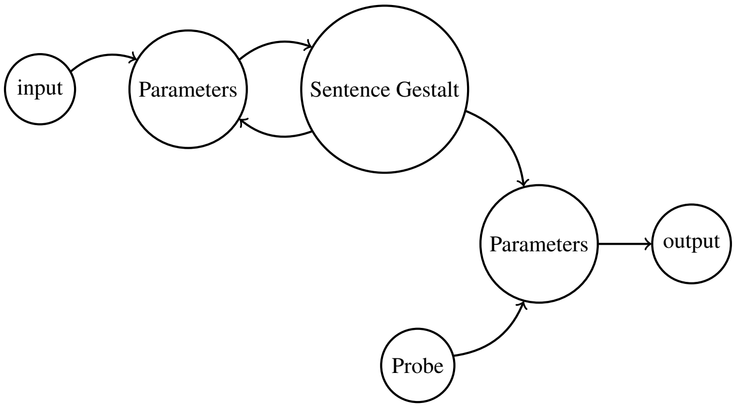


Figure 5.16: McClelland model

An example of this approach is David McClelland's "sentence gestalt model" in Figure 5.16 (McClelland et al. (1989)). In this model, the output of the processing level, *the sentence gestalt*, allows comprehender "to respond correctly when probed in various ways." For example, the system should give correct answers to the following questions: Who is the agent? What is the theme? And so on (McClelland et al. (1989), pp. 296-298).

In the same way, Figure 5.17 illustrates the toy version of the position I defend. In this toy model, "The mother of science is philosophy" and "Her mother likes music" contain the same work, "mother" with different but related uses. In other words, it is example of polysemy. In this simplified model, each word and the initial context, $context_0$ corresponds to a vector. They are processed through a recurrent neural network. Best parameters are the ones that reflect fine shades of meanings. For example, the difference and the relation between different uses of the word "mother" must be reflected by the system. Backpropagation technique that we discussed in section Perceptron can be used learn a good set of parameters.

In this model, we do not need, meaning of the words, semantic combination rules which trace the surface syntax etc. All the process is done at the micro level. At the end, sytems outputs a vector. The output value controls the behavior of the system. As it is seen in the figure, output value is where we can extract required information.

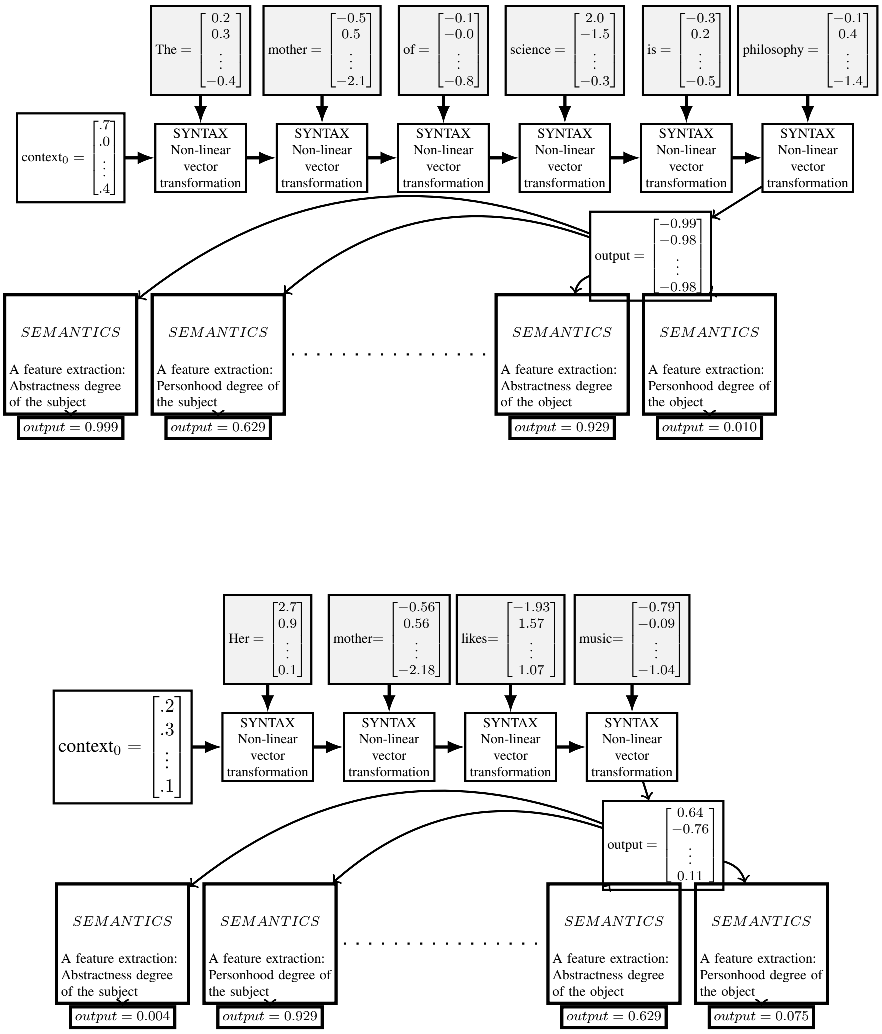


Figure 5.17: A toy example that illustrates how we understand “The mother of science is philosophy" and “Her mother likes music" are computed

CHAPTER 6

FURTHER ISSUES

6.1 Normativity in language use

The approach I defend, as a naturalist one, confronts the normativity issue. According to the normativity of meaning thesis, "red" means RED is another way of saying we ought to use "red" only for red objects. The reason behind this is thought to be the correctness in language use. Kripke's argument against dispositionalism summarizes the normativity claim.

> Suppose I do mean addition by '+'. What is the relation of this supposition to the question how I will respond to the problem '68+57'? The dispositionalist gives a descriptive account of this relation: if '+' meant addition, then I will answer '125'. But this is not the proper account of the relation, which is normative, not descriptive. The point is not that, if I meant addition by '+', I will answer '125', but that, if I intend to accord with my past meaning of '+', I should answer '125'. Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be disposed to respond as I should, but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is normative, not descriptive. (Kripke (1982), p. 37)

Even though philosophers agree on the correct/incorrect distinction (Glüer and Wikforss (2015), p. 66), there is no consensus on whether normativity follows. It consists of sentences in "S ought to do X" form. "S ought to mean RED by 'red'" is an instance of the normativity of meaning thesis. However, it is not justified why people ought to use words correctly (Glüer and Wikforss (2015)) (Boghossian (2005)) (Hattiangadi (2006)).

> Kripke says: If I mean addition by '+' then it doesn't follow that I will say that '68+57=125', but only that I ought to say that it does. But it seems to me that neither claim follows. In particular, the ought claim doesn't follow because, even

> though I mean addition by '+' and know therefore that it would only be correct to say that '68+57', I might still not choose to say it because I might deliberately not choose to say what I know to be correct. Deciding knowingly to assert what is false is not to undermine the very possibility of assertion (Boghossian (2005), p. 212)

While the normativity thesis is controversial, there is more agreement on the correct/incorrect distinction. Now the question is how the system I defend follows the correct/incorrect distinction. The behavior of the system I propose is determined solely by a large number of parameters. Does it mean that correct use/understanding of linguistic expressions is determined by the behavior of the system? Is correctness trivial in such a system? In other words, there is no incorrect use in it?

To answer these questions, let's first see how we use concepts like correctness, mistake, error, etc. When we talk about the behavior of a calculator, a computer, a dish machine, or an automobile, we use these concepts to evaluate their behavior. On the other hand, when we talk about the behavior of a planet, a star, a quark, etc., we don't use these words. It is not just that we don't use these concepts, but we should not use them when talking about a star. A star does not give an error, or a quark does not make mistakes. On the other hand, there are some entities for which we use these concepts only depending on the situation. For example, for humans, we talk about misunderstanding, miscalculating, misbehaving, etc. We believe that it is correct to answer 1001 when we are asked to calculate $1000 + 1$. Similarly, in IQ tests, there are correct and incorrect answers for pattern-driven questions. If a child cannot correctly apply the word "green" to green objects, we believe there must be some problem with his brain. However, when a biologist studies a human body, she does not use normative concepts. At most, she may talk about statistical deviance. The reason is that human beings are both natural systems like stars and cultural objects like calculators and televisions. Human linguistic behavior as a natural system, I think, is exempt from error. However, language as a cultural, social construct can be correctly or incorrectly used. For this reason, my aim is not to give metaphysical justification of the notion of correctness in linguistic behavior. Instead, my aim is to show how a connectionist system I advocate explains these observations.

Given an input (e.g., "pick up colored objects," "bring me a coffee," etc.), a connectionist system (e.g., human) outputs a set of behaviors. Since a linguistic system is part of a community, her responses are evaluated by community members. She constantly gets feedback from other members. For example, if the agent picks up rectangular objects when asked to pick up red ones, she gets negative feedback. If the feedback updates her behavior in the way other members endorse, she adapts to the community. She becomes a part of that linguistic community. Otherwise, she may be isolated. In other words, instead of a primitive notion of correctness, we have ideas like parameter update and social adaptability.

In one sense, it is similar to Merril B. Hintikka and Jaakko Hintikka's interpretation of Wittgenstein's language games. According to this interpretation, there are two kinds of language games: Primary and secondary language games. Natural expressions of a child that state her sensations is an instance of the primary language game. The relation between pain and a certain pain expression in a primary language game is not epistemological but grammatical. It is incorrigible. However, the relation is not necessary, but it is, indeed, contingent. The reason is that a player of this game, say a child, may lie. At this point, we have the secondary language game. In the secondary language game, the relation is between different language games. A secondary language game is corrigible. There is the possibility of error in a secondary language game. While moves in a primary language are spontaneous, there are rules in a secondary language game. For example, a past tense talk of pain in a secondary language game brings persistence criteria (Hintikka and Hintikka (1986), pp. 279-286) (Hintikka (1996), pp. 338–340). The similarity between my approach and Hintikkas' Wittgenstein is the smooth transition from the natural structure of the linguistic agent to its social role. Hintikkas' quote from *Philosophical Investigations* shows the naturalist start.

> Why can't a dog simulate pain? Is he too honest? Could one teach a dog to simulate pain? Perhaps it is possible to teach him to howl on particular occasions as if here were in pain, even when he is not. But the surroundings which are necessary for this behaviour to be real simulation are missing (Wittgenstein (1953), §250).

A linguistic system (e.g., a robot, an animal, etc.), just as a dog, is determined by its parameters. In the absolute sense, there is no correct/incorrect behavior distinction. However, a linguistic system as a social being needs to have good adaptation skills. I will use the following toy model to illustrate this mechanism. Consider Mary as the subject, and Socrates, Plato, and Aristotle as members of her linguistic community. The following five steps summarize her adaptation on the basis of correct/incorrect language use.

1. get inp0: {input0Socrates, input0Plato, input0Aristotle}
2. (parameters0, inp0, context0) $\rightarrow$ output1
3. get feed0: {feedback0Socrates(output1), feedback0Plato(output1), feedback0Aristotle(output1)}
4. evaluate feed0
5. evaluate(feed0) $\rightarrow$ update(parameters0 to parameters1)

So, the answer is straightforward; there is no correctness in the absolute sense. Instead, there is the concept of social adaptation. Social adaptation can be modeled with a connectionist architecture. For example, deep reinforcement learning models can be used for this task. Given the input, "pick up colored objects", Mary acts in a certain way, a. There is a policy, π, that governs Mary's behavior. The policy assigns values to possible actions given Mary's state. For her actions, she gets a reward from other agents. If her action, a, is endorsed, the reward is a positive value. Otherwise, it is a negative value. The aim is to maximize the reward value. The rest is the optimization problem. Once the parameters are optimized, Mary joins the linguistic community. Otherwise, she is isolated from the linguistic community.

The reason I give such a simple answer is that I do not postulate meanings. Once we have meanings, there come questions about their identity conditions. The presence of meanings blocks suck simple solutions with objections like the following: This may explain linguistic behavior, but what about meanings themselves? For example, the answer does not tell us what makes the meaning of “green” self-identical, persistent over time. Since I do not postulate meanings, I do not have to deal with such questions. In my approach, there are inputs, outputs and in between them, there are millions of equations free from meanings. Additionally, there are feedbacks from other members of the society which trigger learning and adaptation processes.

6.2 The problem of explanation

In explaining language understanding, I do not postulate macro entities like meaning, meaning components, representations, etc. In other words, an ordinary English language user's understanding of James Baldwin's words "home is not a place, but simply an irrevocable condition" is not explained through meaningful concepts associated with *home, is, negation,...* etc. though that would be a good literature analysis. Instead, each word is considered nothing but physical entities with smaller physical parts. For this reason, they are represented with vectors. Parameters to process these inputs are represented with huge dimensional matrices. The output is a result of arithmetical operations on these matrices. Suppose we have a set of parameters that approximates an ordinary language user's behavior. Can we say that there is no explanation even though it is successful in terms of prediction? For Chomsky, the answer is yes. He criticizes approaches like mine as having a new notion of success that deviates from the traditional scientific practice. (Chomsky (2011)) (Norvig (2012))

> Take the study of bee communication. According to this conception, the way it is generally conducted is seriously flawed. Instead of difficult experiments devising circumstances that never occur in nature – say, having bees fly to flowers on an island (see Gallistel, this issue) – bee scientists should be carrying out statistical analysis of massive collections of videotapes of bees swarming, achieving greater and greater success in approximating the videotapes, and getting a tolerably good prediction of what is likely to happen next, doubtless better than bee scientists could give (or would care about). Perhaps physics should be revised the same way. No balls rolling down frictionless planes and other such abstractions and idealizations that have virtually defined the subject for centuries: rather, extensive statistical analysis of videotapes of leaves blowing in the wind and other natural events, which will surely give more successful predictions of what will happen outside the window that what the physics department can provide" (Chomsky (2011), p. 266).

I disagree with Chomsky. As Norvig states, the approach Chomsky criticizes is part of the scientific practice. As "languages are complex, random biological processes that are subject to the whims of evolutions and cultural change," "what constitutes a language is not an eternal ideal form, represented by the settings of a small number of parameters, but rather is the contingent outcome of complex processes" (Norvig (2012), p. 33). By appealing to Leo Breiman's distinction of two different statistical cultures, Norvig claims that there are two kinds of scientific practices: simple and complex models.

For Breiman, there are two different cultures of analyzing data:

1. The Data Modeling Culture

2. The Algorithmic Modeling Culture

In the first approach, data is modeled using relatively a few parameters. In other words, the model is a very simple one. Chomsky's Minimalist Program and its precursors can be considered examples of this approach. Models are ideal, simple and the phenomena they aim to explain are noisy.

The second approach is Algorithmic Modeling Culture, according to which "nature produces data in a black box whose insides are complex, mysterious, and at least, partly unknowable. What is observed is a set of x's that go in and a subsequent set of y's that come out. The problem is to find an algorithm $f(x)$ such that for future x in a test set, $f(x)$ will be a good predictor of y" (Breiman (2001), p. 205). For this approach, the nature that produces outputs is complex and inscrutable. To deal with it, we need to find tools that are "only slightly less inscrutable than nature's" (Breiman (2001), p. 209). On the other hand, Breiman criticizes using simple models to deal with complex phenomena referring to an old saying: "If all a man has is a hammer, then every problem looks like a nail" (Breiman (2001), p. 204).

