

Phenomenological References: Arguments for Mentalistic Natural Language Semantics

Jong Sup Jun*†

Seoul National University

Jong Sup Jun. 2004. Phenomenological References: Arguments for Mentalistic Natural Language Semantics. *Language and Information* 8.2, 113–130. In a prevailing view of meaning and reference (cf. Frege 1892), words pick out entities in the physical world by virtue of meaning. Linguists and philosophers have argued whether the meaning of a word is *inside* or *outside* language users' mind; but, in general, they have taken it for granted that words refer to entities in the physical world. Hilary Putnam (1975), based on his famous twin-earth thought experiment, argued that the meaning of a word could not be inside language users' head. In this paper, I point out that Putnam's argument makes sense only if words refer to entities in the physical world. That is, Putnam did not provide any argument against mentalistic semantics, since he erroneously assumed that *meaning*, but not *reference*, was inside our mind in mentalistic semantics. Mentalistic semanticists, however, assume that words pick out their references inside our head (instead of a possible outside world). A number of arguments for the mentalistic position come from psychology: studies on emotion and visual perception provide numerous cases where words cannot pick out entities from the physical world, but inside our head. The mentalistic theory has desirable consequences for the philosophy of language in that some classical puzzles of language (e.g. Russell's (1919) well-known puzzle of *excluded middle*) are explained well in the proposed theory. (Seoul National University)

Key words: meaning, reference, generative lexicon, conceptual semantics

* College English Program, Seoul National University, San 56-1, Sillim-dong, Gwanak-gu, Seoul 151-742, Korea. E-mail: jongsupjun@korea.com

† I am grateful to Ray Jackendoff, James Pustejovsky, Alan Berger, Maurice Hershenson, and four anonymous reviewers for discussion and comments on the idea developed in this paper.

1. Introduction: Basic Issues and Problems

Since Frege (1892) first made the well-known distinction between meaning/sense¹ and reference, there have been linguistic and philosophical proposals for what constitutes meaning or what we mean by *meaning*. Among variety of schools and thoughts, two competing approaches have always been in the arena, debating upon many interesting questions relevant to linguistics, philosophy, and psychology. One of the two approaches is devoted to a thesis called *psychologism/mentalism* for language and meaning (cf. Jackendoff 1983, 1990). In this approach, the meaning of an expression is simply in the head: a competent language user has a full access to the correct meaning of an expression, and determines the reference based on the meaning in the head.

On the other side of the arena lies a view against mentalism. Putnam (1975, 1981) has been adamant in his position against the mentalistic perspective of language and meaning. Based on thought experiments like the meaning of *water* in the twin-earth, he makes a strong claim that meanings are not in our heads; in fact, there is no such thing as *meaning* that a language user has a full access to in order to determine the reference of a word. I call this view *externalism*, an unsatisfactory blanket term for all anti-mentalistic views of language, to contrast it with *psychologism/mentalism*.

Under the two competing hypotheses lies the Fregean principle that meaning determines reference. Although the Fregean assumption has been challenged by some local linguistic examples like Kripke's (1972) treatment of proper names as rigid designators and Kaplan's (1989) theory of directly referential terms, the challenges and attacks have always been local, and have never reached the global treatment of *all/most* possible linguistic types and categories. That is, Kripke's treatment was confined to proper names and some natural kind terms; and Kaplan's major concern was demonstratives. Except for these few cases, the Fregean principle has been quite stable in many philosophical theories of language.

In Putnam's (1975, 1981) famous twin-earth thought experiment, the word *water* refers to H₂O on the Earth and XYZ on the twin-earth. In other words, the chemical composition of the earthian water is different from that of the twin-earthian water in extension. The thrust of this argument is that if meanings were in our heads, the mentalistic meaning of *water* would fail to pick out proper references – i.e. either the earthian H₂O or the twin-earthian XYZ – in the world, and that meanings, therefore, cannot be in our heads. In this thought experiment, Putnam tries to show cases in which mentalistic meanings fail to pick out (or determine) proper references, and then to use these cases as counterexamples against the mentalistic perspective. Thus, the Fregean principle that meaning determines reference is still at work in Putnam's arguments.²

¹ Despite the confusion in translating the German word *Bedeutung* also into *meaning* in English, I follow the convention of translating Frege's *Sinn und Bedeutung* into *Meaning/Sense and Reference*.