According to the distinction, my approach falls in the second category, the Algorithmic Modeling Culture. Let me now show that there is no explanation problem in an approach with a massive number of parameters. Consider how Newton's laws of motion and law of universal gravitation explain Kepler's three laws. Let's see the derivation of Kepler's third law from Newton's system. Kepler's third law: "The periods of the planets are proportional to the $3/2$ powers of the major axis lengths of their orbits" (Hugh D. Young (2007), p. 397)

Its application on Earth's orbiting around Sun can be explained through Newton's laws.

<u>Newton's law of universal gravitation:</u>

$F_{NET} = (G \cdot mass_{sun} \cdot mass_{earth})/r^2$

Newton's second law of motion:

$F_{NET} = d\,momentum/dt = mass_{earth} \cdot a$

The derivation:

$(G \cdot mass_{sun} \cdot mass_{earth})/r^2 = mass_{earth} \cdot a$

$(G \cdot mass_{sun} \cdot mass_{earth})/r^2 = mass_{earth} \cdot v^2/r$

$(G \cdot mass_{sun} \cdot mass_{earth})/r^2 = mass_{earth} \cdot (2\pi r/T)^2/r$

$(G \cdot mass_{sun})/r_2 = 4 \cdot \pi^2 \cdot r/T^2$

$T = (4 \cdot \pi^2/G \cdot mass_{sun}) \cdot r^{3/2}$

$T \propto r^{3/2}$

The derivation shows that Kepler's third law or its application to Earth's orbit around Sun can be explained through Newton's second law of motion and his universal law of gravitation. Since there are a few equations, it seems easy to interpret them. However, epistemologically we do not need to be familiar with concepts like gravitation or mass. The familiarity is a psychological issue. I am not even sure that I really understand these concepts. However, I know how to use these concepts in equations. What we need is to be able to do measurements and use these concepts in equations. Consider the following case. Three students give correct answers to a hundred physics questions. The first one uses a few equations. The second one uses a system with one million equations. The third one does not use any equations but makes accurate predictions by chance or mysterious powers. In this situation, the third student's approach is not scientific, even though she makes good predictions. It is because there is no explanation. However, we can't think of the same problem for the second student. The first two students are both scientific and explanatory, even though the second one's equations are far from interpretable.

Interpretability is not a mandatory element of scientific practice. Depending on the practice, researchers may or may not need it. As Breiman says, "there is no way they [doctors] can interpret a black box containing fifty trees hooked together. In a choice between accuracy and interpretability, they'll go for interpretability" (Breiman (2001), p. 209). Interpreting every single variable is not what I am interested in. It does not mean that there should be only one approach to language understanding. Just as doctors, other researcher on language depending on their interest, in a choice between accuracy and interpretability, may go for interpretability. To sum up, even though there is less interpretability in my approach, there is the explanation. Parameters in the form below explain the linguistic behavior.

$$\hat{y} = h(\cdots g(\begin{bmatrix} a_{11} & a_{12}\cdots & a_{1m} \\ a_{21} & a_{22}\cdots & a_{2m} \\ \cdots & & \\ a_{n1} & a_{n2}\cdots & a_{nm} \end{bmatrix} \bullet f(\begin{bmatrix} w_{11} & w_{12}\cdots & w_{1m} \\ w_{21} & w_{22}\cdots & w_{2m} \\ \cdots & & \\ w_{n1} & w_{n2}\cdots & w_{nm} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix})))$$

where $\hat{y}$ is the predicted value and $\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{bmatrix}$ is the input.

6.3 Modeling Simple Inferences

In addition to its complex, dynamic, holistic side, human language has relatively simple, regular sides such as inference, causal reasoning, compositionality, etc. The eliminativist approach I propose is not based on meanings. Rather, it is based on a massive number of parameters. How effective would be such a complex system on simple, rule-governed, precise abilities? There are different types of inference rules. These are deductive, inductive, and abductive rules. Inductive and abductive systems are not in apparent contrast with the holistic, dynamic, and context-dependent approach to natural languages. In empirical sciences or daily practices, we usually appeal to these methods. For example, if my car is wet, I would think that it probably rained yesterday. Similarly, medical doctors make inductive generalizations about eating habits and cancer. These holistic, dynamic, and context-dependent practices are where complex models, say neural networks, are successful. However, the problem is to model these high-level reasonings with a low-level approach where high-level entities are not used. Even though in the approach I defend, “rain” or “ground” does not have any meaning, there is a relation between rain and ground in the abductive reasoning example above. These relations are used in causal reasoning, pragmatic inference. For example, due to the relation between tiger and danger, the expression “there is a tiger” may be understood as a warning. The problem is more evident in another type of reasoning, deductive reasoning, which is neither probabilistic nor context-dependent. For instance, Modus Ponens is a simple, rule-governed, and exact deductive method. The truth of the premises p and $p \to q$ guarantees the truth of q. How does a neural network arrive at such exact conclusions?

First of all, it is not easy to apply deductive rules to linguistic data. In other words, it isn’t easy to find p and $p \to q$ examples in any book, journal, newspaper, or conversation. A prominent logician, Vann McGee, argues against the validity of modus ponens in natural language use. He states that modus ponens "is not strictly valid; there are occasions on which one has good grounds for believing the premises of an application of modus ponens, but yet one is not justified in accepting the conclusion" (McGee (1985), p. 462). He provides us with a counterexample to the modus ponens rule. Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:

If a Republican wins the election, then if it’s not Reagan who wins it will be Anderson.
A Republican will win the election.
Yet they did not have reason to believe If it’s not Reagan who wins, it will be Anderson. (McGee (1985), p. 462)

The first premise is in the following form: $\phi \to (\psi \to \theta)$. In the example above, ϕ stands for the second premise. By modus ponens, the truth of $(\psi \to \theta)$ must be guaranteed. However, it seems that people do not follow this line of reasoning. McGee states that we may think that the problem arises from different uses of conditional in classical logic and the English language. For this reason, he considers Robert Stalnaker’s account of conditionals. For Stalnaker, $\phi \to \psi$ means that ψ is true in the closest possible world where phi holds. Even though it satisfies modus ponens, it violates a rule ordinary language users follow in many cases, the law of exportation: $\{(\phi \wedge \psi) \to \theta\} \vdash (\phi \to (\psi \to \theta))$ According to the law of exportation, the second sentence below follows from the first sentence. However, according to Stalnaker’s account of the conditional the third sentence follows. It is an unacceptable conclusion.

1. If Reagan hadn't won the election and a Republican had won, it would have been Anderson.

2. If Reagan hadn't won the election, then if a Republican had won, it would have been Anderson.

3. If Reagan hadn't won the election, then if a Republican had won, it would have been Reagan.

To fix the problem, we may use modified versions of Stalnaker's definition of conditional above. However, as McGee states, we don't have any reason to believe that they solve the problem.

> The selective use of unnatural translations is a powerful technique for improving the fit between the logic of the natural language and the logic of a formal language. In fact, it is a little too powerful. One suspects that, if one is sly enough in giving translations, one can enable almost any logic to survive almost any counterexample. What is needed is a systematic account of how to give the translations. In the absence of such an account, the unnatural translations will seem like merely an ad hoc device for evading counterexamples. There is no guarantee that any approach will work (McGee (1985), p. 471).

I think the lesson we draw from McGee's work is that even a straightforward rule like modus ponens is not easy to apply outside of logic textbooks. As he states, "the methodological moral to be drawn from this is that, when we formulate general laws of logic, we ought to exercise the same sort of caution we exercise when we make inductive generalizations in the empirical sciences. We must take care that the instances we look at in evaluating a proposed generalization are diver as well as numerous" (McGee (1985), p. 468).

Let's consider the famous textbook example of modus ponens.

- All human beings are mortal.

- Socrates is a human being.

- Socrates is mortal.

In this derivation, we use modus ponens and the universal instantiation rule. Consider, somebody who believes that Jesus is immortal utters “Jesus is a human being” and confirms the first premise above. In that case, the person would probably say that what I mean by "all" does not quantify over Jesus. Alternatively, he may say "human" in "Jesus is a human being" and “All human beings are mortal” have different meanings. In other words, in order to apply logical rules to an ordinary discourse in a deductive, exact, non-probabilistic way, first, we need to be able to translate natural language sentences to precise logical formulas. This, I think, is a reason to use complex, dynamic language processing models instead of simpler, logic-based ones.

However, even though there are such problems, human beings use simple reasoning tools most of the time. We can repeat the question one more time. How can a complex system account for such simple daily life practices? Let’s begin with the easy part of this question. Can a complex model like a neural network implement modus ponens or any other deductive rule? The answer is yes. The origin of neural networks is to implement simple Boolean functions. It is already proven that a neural network can implement any logical function by adding more layers. Let’s consider an inference rule set theoretically. Suppose P is the power set of one’s belief set. A deductive system is a function that is defined from P to P. For example, {All human beings are mortal, Socrates is a man} is a member of P. Its value {All human beings are mortal, Socrates is a man, Socrates is mortal} is a member of P as well. The problem now is to distinguish these argument-value pairs from other pairs. Once a neural network, say a recurrent neural network, turns these sets to points in the multidimensional space, the problem shifts to classify these points. It is already proven that any region in a multidimensional space can be separated from other regions with a neural network with a sufficient number of layers (Siegelmann and Sontag (1995)) (Fodor and Pylyshyn (1995)).

Let’s now move from the mere theoretical possibility of implementing natural language inferences to a more complicated issue. As I emphasized above, even though natural language is polysemous, flexible, dynamic, etc., there are still apparently regular processes. For example, even though there are questions about the use of modus ponens rule in natural language reasoning, it is not completely eliminated. Similarly, language users make use of causality at the macro level. In their reasoning, they are able to relate an agent, say a cat, with an object. Similarly, language users make pragmatic inferences. "It's late" may imply that the speaker wants to leave. Therefore, there must be a mechanism to distinguish explicit, reasoning-based activities from implicit ones like recognizing a person in a photo. Daniel Kahneman distinguishes these two kinds of processes as System 1 and 2 abilities.

> - System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control.
> - System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration.
>
> (Kahneman (2011), p. 20)

The first three processes below belong to System 1, while the last three instantiates System 2.

- Detect that one object is more distant than another.
- Orient to the source of a sudden sound.
- Drive a car on an empty road.
- Check the validity of a complex logical argument.
- Compare two washing machines for overall value.
- Search memory to identify a surprising sound.

(Kahneman (2011), p. 21)

Kahneman says “the capabilities of System 1 include innate skills that we share with other animals. We are born prepared to perceive the world around us, recognize objects, orient attention, avoid losses, and fear spiders.” In the same naturalist spirit, I treat language understanding as a System 1 process. However, deductive rules, pragmatic inference, causal reasoning in natural language use is part of System 2.

Indeed, System 1 and System 2 are neither rival theories nor independent mechanisms. Kahneman emphasizes on the interaction between these two systems.

> System 1 and 2 are both active whenever we are awake. System 1 runs automatically and System 2 is normally in a comfortable low-effort mode, in which only a fraction of its capacity is engaged. System 1 continuously generates suggestions for System 2: impressions, intuitions, intentions, and feelings. If endorsed by System 2, impressions and intuitions turn into beliefs, and impulses turn into voluntary actions. When all goes smoothly, which is most of the time, System 2 adopts the suggestions of System 1 with little or no modification. You generally believe your impressions and act on your desires, and that is fine – usually (Kahneman (2011), p. 24).

I believe that System 2 properties emerge from a neural network through dimension reduction. Even though the semantic state we arrive at upon hearing a word is high dimensional, reducing it to a lower-dimensional space can provide us with System 2 properties. Yoshua Bengio, a prominent researcher in the connectionist paradigm, in his late works gives us a possible explanation of this emergence. Let me summarize his point.

Upon hearing a sequence of words, we enter into an unconscious state, which I call overall semantic state.

$$h1 = F(x_t, h_{t-1})$$

An unconscious state is a high-dimensional one. An attention-based function extracts consciously accessible elements from the unconscious state. It results in a low-dimensional vector that represents a conscious state. It is low dimensional because, among all the ingredients of the unconscious state, only a few of them are consciously accessible.

$$ct = C(h_t, c_{t-1}, m_{t-1}, z_t)$$

where m_t, memory at t, is $M(m_{t-1}, c_t)$

An uttered expression is even poorer than the relevant conscious state. "The conscious state is generally a richer object than the uttered sentence, i.e., mapping from conscious states to sentences loses information (think about visual imagery, or artistic expression, which are difficult to put in words), and the same sentence could thus be interpreted differently depending on context and the particulars of the agent who reads that sentence" (Bengio (2017), p. 5)

$$u_t = U(c_t, u_{t-1})$$

Here, the point is to extract high-level variables used in inferences or causal reasonings. For example, the relation between these variables can be modeled with sparse factor graphs (Bengio (2017), pp. 3-4). High-level variables are important not only in natural language reasoning but also in an agents relation to its environment.

> ... for AI agents such as robots trying to make sense of their environment, the only observables are low-level variables like pixels in images or low-level motor actions. To generalize well, an agent must induce high-level variables, particularly those which are causal (i.e., can play the role of cause or effect)
>
> ...in realistic settings such as those experienced by a child or a robot, the agent typically does not have full knowledge of what abstract action was performed and needs to perform inference over that too: in addition to discovering the causal variables (from the low-level observations) and how they are related to each other, the agent needs to learn how high-level intentions and actions are linked with low-level observations and with interventions on the high-level causal variables (Goyal and Bengio (2020), pp. 21-22)

This approach is not a hybrid one. In other words, traditional tools such as symbolic logic, which are good at System 2 processes, are not put on top of connectionist systems (Bengio (2017), p. 6). At this point, I share Yoshua Bengio, Yann Lecun, and Geoffrey Hinton's comment below:

> Clearly, this symbolic AI program aimed at achieving system 2 abilities, such as reasoning, being able to factorize knowledge into pieces which can easily recombined in a sequence of computational steps, and being able to manipulate abstract variables, types and instances. We would like to design neural networks which can do all these things while working with real-valued vectors so as to preserve the strengths of deep learning which include efficient large-scale learning using differentiable computation and gradient-based adaptation, grounding of high-level concepts in low-level perception and action, handling uncertain data, and using distributed representations (Bengio et al. (2021), p. 65)

As we discussed in the previous chapters, even though the symbolic approach is good at modeling compositionality, systematicity and productivity, it falls short of representing the uncertain, flexible, dynamic, holistic parts of language understanding. (Bengio (2017), p.6) (Russin et al. (2020)) (Goyal and Bengio (2020), pp. 33-34)

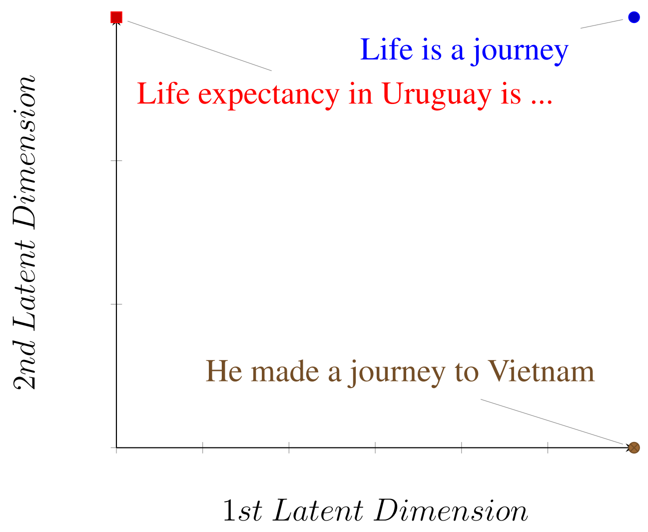
To summarize, first of all, system 2 abilities in language understanding, such as causal reasoning, inference, compositionality, etc., are not easily processed as they seem. However, the difficulty does not abolish the distinction between system 1 and 2 processes. The approach I defend is mainly interested in the system 1 abilities. Inference is a system 2 process. As discussed above, the emergence of system 2 abilities from the system 1 abilities is possible. It is an active research area in connectionism.

6.4 Literal - Metaphorical distinction

According to the compositional approach, the meaning of a sentence is determined by the meanings of its constituents. In computing the meaning of "Necessity is the mother of invention" and "Jane is her mother," two different concepts for mother are used. One is the metaphorical $MOTHER$, and the other is the literal $MOTHER$. On the other hand, in my approach, there is no such distinction since there is no meaning or meanings assigned to the word "mother."

Can we say that there is no literal metaphorical distinction in the approach I defend? The answer is both yes and no. In terms of processing, there is no literal metaphorical distinction since no meaning is assigned to the word "mother." However, there is a psycholinguistic distinction that English language users can make. There are psychiatric conditions regarding our capability of making this distinction (**?**). Therefore, even though the difference is evident neither in inputs nor in computational steps, there must be a categorical distinction between outputs. Therefore, the metaphorical literal distinction must be represented at the output level.

Let me show one way to show distinction at the output level. Suppose the outputs of two expressions, S_1 and S_2, are o_1 and o_2, respectively. English language users consider S_1 as metaphorical while S_2 as literal. In my proposal, o_1 must be richer than o_2 in terms of their conceptual analysis. In a constructed space with dimensions consisting of various topics (e.g., sexuality, religion, politics, knowledge, etc.), there are more dimensions where o_2 has significant values. Let me illustrate the distinction with the diagram below.



For the sake of brevity, the illustration above is a simplified one. Uses of expressions determine their location in the graph above. Utterances of the same expression type may be located at different places, depending on their use. For example, a metaphor's position shifts when it is frequently used. In other words, live and dead metaphors are located at different positions. Similarly, the context may shift a prima facie literal sentence's position.

Science/art distinction can be considered an example to understand the graph above. In scientific practice, we usually try to move in single dimensions as much as possible. This minimalist strategy aims to control what we talk about. We want to be safe.

On the other hand, in art, we tend to go on in more dimensions simultaneously. For this reason, it is not easy to paraphrase an artwork with words. To note, I do not presuppose a sharp line between science and art in terms of metaphor/literal distinction. As I argue in the previous chapters, scientific expressions are not rule-governed as well. However, the prototypical understanding of science and art aligns with this distinction. Prototypically, a poem, a painting, or a movie is rich in terms of dimensions. On the other hand, a scientific paper tends to be as minimal as possible.

Similarly, in pragmatic disorders, individuals do not have a tendency to move in multiple dimensions simultaneously. Alternatively, the situation can be described as neurotypical individuals cannot prevent themselves from moving in multiple dimensions given certain inputs. Here, my aim is not to provide a model for psychiatric conditions. Similarly, I do not develop a theory of science art distinction. Instead, these issues are used to illustrate how literal/metaphorical distinction is represented in the approach I propose.

6.5 Frege's puzzles from the eliminativist perspective

In any philosophy of language textbook, Frege's puzzles take a central place. Major theories like Millianism, Descriptivism, etc., are centered around these puzzles. Let me repeat the puzzles.

The Identity Puzzle:

Consider the identity sentence below:

- Hesperus is Hesperus.
- Hesperus is Phosphorus.

The first sentence is in $\ulcorner a\ is\ a \urcorner$ form, while the second is in $\ulcorner a\ is\ b \urcorner$ form. According to the compositional approach, the meanings of these sentences depend on the meaning of their constituent parts. In other words, their meanings depend on the meanings of "Hesperus" and "Phosphorus." According to one approach, the meaning is the object referred because when you say "Hesperus is bright," what is bright is the object "Hesperus" refers to. Then, the sentences above have the same meaning. Since "Hesperus" and "Phosphorus" refer to the same object, both sentence says nothing other than the self-identity of that object. However, while the second sentence is informative, the first one is not.

The Propositional Attitude Puzzle:

Consider the inference below.

John believes that Hesperus star is bright.
Hesperus is Phosphorus.

$\therefore$ John believes that Phosphorus is bright.

By the same line of reasoning above, the truth of the first two sentences must guarantee the truth of the last sentences. However, it may be the case that the first two sentences are true while the last one is false.

There are several proposals in the contemporary philosophy of language literature to overcome these problems. These solutions bring new notions of meaning, identity, psychological predicates, etc.

How does the position I defend evaluates these puzzles? First of all, these puzzles belong to the compositional, symbolic approach, according to which these sentences and sentence parts correspond to meanings. Without meanings, we do not have these puzzles. Let me briefly illustrate how a simple eliminativist model processes the sentences above.

Let' f be the non-linearity function that we use in this model, and W is the set of weights. Given the sentences, “Hesperus is Hesperus" and “Hesperus is Phosphorus," the system outputs the following values

$$f(W*vector_{Hesperus}+W*f(W*vector_{is}+W*f(W*vector_{Hesperus}+W*context)))$$

$$f(W*vector_{Hesperus}+W*f(W*vector_{is}+W*f(W*vector_{Phosphorus}+W*context)))$$

Since these may correspond to different values in a high dimensional space, we don’t expect them to have the same properties (e.g., informativity).

Similarly, the propositional attitude puzzle does not arise in the eliminativist system since there is no compositionality-based inference mechanism. Instead, as I explained earlier in this chapter, inferences are governed by a few variables at the surface level. Those variables are not word meanings.

CHAPTER 7

CONCLUSION

The aim of both my approach and compositional approaches is to find a mechanical procedure to imitate language understanding. The defenders of the compositional approach believe that there is a homomorphism from the surface structure of a linguistic expression to its meaning. If there is such a structural similarity that maps surface syntax to semantics, we can use surface structure to understand meaning. There are two main schools of thought that follows the compositionality principle. These are the rule-based approach and the pragmatics-oriented approach.

For the rule-based version, we need to map constituents of the surface structure to constituents of their meanings in a non-trivial, systematic way. In short, the interpretation of $F_{ß}(\alpha_1....\alpha_n)$ is computed below:

$$h(F_{ß}(\alpha_1....\alpha_n)) = G_{ß}(h(\alpha_1)....h(\alpha_n)).$$

where F is a syntactic combination function, h is an interpretation function and $G_{ß}$ is a semantic combination function.

In the third chapter, I argue that this is far from possible. Polysemy shows the structural discrepancy between the surface syntax and meaning. Once we ignore this distinction, we confront profound conceptual issues.

The problems with the rule-based approach is already stated by the other compositional approach, pragmatics-oriented approach. For the pragmatics-oriented version, context representations and meanings must work together. Meanings of expressions are built upon, word meanings and contextual ingredients. Inferential procesees, associations play an important role. Computation of the meaning of $F_{ß}(\alpha_1....\alpha_n)$ is simplified below:

$$h(F_{ß}(\alpha_1....\alpha_n)) = G_{ß}(p_i(h(\alpha_1))....p_j(h(\alpha_n))).$$

where p_i and p_j are pragmatic functions.

In the fourth chapter, I argue that using context representations, meanings in an inferential or associationist mechanism is not realistic. Both of these approaches are interested in building a semantic state through smaller semantic entities.

In this work, I defend Meaning Eliminativism according to which there words do not have any meaning which implies that the sentence meaning is not built upon the word meaning. To make it a tenable position, I propose to use connectionist architectures instead of classical, symbolic ones. Therefore, to understand the meaning of a complex expression, instead of using expressions' surface structure, we should look at the rich structure of the semantic space that language users (e.g., a silicon computer, a human being, etc.) enter upon receiving linguistic inputs. In other words, linguistic expressions' role is to cause language users to enter into a new semantic state.

The difficulties we are having with compositional approaches is that they aim to build a semantic state with a set of ingredients. Semantic combination rules guide where to put these ingredients. Pragmatic functions modulate the ingredients. It would be a very, simple, elegant approach to language undertanding only if it was practical. However, the way language users understand linguistic expressions is too complicated to be modeled with these tools.

On the other hand, I do not propose a way to build a semantic state. Instead, words move language users from one state to another. Sub-symbolic tools give us insight to go further in this framework. According to this alternative approach, the sub-symbolic paradigm, high dimensional space is the only tool to approximate the rich structure of language understanding. That is the reason to assign real-valued vectors to words. To conclude, words do not have any meaning. Rather, they cause us, language users, to move in the semantic space where we can find different shades of meaning.

In the last chapter, we deal with some further issues. The first one is the normativity in language use. Normativity is the key issue not only in ethics but also in the philosophy of language. Starting with Kripke's book Wittgensteins's private language argument, normativity has become a major issue in meaning studies. In this section, I give a connecitonist account of normativity on the basis of social adaptation.

The second issue is the problem of explanation. The concept of explanation is an essential feature of scientific theories. A connectionist language understanding model with no meanings, context representations, semantic combination rules etc. seems raise doubts about its explanatory power. Through referring to Leo Breiman's two different scientific cultures, I give an answer to this question.

The fourth issue is how to model simple rule-guided behaviors like inferences. Deductive, abductive, inductive, pragmatic, causal inference types are frequent in everyday language use. Anyone who understands, “It rains" may infer that the ground will be wet. In such a reasoning, we see relations between $rain$, $wetness$, $ground$, etc. On the other hand, the model I propose is not based on such macro entities. Instead, millions of micro entities do the job. Through pointing Yoshua Bengio's late works on the subject, I claim that these apparently rule-based behavior emerge from a connectionist language understanding mechanism.

The fifth issue is the distinction between the literal and metaphorical uses of expressions. Is there such distinction in the model I propose? There is a distinction, as the distinction is psychologically real. However, the distinction is not seen at the processing level. The distinction is obvious at the output level. Metaphorical uses output are richer than the literal uses in terms their significant dimensions.

REFERENCES

Ainsworth, T. (2020). Form vs. Matter. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2020 edition.

Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, 12(142):5–32.

Arora, S., Li, Y., Liang, Y., Ma, T., and Risteski, A. (2018). Linear algebraic structure of word senses, with applications to polysemy. *Transactions of the Association for Computational Linguistics*, 6:483–495.

Austin, J. L. (1961). The meaning of a word. In J. O. Urmson, G. J. W., editor, *Philosophical Papers*, pages 23–43. Oxford University Press.

Baker, G. P. and Hacker, P. M. S. (1984). On misunderstanding wittgenstein: Kripke's private language argument. *Synthese*, 58(3):407–450.

Bar-Hillel, Y. (1964). A demonstration of the nonfeasibility of fully automatic high quality machine translation. In *Language and Information: Selected Essays on Their Theory and Application.*, pages 174–179. Addison-Wesley, Reading, Mass.

Barbara H. Partee, Alice Ter Meulen, R. E. W. (1990). *Mathematical methods in linguistics*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Barker, C. and Jacobson, P. (2007). Introduction: Direct compositionality. In Baker, C. and Jacobson, P., editors, *Direct Compositionality*. Oxford University Press, Oxford.

Baroni, M., Bernardi, R., Zamparelli, R., et al. (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in language technology*, 9(6):5–110.

Bengio, Y. (2017). The consciousness prior. *CoRR*, abs/1709.08568.

Bengio, Y., Lecun, Y., and Hinton, G. (2021). Deep learning for ai. *Communications of the ACM*, 64(7):58–65.

Berkeley, G. (2009). *Philosophical Writings*. Cambridge University Press.

Boghossian, P. (2005). Is meaning normative? In Christian, N. and Ansgar, B., editors, *Philosophy – Science – Scientific Philosophy. Main Lectures and Colloquia of GAP.5, Fifth International Congress of the Society for Analytical Philosophy, Bielefeld, 22–26 September 2003*, pages 205–218. Paperborn.

Borg, E. (2004). *Minimal semantics*. Oxford University Press, Oxford.

Boyle, R. (2009). Of the excellency and grounds of the corpuscular or mechanical philosophy. In Ariew, R. and Watkins, E., editors, *Modern Philosophy: An Anthology of Primary Sources*. Hackett Publishing Company, Indianapolis, Indiana.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3):199–231.

Brugman, C. and Lakoff, G. (1988). Cognitive topology and lexical networks. In Steven L. Small, G. W. C. and Tanenhaus, M. K., editors, *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence*, pages 477–508. Morgan Kaufmann, San Mateo, CA.

Cappelen, H. and Lepore, E. (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Wiley-Blackwell.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue internationale de philosophie*, 4(11):20–40.

Carston, R. (2002). *Thoughts and Utterances The Pragmatics of Explicit Communication*. Blackwell Publishing, Malden, MA.

Carston, R. (2007). How many pragmatic systems are there? In Frápolli, M. J., editor, *Saying, Meaning and Referring: Essays on François Recanati's Philosophy of Language*. Palgrave-Macmillan.

Carston, R. (2012). Word meaning and concept expressed. *The Linguistic Review*, 29:607– 23.

Carston, R. (2015). Contextual adjustment of meaning. In *The Routledge Handbook of Semantics*. Routledge, London.

Carston, R. (2019). Ad hoc concepts, polysemy and the lexicon. In Scott, K., Clark, B., and Carston, R., editors, *Relevance, Pragmatics and Interpretation*, pages 150–162. Cambridge University Press.

Carston, R. (2021). Polysemy: Pragmatics and sense conventions. *Mind & Language*, 36(1):108–133.

Chalmers, D. J. (1993). Connectionism and compositionality: Why fodor and pylyshyn were wrong. *Philosophical Psychology*, 6(3):305–319.

Chomsky, N. (2002). *On nature and language*. Cambridge University Press, New York.

Chomsky, N. (2011). Language and other cognitive systems. what is special about language? *Language Learning and Development*, 7(4):263–278.

Churchland, P. (2007). *Neurophilosophy at Work*. Cambridge University Press, New York.

Churchland, P. M. (1995). *The Engine of Reason, the Seat of the Soul: A Philosophical Journey Into the Brain*. MIT Press.

Clark, H. (1996). *Using Language*. Cambridge University Press, Cambridge.

Clark, H. H. (1983). Making sense of nonce sense. In D'Arcais, G. B. F. and Jarvella, R. J., editors, *The Process of Language Understanding*, pages 297–331. John Wiley & Sons Ltd.

Clark, H. H. and Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning and Verbal Behavior*, 22(5):591–608.

Cohen, J. (1985). A problem about ambiguity in truth-conditional semantics. *Analysis*, 45:129–135.

Davidson, D. (1967). Truth and meaning. *Synthese*, 17(1):304–323.

Davidson, D. (2001a). In defence of convention t (1973). In Davidson, D., editor, *Inquiries into Truth and Interpretation: Philosophical essays*, pages 65–74. Oxford University Press, Oxford.

Davidson, D. (2001b). Semantics for natural languages (1970). In Davidson, D., editor, *Inquiries into Truth and Interpretation: Philosophical essays*, pages 55–64. Oxford University Press, Oxford.

Davidson, D. (2001c). Theories of meaning and learnable languages (1965). In Davidson, D., editor, *Inquiries into Truth and Interpretation: Philosophical essays*, pages 3–16. Oxford University Press, Oxford.

Davidson, D. (2001d). What metaphors mean. In Davidson, D., editor, *Inquiries into truth and interpretation: Philosophical essays*, pages 245–264. Oxford University Press, Oxford.

Davidson, D. (2005a). A nice derangement of epitaphs. In *Truth, Language and History*, chapter 7, pages 89–107. Oxford University Press, New York.

Davidson, D. (2005b). *Truth and Predication*. Harvard University Press, Cambridge, MA.

Davies, M. (2003). Philosophy of language. In Bunnin, N. and Tsui-James, E. P., editors, *The Blackwell Companion to Philosophy*, pages 90–146. Blackwell Publishing, Malden, MA.

de Sola Pool, I. and Kochen, M. (1978). Contacts and influence. *Social networks*, 1(1):5–51.

Dennett, D. C. (1997). True believers: the intentional strategy and why it works. in j. haugeland, editor, mind design ii: Philosophy psychology artificial intelligence. In Haugeland, J., editor, *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. MIT Press, Cambridge, Massachusetts.

Descartes, R. (1968). *Discourse on Methods and the Meditations (translated by F. E. Sutcliffe)*. Penguin Books, London.

Dever, J. (1999). Compositionality as methodology. *Linguistics and Philosophy*, 22(3):311–326.

Devitt, M. (2013). Is there a place for truth-conditional pragmatics? *Teorema: Revista Internacional de Filosofía*, pages 85–102.

Dowty, D. (2007). Compositionality as an empirical problem. In Baker, C. and Jacobson, P. I., editors, *Direct Compositionality*, pages 23–101. Oxford University Press, New York.

Dowty, D. R. (1991). *Word Meaning and Montague Grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*, volume 7 of *Synthese Language Library Text and Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, Dodrecht, Holland.

Dumais, S. T. (2004). Latent semantic analysis. *Annual review of information science and technology*, 38(1):188–230.

Elgin, C. Z. (1996). *Considered judgment*. Princeton University Press, Princeton, New Jersey.

Eliasmith, C. and Bechtel, W. (2006). Symbolic versus subsymbolic. *Encyclopedia of cognitive science*.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2):179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2):195–225.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4):547–582.

Ettinger, A., Elgohary, A., Phillips, C., and Resnik, P. (2018). Assessing composition in sentence vector representations. In *The 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico.

Falkum, I. L. (2011). *The Semantics and Pragmatics of Polysemy: A Relevance-Theoretic Account*. PhD thesis, University College London.

Falkum, I. L. (2015). The how and why of polysemy: A pragmatic account. *Lingua*, 157:83–99.

Falkum, I. L. and Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*, 157:1–16.

Fellbaum, C. (2013). Wordnet. In Chapelle, C. A., editor, *The Encyclopedia of Applied Linguistics*, pages 707–728. Wiley/Blackwell.

Ferrer i Cancho, R. and Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791.

Fillmore, C. J. (2006). Frame semantics. In Geeraerts, D., editor, *Cognitive Linguistics: Basic Readings*, volume 34, pages 373–400. Mouton de Gruyter, Berlin.

Fine, G. (1992). Inquiry in the meno. In Kraut, R., editor, *Cambridge Companion to Plato*. Cambridge University Press, New York.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. The Mit Press, Cambridge.

Fodor, J. (2008). *LOT 2 The Language of Thought Revisited*. Clarendon Press, Oxford.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press, New York.

Fodor, J. A. and Pylyshyn, Z. W. (1995). Connectionism and cognitive architecture: A critical analysis. In Cynthia MacDonald, G. M., editor, *Connectionism: Debates on Psychological Explanation*. Basil Blackwell.

Frege, G. (1879/1967). Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In *From Frege to Gödel: A source book in mathematical logic*, pages 1–82. Harvard University Press, Cambridge, Massachusetts.

Frege, G. (1948). Sense and reference. *The philosophical review*, 57(3):209–230.

Frege, G. (1953). *Foundations of Arithmetic (Transl. by J. L. Austin)*. Basil Blackwell, Oxford.

Frege, G. (1963). Compound thoughts. *Mind*, 72(285):1–17.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, 3(1):111–127.

Frisson, S. (2015). About bound and scary books: The processing of book polysemies. *Lingua*, 157:17–35.

Gamut, L. T. F. (1990). *Logic, Language, and Meaning, Volume 2: Intensional Logic and Logical Grammar*. University of Chicago Press.

Gärdenfors, P. (1999). Some tenets of cognitive semantics. *Cognitive semantics: Meaning and cognition*, 55:19–36.

Gasparri, L. (2013). A look at meaning eliminativism. *Philosophy Study*, 3(11):1018.

Gasparri, L. (2014). Minimal semantics and word sense disambiguation. *Disputatio*, 6(39):147–171.

Geeraerts, D. (2010). *Theories of Lexical Semantics*. Oxford University Press.

Geeraerts, D. (2013). Post-structuralist and cognitive approaches to meaning. In Allan, K., editor, *The Oxford Handbook of the History of Linguistics*. Oxford University Press, Oxford.

Geeraerts, D. and Cuyckens, H. (2007). *The Oxford handbook of cognitive linguistics*. Oxford University Press.

Glüer, K. and Wikforss, Å. (2015). Meaning normativism: Against the simple argument. *Organon F*, 22(1):63–73.

Goldberg, A. E. (2016). Compositionality. In *The Routledge handbook of semantics*, pages 419–433. Routledge.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Goodman, N. (1968). *Languages of art: An approach to a theory of symbols*. Hackett publishing.

Goyal, A. and Bengio, Y. (2020). Inductive biases for deep learning of higher-level cognition. *CoRR*, abs/2011.15091.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66(3):377–388.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Grice, P. (1991a). Further notes on logic and coversation. In *Studies in the Way of Words*, pages 41–57. Harvard University Press.

Grice, P. (1991b). Logic and coversation. In *Studies in the Way of Words*, pages 22–40. Harvard University Press.

Harris, C. L. (1994). Coarse coding and the lexicon. In *Continuity in Linguistic Semantics*, pages 205–229. John Benjamins Publishing Company, Amsterdam.

Hattiangadi, A. (2006). Is meaning normative? *Mind and Language*, 21(2):220–240.

Heck, R. (2013). Is compositionality a trivial principle? *Frontiers of Philosophy in China*, 8(1):140–55.

Hintikka, J. (1996). *Ludwig Wittgenstein: Half-Truths and One-and-a-Half-Truths*. Kluwer Academic Publishers.

Hintikka, M. and Hintikka, J. (1986). *Investigating Wittgenstein*. Basil Blackwell, Oxford.

Hirsch, E. (2002). Quantifier variance and realism. *Philosophical Issues*, 12:51–73.

Hobbes, T. (1839/1655). Elements of philosophy, part i-iv. In *The English Works of Thomas Hobbes*, volume I. John Bohn, London.

Hodges, W. (2001). Formal features of compositionality. *Journal of Logic, Language and Information*, 10(1):7–28.

Hodges, W. (2012). Formalizing the relationship between meaning and syntax. In *Oxford Handbook of Compositionality*, pages 245–261. Oxford University Press.

Hofmann, T. (2013). Probabilistic latent semantic analysis. *arXiv preprint arXiv:1301.6705*.

Horwich, P. (1997). The composition of meanings. *The Philosophical Review*, 106(4):503–532.

Hugh D. Young, R. A. F. (2007). *University Physics*, volume 1. Pearson, San Francisco.

Hylton, P. (2019). Carnap and quine on analyticity: The nature of the disagreement. *Noûs*.

Iacobacci, I., Pilehvar, M. T., and Navigli, R. (2015). SensEmbed: Learning sense embeddings for word and relational similarity. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 95–105, Beijing, China. Association for Computational Linguistics.

Jackendoff, R. (1995). *Languages of the Mind Essays on Mental Representation*. The Mit Press, Cambridge.

Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press, Oxford.

Janssen, T. M. V. (2001). Frege, contextuality and compositionality. *Journal of Logic, Language and Information*, 10(1):115–136.

Johnson, M. (2013). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.

Kant, I. (1787/1998). *Critique of Pure Reason, translated/edited by P. Guyer and A. Wood*. Cambridge University Press, Cambridge.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes From Kaplan*, pages 481–563. Oxford University Press.

Katz, J. and Fodor, J. (1963). The structure of a semantic theory. *Language*, 39:170–210.

Kazmi, A. and Pelletier, F. J. (1998). Is compositionality formally vacuous? *Linguistics and Philosophy*, 21(6):629–633.

King, P. (2007). Abelard on mental language. *American Catholic Philosophical Quarterly*, 81(2):169–187.

King, P. and Arlig, A. (2018). Peter Abelard. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2018 edition.

Kleene, S. C. (1956). Kleene. representation of events in nerve nets and finite automata. *Automata studies*.

Kossak, R. (2018). *Mathematical Logic*. Springer.

Kripke, S. A. (1982). *Wittgenstein on rules and private language: An elementary exposition*. Harvard University Press.

Kuhn, T. S. (1979). Metaphor in science. In Ortony, A., editor, *Metaphor and Thought*, pages 409–19. Cambridge University Press.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago, third edition.

Lakoff, G. (1987). *Women, Fire and Dangerous Thing: What Catergories Reveal About the Mind*. University of Chicago Press.

Lakoff, G. (1988). Cognitive semantics. In Eco, U., editor, *Meaning and Mental Representations*, pages 119–154. Bloomington: Indiana University Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In Ortony, A., editor, *Metaphor and Thought*, pages 202–251. Cambridge University Press.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh*, volume 4. New york: Basic books.

Lakoff, G. and Johnson, M. (2008). *Metaphors We Live By*. University of Chicago press.

Langacker, R. (1984). Active zones. In *Annual Meeting of the Berkeley Linguistics Society*, volume 10, pages 172–188.

Lewis, D. (1975). Languages and language. In Gunderson, K., editor, *Minnesota Studies in the Philosophy of Science*, volume VII, pages 3–35. University of Minnesota Press, Minneapolis.

Lewis, D. K. (1969/2002). *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell.

Lewis, M. (2019). Compositionality for recursive neural networks. *arXiv preprint arXiv:1901.10723*.

Linnebo, Ø. (2018). *Thin Objects: An Abstractionist Account*. Oxford University Press.

Locke, J. (1975). *An Essay Concerning Human Understanding (Peter H. Nidditch ed.)*. Oxford University Press, Oxford.

Ludlow, P. (1989). Implicit comparison classes. *Linguistics and Philosophy*, 12(4):519–533.

Manning, C. D., Manning, C. D., and Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

McClelland, J. L., St. John, M., and Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, 4(3-4):287–335.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

McGee, V. (1985). A counterexample to modus ponens. *The Journal of Philosophy*, 82(9):462–471.

McKeon, R. (2001). *The Basic Works of Aristotle*. The Modern Library, New York.

Mervis, C. B. and Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, 32(1):89–115.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Milgram, S. (1967). The small world problem. *Psychology today*, 2(1):60–67.

Mill, J. S. (1882). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Princilples of Evidence and the Methods of Scientific Investigation*. Harper and brothers.

Miller, G. A. (1951). *Language and communication*. McGraw-Hill.

Miller, G. A. (1999). On knowing a word. *Annual Review of Psychology*, 50(1):1–19.

Minsky, M. and Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, USA.

Montague, R. (1974a). English as a formal language. In *Formal Philosophy: Selected Papers of Richard Montague. Ed. and with an Introd. by Richmond H. Thomason*. Yale University Press.

Montague, R. (1974b). Universal grammar. In *Formal Philosophy: Selected Papers of Richard Montague. Ed. and with an Introd. by Richmond H. Thomason*. Yale University Press.

Neale, S. (1990). *Descriptions*. MIT Press.

Nietzsche, F. W. (2009). On truth and lie in an extra-moral sense. In *Writings From the Early Notebooks*. Cambridge University Press.

Norvig, P. (2012). Colorless green ideas learn furiously: Chomsky and the two cultures of statistical learning. *Significance*, 9(4):30–33.

Norvig, P. and Lakoff, G. (1987). Taking: A study in lexical network theory. In *Annual Meeting of the Berkeley Linguistics Society*, volume 13, pages 195–206.

Pagin, P. (2003). Communication and strong compositionality. *Journal of Philosophical Logic*, 32(3):287–322.

Pagin, P. and Westerståhl, D. (2010a). Compositionality i: Definitions and variants. *Philosophy Compass*, 5(3):250–264.

Pagin, P. and Westerståhl, D. (2010b). Compositionality ii: Arguments and problems. *Philosophy Compass*, 5(3):265–282.

Partee, B. (1984). Compositionality. varieties of formal semantics. In Veltman, F., editor, *Proceedings of the fourth Amsterdam colloquium*, pages 281–311. Dordrecht: Foris Publications.

Pelczar, M. (2000). Wittgensteinian semantics. *Noûs*, 34(4):483–516.

Pelletier, F. J. (2001). Did frege believe frege's principle? *Journal of Logic, Language and Information*, 10(1):87–114.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Peters, M. E., Ammar, W., Bhagavatula, C., and Power, R. (2017). Semi-supervised sequence tagging with bidirectional language models. *arXiv preprint arXiv:1705.00108*.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.

Piantadosi, S. T., Tily, H., and Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3):280–291.

Piccinini, G. (2004). The first computational theory of mind and brain: A close look at mcculloch and pitts's "logical calculus of ideas immanent in nervous activity". *Synthese*, 141(2):175–215.

Pustejovsky, J. (1996). *The generative lexicon*. MIT Press.

Pustejovsky, J. (1998). Generativity and explanation in semantics: A reply to fodor and lepore. *Linguistic Inquiry*, 29(2):289–311.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological review*, 88(1):16.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1):20–43.

Quine, W. V. O. (2013). *Word and object*. MIT press.

Quine, W. V. O. and Ullian, J. S. (1970). *The Web of Belief*. New York: Random House.

Rademaker, A. (2005). *Sophrosyne and the Rhetoric of Self-Restraint: Polysemy & Persuasive Use of an Ancient Greek Value Term*. Brill, Leiden.

Rakova, M. (2003). *The Extent of the Literal: Metaphor, Polysemy and the Theories of Concepts*. Palgrave-Macmillan.

Ravin, Y. and Leacock, C. (2000). *Polysemy: Theoretical and computational approaches*. OUP Oxford.

Rayo, A. (2013). A plea for semantic localism. *Noûs*, 47(4):647–679.

Recanati, F. (2004). *Literal meaning*. Cambridge University Press.

Recanati, F. (2010). *Truth-Conditional Pragmatics*. Oxford University Press.

Recanati, F. (2017). Contextualism and polysemy. *Dialectica*, 71(3):379–397.

Recanati, F. (2019). Why polysemy supports radical contextualism. In *International and Interdisciplinary Conference on Modeling and Using Context*, pages 216–222. Springer.

Rosch, E. (1978). Principles of categorization. In Rosch, E. and Lloyd, B., editors, *Cognition and Categorization*. Lawrence Elbaum Associates.

Rosch, E. (2013). Neither concepts nor lotfi zadeh are fuzzy sets. In *On Fuzziness*, pages 591–596. Springer.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976).

Basic objects in natural categories. *Cognitive psychology*, 8(3):382–439.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Rumelhart, D. E. (1975). Notes on a schema for stories. In *Representation and understanding*, pages 211–236. Elsevier.

Rumelhart, D. E. (1979). Some problems with the notion of literal meanings. *Metaphor and thought*, 2:71–82.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986a). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. (1986b). Schemata and sequential thought processes in pdp models. In James L. McClelland, D. E. R. and Group, T. P. R., editors, *Parallel Distributed Processing Explorations in the Microstructures of Cognition*, volume 2 Psychological and Biological Models, pages 7–57. The MIT Press, Cambridge, Massachusetts.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Russell, B. (1910). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society*, 11:108–128.

Russell, B. (2010). *The Philosophy of Logical Atomism*. Routledge, London and New York.

Russell, S. and Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson.

Russin, J., O'Reilly, R. C., and Bengio, Y. (2020). Deep learning needs a prefrontal cortex. *Work Bridging AI Cogn Sci*, 107:603–616.

Ryle, G. (2009). Philosophical arguments. In *Collected Essays 1929 - 1968: Collected*

Papers Volume 2. Routledge.

Sadrzadeh, M. and Muskens, R. (2019). Static and dynamic vector semantics for lambda calculus models of natural language. *Journal of Language Modelling*, 6(2):319–351.

Saeed, J. I. (2009). *Semantics*. Wiley-Blackwell, Oxford, third edition.

Salmon, N. (2005). Two conceptions of semantics. In Szabó, Z. G., editor, *Semantics Versus Pragmatics*, pages 317–328. Oxford University Press.

Schank, R. C. and Abelson, R. P. (2013). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.

Schneider, W. (1987). Connectionism: Is it a paradigm shift for psychology? *Behavior Research Methods, Instruments, & Computers*, 19(2):73–83.

Searle, J. R. (1978). Literal meaning. *Erkenntnis*, pages 207–224.

Searle, J. R. (1993). Metaphor. In Ortony, A., editor, *Metaphor and thought*, pages 83–111. Cambridge University Press.

Searle, J. R., Willis, S., et al. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge university press.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Siegelmann, H. T. and Sontag, E. D. (1995). On the computational power of neural nets. *Journal of computer and system sciences*, 50(1):132–150.

Smolensky, P. (1995a). Connectionism, constituency and the language of thought. In Cynthia MacDonald, G. M., editor, *Connectionism: Debates on Psychological Explanation*. Basil Blackwell.

Smolensky, P. (1995b). On the proper treatment of connectionism. In Cynthia MacDonald, G. M., editor, *Connectionism: Debates on Psychological Explanation*. Basil Blackwell.

Solé, R. V., Corominas-Murtra, B., Valverde, S., and Steels, L. (2010). Language networks: Their structure, function, and evolution. *Complexity*, 15(6):20–26.

Solé, R. V. and Seoane, L. F. (2015). Ambiguity in language networks. *The Linguistic Review*, 32(1):5–35.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication & Cognition*. Blackwell, Oxford.

Srinivasan, M. and Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, 157:124–152.

Stalnaker, R. (1979). Anti-essentialism. *Midwest Studies in Philosophy*, 4(1):343–355.

Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, 25(5-6):701–721.

Stanley, J. (2000). Context and logical form. *Linguistics and Philosophy*, 23(4):391–434.

Strawson, P. F. (1952). *Introduction to Logical Theory*. Methuen, London.

Sudkamp, T. A. (1997). *Languages and Machines: An Introduction to the Theory of Computer Science*. Addison-Wesley, Reading, Mass., 2 edition.

Szabó, Z. G. (2000). Compositionality as supervenience. *Linguistics and Philosophy*, 23(5):475–505.

Szabó, Z. G. (2001). Adjectives in context. In Jucker, A. H., editor, *Perspectives on Semantics, Pragmatics, and Discourse A Festschrift for Ferenc Kiefer*, pages 119–

146. John Benjamins Publishing.

Szabó, Z. G. (2006). Sensitivity training. *Mind and Language*, 21(1):31–38.

Szabó, Z. G. (2012). The case for compositionality. In *Oxford Handbook of Compositionality*, pages 175–191. Oxford University Press.

Szabó, Z. G. (2017). Compositionality. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition.

Taylor, J. R. (2003). Polysemy's paradoxes. *Language Sciences*, 25(6):637 – 655. Twenty-fifth Anniversary Issue.

Touretzky, D. S. (1994). Polysemy and representation. In *Continuity in Linguistic Semantics*, pages 231–239. John Benjamins.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(October):433–60.

van Fraassen, B. C. (1967). Meaning relations among predicates. *Noûs*, 1(2):161–179.

Vicente, A. (2018). Polysemy and word meaning: an account of lexical meaning for different kinds of content words. *Philosophical Studies*, 175(4):947–968.

Victorri, B. (1994). The use of continuity in modeling semantic phenomena. In Catherine Fuchs, B. V., editor, *Continuity in Linguistic Semantics*, pages 241–251. Benjamins.

Wasow, T. (2015). Ambiguity avoidance is overrated[1]. *Ambiguity: Language and communication*, pages 29–47.

Wasow, T., Perfors, A., and Beaver, D. (2005). The puzzle of ambiguity. *Morphology and the web of grammar: Essays in memory of Steven G. Lapointe*, pages 265–282.

Weaver, W. (1953). Recent contributions to the mathematical theory of communication. *ETC: A Review of General Semantics*, pages 261–281.

Weinreich, U. and Webster (1964). Webster's third: A critique of its semantics. *International Journal of American Linguistics*, 30(4):405–409.

Wierzbicka, A. (1996). *Semantic Primes and Universals*. Oxford University Press, Oxford.

Wilson, D. and Sperber, D. (2002). Truthfulness and relevance. *Mind*, 111(443):583–632.

Wilson, D. and Sperber, D. (2006). Relevance theory. In Ward, L. R. H. . G., editor, *The Handbook of Pragmatics*, pages 607–632. Blackwell Publishing, Malden, MA.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus Translated by C.K. Ogden*. Routledge & Kegan Paul Ltd.

Wittgenstein, L. (1953). *Philosophical investigations (translated by G.E.M. Anscombe, P.M.S. Hacker, and Joachim Schulte)*. Wiley-Blackwell, 4 edition.

Wittgenstein, L. (1965). *The Blue and Brown Books*. Harper and Row, New York.

Wittgenstein, L. (1977). *Culture and Value translated by Peter Winch*. University of Chicago Press.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zadrozny, W. (1992). On compositional semantics. In *COLING 1992 Volume 1: The 15th International Conference on Computational Linguistics*, pages 260–266.

Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy*, 17(4):329–342.

Zimmermann, F. W. (1981). *Al-Farabi's Commentary and Short Treatise on Aristotle's De Interpretatione*, volume 3. Oxford University Press, London.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Hafner Publishing Company, New York.

APPENDICES

A. CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Toy, Tolgahan

Nationality: Turkish

Date and Place of Birth: 1986 Sivas

Marital Status: Single

Phone:

E-mail: tolgahantoyboun@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
Ph.D.	METU Philosophy	2022
MA	Boğaziçi University Philosophy	2013
BA	Bilkent University Philosophy	2011
High School	UHAL	2004

WORK EXPERIENCE

Year	Place	Enrollment
2014-2021	METU Philosophy	Research Assistant
2013-2014	Boğaziçi University Philosophy	Research Assistant

FOREIGN LANGUAGES

English

B. TURKISH SUMMARY/TÜRKÇE ÖZET

ELEYİCİ ANLAM KURAMININ BİR SAVUNUSU: BAĞLANTISALCI BİR YAKLAŞIM

İnsanın dil kullanma becerisi onu diğer hayvanlardan ayıran en önemli bilişsel özellik olarak görülmektedir. Dolayısıyla, dil becerileri genellikle diğer becerilerinden ayrı tutulmuştur. Hatta birçok felsefi, kültürel yaklaşım dili sadece biyolojik yapımızdan değil; neredeyse fiziksel dünyadan da soyutlamaktadır. Natüralist bir yerden baktığımızda ise, dil, tıpkı diğer bilişsel özelliklerimiz gibi, biyolojik yapımızın bir uzantısıdır. Ancak, bu durum, dilin diğer özelliklerimizden soyutlanarak incelenebilmesinin önünde bir engel değildir.

Dil üzerine yapılan çalışmaların alanı oldukça geniştir. Belirli bir doğal dile ait seslerin nasıl çıkarıldığı, birleştirildiği, yazı dile ile nasıl eşleştirildiği, harflerin/seslerin kelimelerle, kelimelerin daha üst dilsel ifadelerle ilişkisi, anlam, imalar vs. dil üzerine çalışan araştırmacıların ilgilendiği konular arasındadır. Bu konular, evrimsel, gelişimsel, genetik, sosyolojik, antropolojik açılardan incelenebildiği gibi zamandan ve meakandan bağımsız olarak soyut bir şekilde de çalışılmaktadır. Bu çalışmalar birbirlerini beslemektedirler. Mesela, dil edinimi üzerine çalışan bir araştırmacı, morfoloji üzerine çalışan bir araştırmacının görüşlerini kendi çalışmalarında kullanmaktadır. Aynı şekilde Almanca'nın fonolojik yapısı üzerine çalışan bir araştırmacı, ana dili Almanca olan çocukların dil edinimi üzerine yapılmış çalışmalardan yararlanmaktadır.

Felsefecilerin bu konudaki rolü ise, genel çerçevenin soruşturulması, temel postulatların sorgulanması, temellendirilmesidir. Yapay zeka çalışmaları, bilgisayar bilimi, bilişsel bilimler, psikoloji, sosyoloji, antropoloji ve mantık biliminin ilgi alanında yer alan dil çalışmaları henüz gelişmekte olan bir alan olduğu için felsefi müdahalelere ihtiyaç duymaktadır.

Bu tezin konusu ise herhangi bir doğal dil kullanıcısının o doğal dile ait ifadeleri nasıl anladığıdır. İnsanların doğal dilleri hangi süreçleri izleyerek anladıkları felsefenin ve bilimin kesişim alanında olan güncel konulardan birisidir. Bu konuda iki temel çerçeveden bahsetmek mümkündür. Bunlardan birincisi insanların doğal dilleri nasıl anladığını mekanik süreçlere indirgemenin mümkün olmadığını iddia eden görüştür. Diğer görüş ise bilimin konusu olan diğer doğal süreçler gibi dil anlamanın da mekanik süreçlerden ibaret olduğu şeklindedir.

Rene Descartes kötümser olan ilk görüşün önemli bir temsilcisidir. Descartes herhangi bir doğal dili anlamanın insanı makinelerden ayıran en önemli özellik olduğunu iddia etmektedir. Dolayısıyla, ona göre insanların dil anlama süreçleri mekanik bir şekilde işlememektedir.

> Bedenlerimize benzeyen ve eylemlerimizi taklit eden makineler olsaydı, yine de onların gerçek insan olmadıklarını anlamamız için çok kesin iki sebebimiz olurdu. ...onlar asla kelimeleri ve işaretleri bizim düşüncelerimizi başkalarına bildirmek için yaptığımız gibi kullanamazlardı. Bir makine sözcükleri ifade edecek şekilde yapılmış olabilir. Mesela belirli bir yerine dokunulduğunda, ona ne söylemek istediğinizi sorabilir; başka bir yerine dokunulursa, incindiğini haykırabilir. Fakat, onun, en akılsız insanın bile yapabileceği şekilde, kelimeleri kendisi düzenleyerek her şeye cevap vermesi beklenemez (Descartes (1968), pp. 73-74).

Descartes'in çağdaşı olan Thomas Hobbes ise karşı görüşün önemli bir temsilcisi olarak gösterilebilir. Hobbes bütün bilişsel süreçlerin mekanik bir şekilde işlediğini iddia etmektedir. Bütün bu süreçler temelde toplama çıkarma gibi basit aritmetik işlemlere indirgenmektedir.

> Muhakeme derken, hesaplamayı kastediyorum. Şimdi hesaplamak, birbirine eklenen birçok şeyi toplamak; bir şey diğerinden çıkarıldığında geriye ne kaldığını bilmektir. Bu nedenle, muhakeme, toplama ve çıkarma ile aynıdır; ve herhangi birisi çarpma ve bölmeyi de eklerse, buna karşı olmayacağım. Zira çarpmanın birbirine eşitlerin toplanmasından başka bir şey olmadığını ve bölmenin mümkün olduğu kadar sık bir şekilde eşitlerin birbirinden çıkarılmasından başka bir şey olmadığını görmekteyiz. Böylece tüm muhakeme, zihnin iki işlemi olan toplama ve çıkarma işlemlerinde kavranır" (Hobbes (1655), p. 3).

Bir anlamda bu geleneğin en önemli temsilcisi Alan Turing doğal dil anlama sürecinin makineler tarafından taklit edilebileceğini iddia etmektedir (Turing (1950), p. 442). Bu çalışmada doğal dil anlamanın tamamen mekanik bir süreç olduğunu iddia eden geleneğin devamı olarak bu sürecin felsefi, mantıksal temelleri üzerinde durulmaktadır.

Bu konu ile ilgili olarak, insanlık düşünce tarihindeki yaygın görüş, doğal dil anlama sürecinin, dilsel ifadelere karşılık geldiği düşünülen "anlam" kavramı üzerinden modellenebileceği şeklindedir. Dolayısıyla, doğal dil anlama sırasındaki mekanik işlemler "anlam" denilen yapıları temsil eden semboller üzerinden ilerlemektedir. Diğer bir ifadeyle, anlam bu yaygın görüşün temel postulatıdır.

Bu tezde, çok anlamlılık, bağlamsallık gibi olgular nedeniyle anlam kavramının modelleme için fazlaca esnek ve belirsiz olduğu iddia edilmektedir. Dolayısıyla da, anlam gibi belirsiz bir postulat üzerine inşa edilen sembolik yaklaşım reddedilmektedir. Alternatif olarak, dil anlama sürecinin anlam kavramına başvurulmadan modellenmesi gerektiği savunulmaktadır. Dil anlama sürecinin anlam kavramı olmadan nasıl modelleneceği konusunda ise bağlantıcı görüşe başvurulmaktadır.

Öncelikle, dil anlama sürecinin anlam üzerinden modellenmesi gerektiğini savunan, sembolik yaklaşım, bileşimsellik ilkesi üzerinden yola çıkmaktadır. Bileşimsellik ilkesi anlamın da ifadeler gibi parça bütün ilişkisi içerisine olduğunu iddia etmektedir. Anlamdaki parça-bütünsel yapı ifadelerdeki parça-bütünsel yapıyı izlemektedir. İfadeler arasındaki ilişkileri bir cebirle, anlamlar arasındaki ilişkileri de başka bir cebirle ifade ettiğimizde, bileşimsellik ilkesine göre birinci cebirden ikinci cebire homomorfik bir ilişki tanımlıdır.

A : ifadeler kümesi

B : anlamlar kümesi

F : ifadeler arasındaki ilişkiler

G : anlamlar arasındaki ilişkiler

Γ : indeksler kümesi olsun.

Bileşimsellik ilkesine göre $< A, F_\gamma >_{\gamma\in\Gamma}$ sözdizimsel cebirinden $< B, G_\gamma >_{\gamma\in\Gamma}$ anlambilimsel cebirine homomorfik bir ilişki tanımlıdır (Dowty (2007), p. 34) (Pagin (2003), p. 299) (Barbara H. Partee (1990), p. 334). Kısaca, bileşimsellik ilkesine göre, karmaşık bir ifadenin anlamı içerdiği basit ifadelerin anlamlarının özel bir fonksiyonudur.

Bileşimsellik ilkesinin dil ile ilgili iki temel olguyu açıklaması, onun yaygın bir şekilde kullanılmasını sağlamaktadır. Bu olgulardan birincisi dilin üretkenliğidir. Sonlu sayıda basit ifadenin (kelime) anlamını hafızasında tutan birisi, potansiyel olarak sonsuz sayıda ifadeyi anlayabilmektedir. Mesela, daha önce hiç duymadığınız "Wittgenstein Antalya'da yaşamaktadır" ifadesini kolayca anlayabilmektesiniz. Bileşimsellik ilkesine göre bu ifadenin sözdizimsel yapısı ve içerisinde geçen basit ifadelerin anlamları kullanılarak bütün bir ifadenin anlamı hesaplanabilir. Dolayısıyla, bileşimsellik ilkesi bu ifadenin dil kullanıcıları tarafından nasıl anlaşıldığını açıklamaktadır. Hatta aynı yöntemi izleyerek, bir makine de potansiyel olarak sonsuz sayıda ifadenin anlamını hesaplayabilir. Yani, bu yaklaşıma göre, kelimelerin anlamları ve anlamların birleşme kuralları makineye girdi olarak verildiğinde, makine bu kelimelerden oluşan herhangi bir ifadenin anlamını, mantıksal formunu çıktı olarak verebilir.

İkinci olgu ise dilin sistematikliğidir. "Ali Ayşe'ye çiçek verdi" ifadesini anlayan birisinin "Ayşe Ali'ye çiçek verdi" ifadesini de anlaması beklenir. Bileşimsellik ilkesi açısından $\ulcorner a$ b'ye çiçek verdi$\urcorner$ ifadesini anlayan birisi a ve b ifadelerinin anlamlarının yerlerini değiştirerek $\ulcorner b$ a'ya çiçek verdi$\urcorner$ ifadesinin anlamını hesaplayabilir.

Bileşimsellik ilkesi bu olguları açıklaması sebebiyle, antik dönemden bugüne insanlık düşüncesindeki yerini korumaktadır. Aristo, Farabi, Ibn-i Sina, Peter Abelard, Thomas Hobbes, Gottfried Wilhelm Leibniz gibi filozoflar bu ilkeye açık bir şekilde dikkat çekmişlerdir. Çağdaş yaklaşımın temellerini ise Gottlob Frege'nin matematiğin temelleri üzerine yaptığı çalışmalar oluşturmaktadır (Frege (1948), p. 217). Frege, matematiğin temelerini ortaya atmayı amaçladığı çalışmalarında öncelikle matermatiksel önermeleri formüle edebileceği bir dile ihtiyaç duymuştur. Bu çalışmalar bir taraftan modern mantığın gelişmesine yol açarken; diğer taraftan da çağdaş anlambilimin doğmasına neden olmuştur.

Frege, matematiksel önermelerin ifade edilebileceği formel dillerden, doğal dillere kadar uzanan çalışmalarında, bileşimsellik ilkesinin anlambilim için önemine dikkat çekmiştir. Bu nedenle, bu ilkeye Frege'nin ilkesi de denilmektedir. Son dönem eserlerinde psikodilbilim alanına girerek dilin üretken yapısının önemini vurgulamıştır.

> Dilin neler yapabileceği şaşırtıcıdır. Sadece birkaç hece ile hesaplanamayacak kadar çok düşünceyi ifade edebilir, öyle ki, ilk kez bir insan tarafından kavranan bir düşünce bile, düşüncenin tamamen yeni olduğu birisinin anlayacağı bir ifade biçimine getirilebilir. Cümlenin yapısına tekabül eden düşünce parçalarını ayırt

> edemeseydik, bu imkansız olurdu, böylece cümlenin yapısı düşüncenin yapısının bir görüntüsü olarak hizmet eder... Düşünceleri basit parçalardan oluşmuş olarak kabul eder ve bunların da cümlelerin basit parçalarına karşılık geldiğini kabul edersek, birkaç cümle parçasının nasıl olup da büyük bir cümle çokluğu oluşturabileceğini anlayabiliriz. Şimdi, düşüncenin nasıl inşa edildiği, parçaların nasıl bir araya geldiği ve bütünün, nasıl olup da tek tek parçalardan daha fazlası olduğu sorusu ortaya çıkmaktadır (Frege (1963), p. 1).

Frege sonrasında, özellikle Ludwig Wittgenstein, Richard Montague, Donald Davidson, Jerry Fodor, Jerrold Katz, Barbara Partee, Irene Heim gibi araştırmacılar dile bileşimsellik ilkesi üzerinden yaklaşmışlardır. Bu yaklaşım dil üzerine çalışan disiplinlerde yaygın hale gelmiştir. Diğer taraftan *deyimler, bağlamsallık, muğlaklık* gibi durumlar bileşimsellik ilkesi için sorun oluşturmaktadır. Bu sorunlara getirilen çözüm önerileri bu alandaki çalışmaların önemli bir bölümünü oluşturmaktadır.

Özellikle kelime anlamı düzeyindeki muğlaklık bu çalışma açısından önemlidir. Kelime anlamı düzeyindeki muğlaklık sorunu bir örnekle ifade edilebilir. Bileşimsellik ilkesine göre “Ahmet çay siparişi verdi” ifadesinin anlamı bu ifadenin parçalarının anlamlarının bir fonksiyonudur. Fakat “çay” kelimesi birden fazla anlama sahiptir. Bir taraftan derenin küçüğü olan akarsu anlamına gelirken, diğer taraftan sıcak içeceği yapılan bir bitki türüne verilen isimdir. Bileşimsel yaklaşım çerçevesinde muğlaklığı çözmenin iki temel yolu vardır. Bunlarda birincisi bileşimsel hesaplama işlemine girmeden önce kelimelerin anlamlarını netleştirmektir. İkincisi ise bütün olası anlamları hesaplayıp, bağlam içerisinde hangisinin uygun olduğuna daha sonra karar vermektir. Mesela, n kelimeden oluşan bir ifade için, her bir kelimenin ortalama m farklı anlamı olduğunu varsayıldığında, m^n farklı aday anlam hesaplanır. Bağlama bakılarak bunlardan bir tanesinin ifadenin anlamı olduğuna karar verilir.

Muğlaklığın bir özel bir türü olan çok anlamlılık sorununda ise çözüm bu kadar kolay görünmemektedir. Çok anlamlılık sorununda bir kelimenin farklı anlamları arasında ilişki bulunmaktadır. Mesela, “çay” kelimesin birbirleriyle ilişkili olan *sıvı içecek, bu içeceğin yapıldığı bitki, bu içecekle dolu olan bardak, bu içeceğin tüketildiği toplantı,* vs. gibi anlamları bulunmaktadır. Birbirleriyle bağlantılı anlamlar listesi yaratıcılığımıza bağlı olarak sınırsız bir şekilde genişletilebilir. Bu soruna getirilen geleneksel çözüm birbirleriyle bağlantılı anlamları farklı bir şekilde sıralamaktır. Mesela, “çay” kelimesinin *akarsu* ve *bitki* anlamları ayrı ayrı sıralandıktan sonra bitki başlığı altında yukarıda saydığımızi birbirleriyle ilişkili olan anlamlar sıralanabilir. Ancak, bu listenin ucu açık olduğu için geleneksel çözüm gerçekçi görünmemektedir (Pustejovsky (1996), p. 127).

Çağdaş literatürde, kelimelerin birbirleriyle yakın anlamlarını tek tek sıralamak yerine, bu anlamların bağlam içerisinde türetilmesine yarayacak bilgilerin girilmesi önerilmektedir. Bu konuda iki farklı yaklaşım bulunmaktadır. Bunlardan ilki, kelimeler için zengin içerikli girdiler kullanmayı öneren dilbilim temelli yaklaşımdır. İkincisi ise girdilerin sadece genel bilgiler sağladığı pragmatik temelli yaklaşımdır. Bu iki görüş bileşimsellik ilkesinin iki farklı uygulamasını ortaya atmaktadır. Birincisi, kural tabanlı, dilbilimsel yaklaşım; ikincisi ise bağlam tabanlı, pragmatik yaklaşımdır. Bu çalışmada her iki görüş de dil anlamanın esnek, dinamik yapısını modellemekten uzak olmaları gerekçesiyle reddedilmektedir.

Kural tabanlı yaklaşım, bağlamsal katkının ifadelerdeki değişkenler üzerinden kontrol edilebileceğini iddia etmektedir. Dolayısıyla, bu görüşe göre bağlam kelime girdilerindeki değişkenleri, yani boşlukları, bir anlamda doyurmaktadır. James Pustejovsky'nin üretken sözlük kuramı bu yaklaşımın en iyi örneklerinden birisidir(Pustejovsky (1996)). Kural tabanlı yaklaşımın önemli bir savunucusu olan Jerry Fodor bu görüşü şöyle özetlemektedir.

> ...kelime anlamının bağlama bağlı varyasyonları konusuna genellikle olduğundan fazla önem verilmektedir, çünkü diğer bağlam duyarlılığı türleri, yanlışlıkla, bileşimsellik ilkesinin ihlalleri olarak yorumlanmaktadır. Örneğin, 'tavuk beslemek' ve 'yemek için tavuk' arasındaki fark, bir bileşimsellik ihlalinden ziyade hayvan/gıda muğlaklığından ibarettir. Çünkü 'besle...' ifadesine argüman olarak tıpkı 'dana besle', 'domuz besle' örneklerinde olduğu gibi bir hayvan gelmesi beklenir. Benzer şekilde, 'iyi kitap', 'iyi dinlenme' ve 'iyi dövüş' arasındaki fark, anlam değişiminden ziyade ifadenin sinkategorematiliğidir... iyi bir kitap, kitaplara olan ilgimize cevap veren kitaptır (yani, okumak güzeldir); iyi bir dinlenme, dinlenmeye olan ilgimize cevap verir (yani, kişiyi tazeler); iyi bir dövüş, dövüşlere olan ilgimize cevap veren bir dövüştür (yani izlemesi ya da içinde olması eğlencelidir)... Benzer şekilde " iyi bir flurg" flurglara olan ilgimize cevap veren flurgtır (Fodor and Pylyshyn (1995), pp. 124-125).

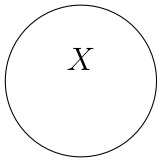
Bu çalışmada kural tabanlı yaklaşım eleştirilerek, çok anlamlılık konusuna gereğinden fazla önem affedilmediği, tam tersine, bu konunun hafife alındığı iddia edilmektedir. Bu iddia üç argümanla desteklenmektedir. İlk olarak, kelimelere karşılık geldiğini düşündüğümüz anlamlar izi sürülemez derecede karmaşıktır. Felsefe tarihindeki çıkmazlar bunun bir örneğidir. Bir disiplin olarak felsefenin konusu kavramlardır. *Adalet, aşk, gerçeklik, bilgi, teknoloji, zaman, uzay*, vs. gibi kavramları tanımlamak felsefe tarihinin en belirgin konularıdır. Felsefe bu listedeki görece daha yaygın kullanılan kavramlardan ziyade akla gelebilecek bütün kavramların analizini kendisine konu edinmektedir.

Fakat Sokrates'ten itibaren kavramları tanımlamanın, sınırlandırmanın pek mümkün olmadığını görmekteyiz. Platon'un diyaloglarında, Sokrates'in tartıştığı insanlar herhangi bir kavram için bir tanım ortaya attığında, Sokrates bağlamı değiştirerek onları belirsizlik, *aporia*, içerisinde bırakmaktadır. Bu soruna çözüm olarak ortaya çıkan Platoncu idealizm kavramların gerçeklerinin bizim yaşadığımız dünyadan uzak olduğunu söylerken, Aristotelesçi realizm ise kavramların içinde bulunduğumuz dünyanın bir bileşeni olduklarını ve ancak uzun metafiziksel soruşturmalar sonucunda ortaya çıkarılabileceklerini iddia etmektedir. Platoncu ve Aristocu görüşlerin bir açılımı olan felsefe tarihinde kavramları netleştirmek konusunda önemli bir ilerleme görülmemiştir. Bu duruma cevap olarak yirminci yüzyılda ortaya çıkan, felsefeyi gündelik dil kullanımı üzerinden inceleyen, J. L. Austin, Ludwig Wittgenstein, Gilbert Ryle gibi filozofların başını çektiği bir grup ise kavramları tanımlama, sınırlandırma arayışının başından yanlış bir uğraş olduğunu iddia etmektedir. Sonuç olarak kelimelere bağlam içerisinde karşılık gelen kavramları birkaç kuralla ortaya çıkarmak gerçekçi görünmemektedir.

Buradaki çıkmazın kaynağı olarak kavramların matematiksel anlamda sık dağılımı gösterilebilir. Herhangi iki kavram arasına bu iki kavrama eşit derecede yakın bir başka kavram eklenebilir. Bu sık yapıyı, sık olmayan kelimelerle ve birkaç dilsel kuralla temsil etmeye çalışmak sorunun kaynağını oluşturmaktadır.

İkinci argüman ise dilin, ya da dilsel anlamın kavram dünyamızdaki karmaşıklıktan bağımsız olmamasıdır. Yani yukardaki eleştiriden, dilsel anlamı kavramsal yapımızdan ayırarak kaçınmak mümkün değildir. Bu argümanın kaynağını bilişsel anlambilim kuramının ortaya çıkardığı veriler oluşturmaktadır. George Lakoff, Mark Johnson, Charles Fillmore, Ronald Langacker, Gilles Fauconnier, Claudia Brugman gibi bilişsel anlambilimciler *imge şemaları, kavramsal metaforlar* ve *ad aktarımları* üzerinden dil ile kavramsal yapımız arasındaki ilişkiyi göstermektedirler.

Analog bir yapıda olan imge şemaları özellikle edatların kullanımını kontrol etmektedir.



(Johnson (2013), p. 23)

- “Kansas eyaletin**de** çok fazla arazi bulunmaktadır."
- “Gemi görüş alanımız**da**."
- “Depresyon**da**."

Yukardaki ifadelerde, *-de, -da, -te, -ta* bulunma ekleri içerme şemasını örneklendirmektedirler. Bu eklerin yakın anlamları ise bir imge şemasının başka bir imge şemasına dönüşmesiyle ortaya çıkmaktadır.

Benzer şekilde metaforlardaki sistematik yapı kavramlarımızı yansıtmaktadır. Mesela aşağıdaki metaforda aşk kavramı yolculuk kavramı üzerinden anlaşılmaktaır.

Aşk yolculuktur

- Ne kadar yol kat ettiğimize bir bak.
- Bir yol ayrımındayız.
- Bu ilişkinin bir yere varacağını düşünmüyorum.
- Uzun ve engebeli bir yoldu.
- Bu ilişki çıkmaz bir sokaktan ibarettir. Sadece tekerleklerimizi döndürüyoruz.
- Yoldan çıktık.
- İlişkimiz batıyor.

Şu örneklerde ise *kuram* ve *bina* kavramları arasındaki sistematik ilişkiler göze çarpmaktadır.

Kuramlar binadır

- Kuramınızın temeli bu mu?
- Kuramınızın daha fazla desteğe ihtiyacı var.
- Argümanlar sallantıda.

- Bunun için güçlü bir argüman oluşturmamız gerekiyor.
- Argüman çöktü.

Diğer taraftan metaforlara göre daha kurallı ve basit görünen ad aktarmaları da kavramsal yapımızın bir uzantısıdır. Mesela, "Wittgenstein okuyorum" ifadesini anlarken kitaplar ve yazarları arasındaki kavramsal ilişki önemli bir rol oynamaktadır.

Üçüncü argüman ise kavramların bu denli karmaşık yapısının kötü bir rastlantıdan ziyade iletişim için faydalı bir özellik olduğudur. Bu argümanın temelleri George Kingley Zipf'in çalışmalarında bulunabilir. Zipf grup olarak yapılan eylemlerde bilişsel ekonominin önemine dikkat çekmiştir. Dilsel iletişim sırasında konuşmacı ve dinleyici belli bir çaba sarf eder. Dinleyici açısından, ifadelerin olabildiğince net olması arzulanırken, konuşmacı ise olabildiğince az sayıda kelime kullanarak kendisini ifade etmek ister. Yorgun olduğumuzda "şey" kelimesini sık kullanmamız buna bir örnek olarak gösterilebilir. Zipf makul düzeyde muğlaklığın bilişsel ekonomi için gerekli olduğunu iddia etmektedir. Daha çağdaş çalışmalarda ise bilişsel ekonomi entropi üzerinden hesaplanmaktadır.

$$H[Anlam] = -\Sigma_{m \in Anlam} P(m) log P(m)$$

$$H[Anlam|Balam] = -\Sigma_{c \in Balam} P(c) \Sigma_{m \in Anlam} P(m|c) log P(m|c)$$

Yukardaki formüllerde görüldüğü üzere koşullu entropi koşulsuzdan yüksek olamayacağı için hazır bir kaynak olan bağlamı kullanmamak ekonomik olmayan bir davranıştır (Piantadosi et al. (2012), p. 283). Almanca ve İngilizce üzerinde yapılan çalışmalar, fonotaktik açıdan kolay tahmin edilebilir kelimelerin daha muğlak olduğunu göstermektedir (Piantadosi et al. (2012), pp. 285-287).

Bunlardan daha önemlisi ise çok anlamlılık tarafından sağlanan küçük dünya ağ modelidir. Küçük dünya hipotezine göre iyi bağlanmış bir ağda herhangi iki birim arasında az sayıda aracı birim bulunmaktadır. Ağ modeline göre bu birimler nöron, bilgisayar, şirket, insan vs. olabilir. Bizim konumuzda ise bu birimler kavramlardır. Bu açıdan bakıldığında, dildeki çok anlamalılık özelliği bir sorundan ziyade kavramların birbirlerine daha iyi bağlanmasına yol açan bir özellik olarak karışımıza çıkmaktadır. Mesela, “kitap” kelimesinin soyut anlamı ile somut anlamı, ağaç ve eğitim bakanlığı kavramlarını birbirlerine bağlamaktadır.

Doğal dil anlama süreçlerini bileşimsellik ilkesi üzerinden çalışan bir başka grup ise radikal bağlamsalcı olan pragmatik temelli yaklaşımdır. Kural tabanlı yaklaşıma getirilen eleştirilerin benzerlerini François Recanati, Robyn Carston, Deirdre Wilson, Dan Sperber, Ingrid Falkum gibi pragmatik temelli yaklaşımın önde gelen savunucuları da getirmektedir. Pragmatik temelli görüş bağlamsal katkının birkaç değişken ve dilsel kuralla aşağıdan yukarı doğru kontrol edilemeyeceğini iddia etmektedir. Bağlamsal katkının genellikle yukarıdan aşağı doğru, çok daha kapsamlı bir şekilde olduğunu iddia etmektedirler. Dolayısıyla, bağlamı dil anlama süreci içerisinde çok önemli bir yerde konumlandırmaktadırlar.

Pragmatik temelli görüşün savunucuları kelimelerin bileşimsel işleme girmeden önce yukarıdan aşağı doğru bir işlem olan modülasyona maruz kaldıklarını iddia etmektedirler. Kelime girdilerindeki genel bilgiler, sözdizimsel yapı, anlık bağlam ve dil anlayan kişinin hafızasında yer alan bilgiler modülasyon işleminde rol oynamaktadırlar.

Pragmatik temelli yaklaşımın iki versiyonu bulunmaktadır. Birincisi, *Bağıntı kuramı* ikincisi ise François Recanati'nin geliştirdiği *çağrışımcı* yaklaşımdır. Bağıntı kuramı Paul Grice'ın, daha sonra “Mantık ve Konuşma” başlığıyla yayımlanan, 1967 tarihli William James derslerine dayanmaktadır. Grice bu derslerde mantıksal bağlaçların kullanımını kendisine konu olarak almıştır. Mesela, *ve* bağlacının mantıktaki kullanımında, bu bağlaç ile bağlanan önermelerden her ikisi de doğru olduğunda sonuç doğru olur. Ancak, *ve* bağlacının doğal dillerdeki bazı kullanımlarında zaman bilgisi de eklenmektedir. İlk önermenin ikinci önermeden daha önce olduğu belirtilmektedir. Bu açıdan aşağıdaki iki cümle farklı doğruluk değerleri taşımaktadırlar.

- Evlendirler ve çocukları oldu.
- Çocukları oldu ve evlendiler.

Aynı şekilde, “çoğu” ifadesenin mantıktaki kullanımı doğal dildeki kullanımından farklıdır. Mesela, “Pastanın çoğunu ben yedim” ifadesini yüklemsel mantıkta varlıksal niceleyici kullanarak yazdığımızda pastanın hepsini yediğim şeklindeki anlamı dışlamış olmam. Ancak, doğal dilde, "Pastanın çoğunu ben yedim" ifadesi pastanın hepsini yemediğim anlamına gelmektedir. Bu duruma getirilen geleneksel yaklaşım bu bağlaçların iki farklı anlamı olduğu, bağlama göre bunlardan birisinin kullanıldığını öne sürmektedir.

Paul Grice ise Occam'ın usturası ilkesinden yola çıkarak, iki farklı anlam ortaya atmak yerine bunlardan birisinin asıl anlam olduğunu; diğerinin ise belirli çıkarımsal mekanizmaların sonucunda ortaya çıktığını iddia etmektedir. Mesela "çoğu" kelimesinin ilk anlamı sembolik mantıkta kullanıldığı halidir. Ancak, dinleyici belirli koşullar sağlanmadığında bu ifade için ikincil anlamlar aramaktadır. Bu koşulları, Grice, iletişimin işbirlikçi yapısı üzerine kurmuştur. Grice'ın modelinde, dilsel bir iletişim sırasında, tarafların rasyonel ve birbirleriyle iş birliği içerisinde oldukları varsayımı yatmaktadır. Bu durum iletişimin taraflarının gerektiği kadar bilgi sağlama, konu ile alakalı olma, doğru olduğuna inandığın şeyleri söyleme vs. gibi ilkelere uymasıyla sağlanır.

Örnek olarak, bir felsefe öğrencisi için yazılan tavsiye mektubunda öğrencinin sadece Türkçe becerilerinden söz ediliyorsa bu durum öğrencinin felsefe açısından başarısız olduğu anlamına gelmektedir. Grice'a göre bunun nedeni, mektubu yazan profesörün rasyonel ve işbirlikçi olduğuna dair inancımızdır. Mektubu yazan profesörün ne az, ne fazla, gerektiği kadar bilgi sağlayacağını varsaydığımızda, öğrencinin felsefi yetenekleri ile ilgili olarak hiçbir yorum yapmamasının arkasından başka bir neden ararız. Dilsel, kültürel konvansiyonları da eklediğimizde profesörün öğrencinin felsefe alanındaki yetenekleriyle ilgili görüşlerinin olumsuz olduğunu düşünürüz.

Bağıntı kuramı Grice'ın ortaya attığı çıkarımsal mekanizmayı, sadece ikincil anlamları hesaplamak için değil, bütün bir anlamın hesaplanması için temel mekanizma olarak görmektedir. Bu açıdan bağıntı kuramı iletişimi kodlama ve kod çözme olarak gören Shannon-Weaver modeline karşı çıkmaktadır. Kodlama modelinde, konuşmacı düşüncelerini kelimeler aracılığıyla kodlayarak karşı tarafa gönderir. Bu mesajı bir paket olarak alan dinleyici ise kodları çözerek, düşünceyi ortaya çıkarır (Weaver (1953), p. 264) (Shannon (1948), p. 381) (Wilson and Sperber (2006), p. 5).

Bir önceki bölümde tartışılan kural tabanlı yaklaşım da aslında iletişimin kod modeline bir örnektir (Falkum (2011), pp. 77-82). Wilson ve Sperber bu model için şöyle demektedirler. "Aristoteles'ten modern göstergebilime kadar, tüm iletişim teorileri, kod modeli olarak adlandıracağımız tek bir modele dayanmaktadır" (Sperber and Wilson (1995), p. 2).

Bağıntı kuramı Grice'ın geliştirdiği kuramın öğelerini kullansa da, Grice'ın yaklaşımı ile bağıntı kuramı arasında çok önemli farklılıklar bulunmaktadır. Her şeyden önce, Grice'ın yaklaşımı kural tabanlı sistemden bir sapma değildir. Tıpkı kural tabanlı sistemde olduğu gibi Grice'a göre de bir ifadenin anlamı, kelime anlamı ve bir dizi bileşimsel kurallar üzerinden hesaplanmaktadır. Grice'ın yaklaşımı iki aşamalı bir yaklaşımdır. İlk adımda, birincil anlamlar hesaplanır. Birincil adımda hesaplanan anlamlar Grice'ın bahsettiği işbirliğine dayalı ilkeleri ihlal ettiğinde ise ikincil anlamlar hesaplanmaktadır. Özetle, Grice'ın kodlamaya dayalı iletişim modeline karşı çıktığı söylenemez.

Diğer taraftan, bağıntı kuramına göre anlam, birincil – ikincil ayrımı yapılmaksızın, sadece çıkarımlar üzerinden hesaplanır. Ayrıca, bağıntı kuramındaki çıkarımsal süreçler temelde Grice'ın ikincil anlamın hesaplanması için önerdiği yöntemden farklıdır. Grice'ın işbirlikçi ilkeleri yerine, temel dayanak noktası yorumun *alaka* düzeyidir (Wilson and Sperber (2006), pp. 614-615, 619-620).

Alaka düzeyine dayalı yaklaşımın arkasındaki motivasyon, evrimsel faydalarıdır. Yani az çaba harcayarak *olumlu bilişsel etkiyi* arttırmaktır. "Olumlu bir bilişsel etki, bireyin dünya temsilindeki dikkate değer bir değişimdir (Wilson and Sperber (2006), p. 608). Benzer şekilde, "bağıntı kuramının diliyle ifade edersek, diğer şeyler eşit olduğunda, İŞLEME EFOR gereksinimi ne kadar fazlaysa, girdi de o derecede alakasızdır" (Wilson and Sperber (2006), p. 609). Bu nedenle, bağıntı kuramındaki çıkarımlar yalnızca bir ilkeye uymalıdır: alaka düzeyini en üst düzeye çıkarmak (Falkum (2011), pp. 98-106).

Bağıntı kuramının birbirleriyle ilişkili iki ilkesi bulunmaktadır. Bunlar

- Bilişsel Bağıntılılık İlkesi
- İletişimsel Bağıntılılık İlkesi

Birincisi, dil kullanıcılarının alaka düzeylerini en üst noktaya çıkarma stratejileridir (Wilson and Sperber (2006), p. 610). İkincisi, konuşmacıların bilgilendirici ve iletişimsel niyetlerinden oluşan, Gösterimsel - Çıkarımsal iletişim modelidir (Wilson and Sperber (2006), ss. 611-612). İkinci ilke, Grice'ın doğal olmayan anlam kuramına dayanmaktadır. 1957 tarihli "Anlam" makalesinde Grice, konuşmacıların niyetleri üzerinden doğal anlamı doğal olmayan anlamdan ayırmaktadır. Örneğin, birinin yüzündeki kırmızı noktalar, o kişinin hasta olduğu anlamına gelebilir. Bu durum doğal anlam için örnek teşkil etmektedir. Ancak dilsel anlamda, konuşmacılar dinleyicileri bilgilendirmek isterken aynı zamanda bu niyetlerinin dinleyici tarafından bilinmesinin de isterler (Grice (1957)). Konuşmacının niyetini bildirme konusundaki isteğini bağıntı kuramcıları da iddia etmektedir. Bağıntı kuramına göre, konuşmacılar niyetlerini iletişimin her iki tarafının da açıkça görebileceği öğeler kullanarak yaparlar. Başka bir deyişle, ortak payda veya ortak bilgi/inanç modellerinin (Clark (1996), ss. 92-121) (Lewis (2002), ss. 52-82) (Stalnaker (2002)) yerini ifadeleri karşılıklı olarak ortaya koyma kavramı almaktadır. Konuşmacının kullandığı kelimeler iletişimin her iki tarafına da açık olarak görünen nesneler kümesine eklenmiş olur (Wilson and Sperber (2006), 1996, pp.58-63) (Falkum (2011), pp. 91-94). Bu noktada zihin kuramına da başvurulmaktadır.

Her iki katılımcının da halihazırda maruz kaldığı anlık algısal girdiye ek olarak, iletişimci, iletişimsel amacını gerçekleştirmek için her iki tarafa da açıkça görünen uyaranlar kümesine eklemeler yapmaktadır. Gösterimsel-çıkarımsal iletişim modeline göre, "her gösterimsel uyaran, kendi optimal bağıntılılığının bir varsayımını taşır" (Wilson and Sperber (2006), s. 256). Dilsel iletişimde her kelime bir gösterimsel uyarıcıdır. Kelime anlamları, bağlamsal bileşenlerle birlikte, dinleyiciye alaka düzeyine dayalı bir çıkarım yapabileceği ipuçları olarak verilir. Dolayısıyla, sistem temelde kombinasyon kuralları ve eksik olan kelime anlamlarının üzerine kurulmuştur. Bu araçlarla dil kullanıcısı öncelikle ifadelerin mantıksal formlarını hesaplamaktadır. Bu işlemin ardından bağıntılılık ilkesi kullanılarak, ifadelerin anlamları hesaplanmaktadır (Sperber and Wilson (1995), ss. 172-183) (Carston (2015), s. 198).

Recanati, bağlamsalcı yaklaşımın bağıntı kuramına alternatif olan diğer bir versiyonunu ortaya atmıştır. Bağıntı kuramına benzer şekilde, kelime anlamı, bileşimsel işleme girmeden önce pragmatik işlemlerden geçmektedir. Karmaşık bir ifadenin bileşimsel anlamı aşağıdaki gibi formüle edilmektedir.

$$I(\alpha^\wedge\beta) = f(g_1(I(\alpha)), g_2(I(\beta)))$$

Yukarıdaki formülde I, yorumlama fonskiyonudur. f, g_1 ve g_2 ise sırasıyla, bileşimsellik ve pragmatik modülasyon fonksiyonlarına karşılık gelmektedirler. Recanati bu işlemleri pragmatik bileşimsellik olarak da adlandırmaktadır (Recanati (2010), s. 128-129) (Recanati (2004), pp. 138-140). Formüldeki modülasyonlar birincil pragmatik süreçler olarak adlandırılır. Birincil pragmatik süreçler asıl ifade edileni hesaplamada kullanılırken; ikincil pragmatik süreçler ise ima edileni vermektedirler. Bağıntı kuramında olduğu gibi, birincil pragmatik süreçler kavram genişletme/daraltma ve kavram aktarımı gibi modülasyonlardan oluşmaktadır (Recanati (2017), p. 379) (Recanati (2004), p. 23).

Pragmatik temelli yaklaşımda, bağlam, kelime girdilerindeki temsili yapılarla etkileşimin mümkün olması için, temsiller üzerinden işlenmek durumundadır. Bağlamın temsili, şu an duyduğum araba seslerinden, televizyonda gördüğüm bir görüntünün çağrıştırdığı çocukluk anılarıma kadar neredeyse herşeyi kapsamaktadır. Bağlam temsilinin pratik olarak mümkün görünmemesi pragmatik temelli görüşü reddetmemiz için önemli bir nedendir. Ayrıca, bağlam temsilleri ile kelime girdilerinin etkileşime geçebilmesi için gerekli olan kuralları ortaya atmak çok daha zor görünmektedir. Önermeler mantığı bu işlemler için uygun değildir. Yüklemsel mantık kullanmaya kalktığımızda ise kural tabanlı sisteme yöneltilen eleştirilerle karşılaşmaktayız.

Sonuç olarak, çalışmamızda, bileşimsel yaklaşımın her iki versiyonu da reddedilerek anlam, temsiliyet gibi kavramların sorunun temelini olduğu kanısına varılmaktadır. Çözüm olarak anlam kavramının kullanılmadığı temsiliyetçi olmayan bir dil anlama modeli önerilmektedir. Bu yaklaşıma göre dilsel ifadeler ve dil kullanıcısının davranışları arasındaki bağlantı temsiller yerine doğrudan kelimelerin ve bağlamın kendisi üzerinden modellenmektedir. Bu sayede dinamik ve karmaşık bir yapı olan bağlamın temsili ve anlam sorunları ortadan kalkmaktadır.

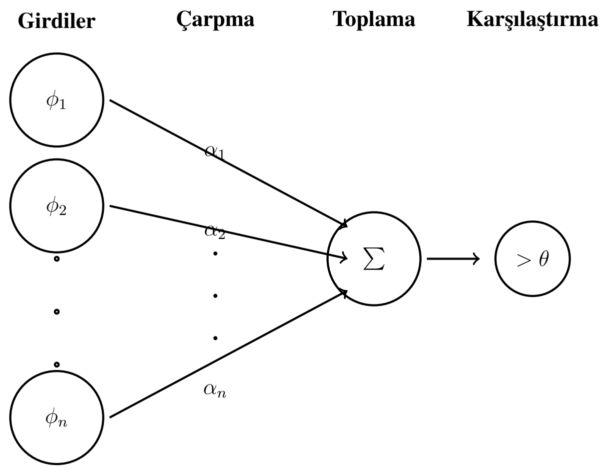
Anlam olmadan dil anlama sürecinin nasıl modelleneceği konusunda ise bağlantıcı görüşe başvurulmaktadır. Sembolik görüşte kurallar açık ve görece az sayıdadır.

Gramer kuralları buna bir örnek olarak gösterilebilir. Kelime anlamları, ve bağlamın temsili için kullanılan atomik parçalar sistemin az sayıdaki parametrelerine karşılık gelmektedir. Bağlantıcı görüş ise bilişsel süreçlerin modellenmesi konusunda, sembolik görüşe alternatif olarak, temsili olmayan çok daha fazla parametre ve örtük kurallar üzerinden çalışmaktadır. Örnek olarak, bir insanın hoşlandığı kitaplar sembolik görüşe göre kitapların özelliklerini kullanarak oluşturduğumuz bir ağaç üzerinden belirlenebilir. Burada olasılıksal bir yöntem de kullanılabilir. Diğer taraftan, bağlantıcı görüş makro düzeydeki parametreler yerine mikro düzeyde doğrudan temsil edemeyeceğimiz parametrelerle bu durumu modellemektedir. Her bir kitap çok yüksek boyutlu reel vektör ile temsil edilir. Sistem milyonlarca ağırlık kullanarak bu vektörleri işleme alır. Sonuç olarak çıkan değerler yorumlanır ve kitapların uygun olup olmadığı belirlenir.

Bağlantıcı ve sembolik görüşler aslında bütün bilişsel süreçlerde rekabet durumundadırlar. Ekonomik, dini, cinsel, politik davranışlarımızın arkasındaki mekanizmanın görece az sayıda, makro, anlamlı yapılardan mı oluştuğu; yoksa sayıca çok daha fazla olan mikro, anlamsız yapılardan mı oluştuğu tartışmanın ana noktasıdır. Bu tezde, bağlantısalcı görüşle uyumlu olarak dil anlama sürecinin anlamlı, makro yapılarla modellenemeyeceği iddia edilmektedir.

Bağlantıcı görüşün tarihine kısaca değinmek gerekirse, bu görüşün temelleri felsefi mantık alanına ilgileri olan Warren S. McCulloch ve Walter Pitts isimlerinde iki psikiyatristin insan zihnini sinirbilimsel süreçlere indirgeme çabalarıyla başlar (McCulloch and Pitts (1943)). Nöron iletimindeki *ya hep ya hiç* ilkesinin sembolik mantıkta bir önermenin sadece doğru ya da yanlış değerlerini alması ilkesi ile örtüşmesi nöronlar arası ilişkinin mantıksal bir şekilde modellenebileceği fikrini doğurmuştur. McCulloch ve Pitts modellerinin Turing makinalarıyla aynı güce sahip olduğunu iddia etse de, Stephen Cole Kleene bu modelin sonlu durum otomatlarıyla aynı güce sahip olduğunu göstermiştir.

McCulloch-Pitts modelinin bağlantıcılık ve sembolik görüşü ayırmak yerine her ikisine de temel oluşturduğunu söylemek mümkündür. Bağlantıcılığın sembolik yaklaşımdan ayrılması Frank Rosenblatt'ın McCulloch-Pitts modelini bir ileri noktaya taşıması ile mümkün olmuştur. Rosenblatt, bilişsel süreçlerimizi modellemek için doğru ve yanlış ya da sıfır ve bir gibi iki değerden oluşan sembolik mantık yerine, ara değerlerin de kullanıldığı olasılık kuramının temel alınması gerektiğini iddia etmiştir. Perceptron olarak bilinen model yapay sinir ağlarının ilk versiyonudur. Bu modelde, girdiler sistemin ağırlıklarıyla çarpılır. Elde edilen değer belli bir sabitle karşılaştırılır. Sonuca göre yapay nöronlar aktive olur (Rosenblatt (1958)). Bu aktivasyon, sistemin amacına göre yorumlanmaktadır. Mesela, bir otomobilin alınıp alınmamasına, fotoğraftaki insanın bir erkek olup olmamasına sistem bu yöntemle karar verebilir. Veri tabanına yeteri kadar örnek girdi ve etiket eklendiğinde, bu girdileri doğru etiketlerle eşleştirecek ağırlıklar aranır.



Bu yöntemle sembolik mantığın belirli fonksiyonları taklit edilebilmektedir. Bu da sistemin bütün mantıksal fonksiyonları taklit edebilecek güçte olduğunun düşünülmesine neden olmuştur. Ancak, sembolik görüşün önemli bir savunucusu olan Marvin Minsky bu modelin XOR bağlacını taklit etmekte yetersiz olduğunu göstermiştir (Minsky and Papert (1969), p. 62). Böyle bir fonksiyonu taklit edebilmek için perceptronun birden fazla tabakadan oluşması gerekmektedir. Bu durumda da doğru parametreleri sistematik olarak bulma sorunu ile karşılaşmaktayız.

Minsky'nin bu gösterimi uzun bir süre bağlantıcı görüşün sembolik görüş karşısında zayıflamasına neden olmuştur. Ancak daha sonraları, David E. Rumelhart geriye yayılım algoritmalarını tanıtarak bu sorunun üstesinden gelmiştir. Geriye yayılım algoritmasında ilk aşama olması gereken değer ile sistemin çıktı olarak verdiği değer üzerinden hatanın hesaplanmasıdır. Bu işlem için farklı hata fonksiyonları mevcuttur. Seçilen hata fonksiyonun her bir ağırlığa göre türevi kullanılarak küçük adımlarla yerel minimum noktasına ulaşmak amaçlanmaktadır (Rumelhart et al. (1986a)).

Doğal dil anlama açısından, bağlantıcılığın geliştirdiği bu yöntemler kullanılarak kelimelerin çok boyutlu uzayda temsilleri yapılmaktadır. Bu temsiller oldukça zengin bir içeriğe sahiptirler. Diğer taraftan özyinelemi yapay sinir ağları doğal dil anlama sırasında devamlı değişen bağlamın modellenmesi için oldukça kullanışlıdır. Özyinelemeli sinir ağlarında sistemin çıktısı tekrar girdi olarak sisteme geri dönmektedir. Bu ağlar bir dilden başka bir dile otomatik çeviri, bir sonraki kelimeyi tahmin etme, fotoğrafları cümleye çevirme gibi karmaşık işlemleri yerine getirebilmektedir.

Doğal dillere ait ifadelerin semantik ve pragmatik değerleri bağlantıcı yaklaşımın güncel çalışma konuları arasındadır. Bu çalışmanın konusu olan doğal dil anlama süreçlerinin anlam olmadan modellenmesi için gerekli olan araçlar yukarıda belirtildiği şekliyle bağlantıcı görüş tarafından sağlanmaktadır.

Bağlantıcı yöntemin doğal dil anlama süreçlerinin anlam olmadan modellenmesine en önemli katkısı sentaktik işlemlerin yorumsal işlemlerden ayrılmasıdır (Smolensky (1995a), s. 167). Sentaktik düzeyde işlemleri reel değerli vektör parçaları yapmaktadır. Her bir kelimeye karşılık gelen vektör işleme alındıktan sonra, sistemin durumu zihinsel uzayda bir noktadan başka bir noktaya taşınmaktadır. Mesela, sistem n boyutlu bir uzayda h_0 noktasında olsun. "Kar" kelimesi girdi olarak verildiğinde lineer olmayan işlemlerle sistem h_1 noktasına geçer. h_1 noktasındayken "beyazdır" kelimesi girdi olarak verildiğine ise sistem h_2 noktasına geçmektedir. Yorumlama düzeyinde ise bu noktalar analiz edilir. İstenilen bilgilerin elde edilmesi için bu noktalar daha düşük boyutlu uzaylara transfer edilmektedirler. Dolayısıyla, sentaktik işlemler ve yorumlama farklı düzeylerde yapılmaktadır.

Bu çalışmada sunulan yaklaşımın ve bileşimsellik ilkesine dayanan yaklaşımların amacı, insanların doğal dil anlama süreçlerini taklit edebilecek mekanik bir prosedür bulmaktır. Bileşimsellik ilkesine dayanan yaklaşımların savunucuları, dilsel bir ifadenin yüzey yapısından anlamına doğru bir homomorfizma olduğuna inanırlar. Eğer bir ifadenin sözdizimini anlamı ile eşleştiren yapısal bir benzerlik varsa, o ifadenin anlamını yüzeydeki yapıyı kullanarak hesaplayabiliriz. Bileşimsellik ilkesini merkeze alan iki ana düşünce okulu vardır. Bunlar kural tabanlı yaklaşım ve pragmatik odaklı yaklaşımlardır. Kural tabanlı versiyon için, yüzey yapısının bileşenlerini, anlamlarının bileşenlerine, sistematik bir şekilde eşleştirmemiz gerekmektedir. Kısaca, $h(F_{ß}(\alpha_1....\alpha_n)$'nin anlamıı aşağıdaki gibi hesaplanır:

$$h(F_{ß}(\alpha_1....\alpha_n)) = G_{ß}(h(\alpha_1)....h(\alpha_n)).$$

Burada F sentaktik kombinasyon fonksiyonu, h yorumlama fonksiyonu ve $G_{ß}$ anlamsal kombinasyon fonksiyonudur. Bu çalışmada, bu yöntemin sonuç vermeyeceği savunulmaktadır. Çok anlamlılık yüzeysel sözdizimi ile anlam arasındaki yapısal uyuşmazlığı göstermektedir. Bu ayrımı bir kez görmezden geldiğimizde, derin kavramsal sorunlarla karşı karşıya kalırız. Kural tabanlı yaklaşımla ilgili sorunlar, diğer bileşimsellik temelli yaklaşım olan pragmatik odaklı yaklaşım tarafından da eleştirilmektedir. Pragmatik odaklı versiyon ise bağlam temsilleri ve anlamlar üzerinden işlemektedir. Karmaşık ifadelerin anlamları, kelime anlamları ve bağlamsal bileşenler üzerine kuruludur. Çıkarımsal süreçler, çağrışımlar bu süreçte önemli bir rol oynamaktadırlar. $F_{ß}(\alpha_1....\alpha_n)$ ifadesinin anlamının hesaplanması aşağıda basitleştirilmiştir:

$$h(F_{ß}(\alpha_1....\alpha_n)) = G_{ß}(p_i(h(\alpha_1))....p_j(h(\alpha_n))).$$

Burada p_i ve p_j pragmatik fonksiyonlardır. Bu çalışmada, pragmatik temelli yöntem de reddedilmiştir. Bu yaklaşımların her ikisi de anlamsal bileşenler aracılığıyla anlamsal bir durum inşa etmek üzerine kuruludurlar. Bu çalışmada, kelimelerin hiçbir anlam taşımadığı, yani doğal dil anlama süreçlerinin kelime anlamı üzerine kurulu olmadığını iddia eden Anlamsal Eleyicilicilik kuramı savunulmaktadır. Bu kuramı uygulanabilir bir hale getirebilmek için klasik, sembolik mimariler yerine bağlantıcı mimarilerin kullanılması önerilmektedir.

Bileşimsel yaklaşımlarla ilgili karşılaştığımız zorluklar, bir dizi bileşenle anlamsal bir durum oluşturmayı amaçlamalarıdır. Semantik kombinasyon kuralları, bu bileşenlerin nasıl yerleştirileceğine rehberlik etmeyi amaçlamaktadır. Pragmatik fonksiyonlar bileşenleri modüle eder. Bu yaklaşımların oldukça basit ve sade olmaları bilimsel açıdan bir erdemdir. Ancak, dil kullanıcılarının dilsel ifadeleri anlama biçimleri bu araçlarla modellenemeyecek kadar karmaşıktır. Öte yandan, bu çalışmada anlamsal bir durum inşa etmenin bir yolunu önerilmemektedir. Bunun yerine kelimeler ve bağlamın tek görevinin dil kullanıcılarını bir zihinsel durumdan diğerine taşıdığı iddia edilmektedir. Sembol-altı araçlar, bu çerçevede daha ileri gitmemiz için bize yol açmaktadırlar. Sunulan yüksek boyutlu semantik uzay, doğal dil anlama sürecinin zengin yapısını taklit etmenin tek aracı olarak görülmektedir. Kelimelere reel vektörler atanmasının nedeni de budur.

C. THESIS PERMISSION FORM / TEZ İZİN FORMU

(Please fill out this form on computer. Double click on the boxes to fill them)

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences ☐

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences ☒

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics ☐

Enformatik Enstitüsü / Graduate School of Informatics ☐

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences ☐

YAZARIN / AUTHOR

Soyadı / Surname : Toy
Adı / Name : Tolgahan
Bölümü / Department : Felsefe / Philosophy

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English): A DEFENSE OF MEANING ELIMINATIVISM: A CONNECTIONIST APPROACH

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☐ **Doktora** / PhD ☒

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☒
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two years. *** ☐
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. ***** ☐

***** *Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir. / A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.*

Yazarın imzası / Signature .... **Tarih** / Date ....
(Kütüphaneye teslim ettiğiniz tarih. Elle doldurulacaktır.)
(Library submission date. Please fill out by hand.)

Tezin son sayfasıdır. / This is the last page of the thesis/dissertation.