² Later in his thesis, Putnam (1975, 1981) develops a socio-linguistic idea of how correct references are determined. He believes that references are determined through the social linguistic division of labor, which obviously differs from the Fregean concept of meaning. In this sense,

It is, therefore, reasonable to ask “What on earth are references?” In Russell’s (1919) argument, definite descriptions are the meanings of proper names, and they determine the references of proper names. Kripke (1972) argues that proper names do not have meanings, and that their references are *fixed* by definite descriptions which are neither necessary nor *a priori*. Putnam argues that the proper reference of *water* cannot be determined by the meaning in our head, and that meaning, thus, cannot be transparent to our observation. A fundamental assumption of these theories is that words pick out entities in the external world, and that these entities are the references of the words. For instance, the reference of a word *chair* is the actual chair as a physical entity in the outside world; i.e. the word *chair* refers to ‘a chair’ in front of us. Formal semanticists in the tradition of Montague (1973) and Carnap (1956) take the reference of *chair* as a set of all chairs in a possible external world. In general, reference is fixed as an outside entity, and it is the concept of meaning that varies with respect to each theory. In Putnam’s argument, mentalistic semantics faces a dilemma in picking out a correct reference, since no mentalistic meaning enables us to pick out either H₂O or XYZ correctly from the actual physical world. This shows that Putnam’s argument is valid only under the assumptions that meaning determines reference, and that reference is something in the physical world.³

It is not clear, however, that the reference of a linguistic expression – or roughly a word – should be a real entity in the physical outside world. Jackendoff (1983, 1987, 1990, 1997, 2002) has constantly opposed this view, maintaining that words “refer to speakers’ construals of the external world (Jackendoff 1987, 128)”. In Jackendoff’s mentalistic semantics or what he calls *conceptual semantics*, words like *chair* do not refer to an actual chair in the world, but the image of a chair in our conscious mind. In other words, the linguistic expression *chair* refers to some psychological state of our mind, i.e. our phenomenological experience of a chair.⁴

Putnam’s thesis seems to be against the Fregean principle that meaning determines reference. On the contrary, Putnam’s socio-linguistic hypothesis has not undermined the Fregean principle. What he proposes in the socio-linguistic hypothesis is to replace the Fregean concept of meaning with the socio-linguistic concept of meaning. Fregean meaning is relatively transparent to each language user, whereas Putnam’s socio-linguistic meaning as *reference fixer* is opaque to each language user. This becomes clear in Putnam’s John Locke lectures in 1978: he claims that “we are partially opaque to ourselves” and that “certain human abilities – language speaking is the paradigm example – may not be theoretically explicable in isolation”. Therefore, the Fregean principle still lies unharmed under Putnam’s thesis: *something* determines reference. That *something* is the opaque meanings of expressions, which are unfortunately inexplicable to us in isolation.

³ An anonymous reviewer correctly points out that Putnam’s claim was merely about *meaning*, namely that there is no such thing as meaning in our head, without making any explicit assumption about *reference*. Even though Putnam did not elaborate his argument with respect to reference, it is impossible to construe his theory without assuming physical references in the outside world: he may have taken it for granted that meaning picks out entities in the physical world following the old tradition of philosophy. Notice that it is the chemical composition that matters to determine the reference of the word *water*. According to Putnam, both the earthian water and the twin-earthian water give us the same psychological experience; but the references are not the same, since one is H₂O and the other is XYZ in chemical composition. Putnam’s reference for *water* is surely the physical entity, namely either H₂O or XYZ: there is no room to believe otherwise.

⁴ The term *phenomenological* is used in the standard sense of the word in philosophy; i.e. phe-

Putnam's argument against meaning in the head based on the twin-earth *water* is not valid at all in Jackendoff's mentalistic view of meaning and reference. Words like *water* refer to our psychological (or phenomenological) experience of actual water. Since Putnam already made an explicit assumption that the psychological experience of actual water is exactly the same for both earthians and twin-earthians, it does not matter at all whether the actual water is H₂O or XYZ. Since the meaning for *water* in our heads pick out the phenomenological experience of water in our head, we can still maintain that the meaning of water successfully picks out the correct reference for *water* for both earthians and twin-earthians.⁵ This idea is, however, not free from counter-arguments and objections: numerous objections to the mentalistic semantics may arise. For example, we do not say that a lump of fake gold is the same as a lump of genuine gold just because both of them give us exactly the same psychological experience. I will return to this issue and related problems in section 4.

The goal of this paper is to provide arguments for Jackendoff's concept of reference as *speakers' construal of the physical outside world* arguing against Putnam's twin-earth experiment. When we face a physical entity in the outside world, we go through psychological experiences. Some of these experiences have direct correspondences with outside entities in the world, and others do not. For instance, when we see an apple, we go through such psychological experiences as visual perception of that apple and (also possibly) the recollection of the theory of gravitation aroused by the famous episode connecting Sir Isaac Newton and the falling apple. Only the former mental experience has a direct correspondence to the outside entity called *apple*. The latter kind of association is aroused by consulting one's encyclopedic knowledge with respect to the former visual experience. In this sense, the former experience is primary, and the latter experience is secondary.

As long as we are not committed to the Quinean (1951) skepticism with our own sense data, we are pretty certain that the primary experiences are stable for most people, and automatic in the sense that they *must* occur in the face of the

phenomenological experience refers to our conscious psychological state of the mind. In this sense, it is analogous to the concept of *qualia*, the subjective feeling.

⁵ An anonymous reviewer wonders if my current proposal would be an appropriate objection to Putnam's thesis, in that Putnam's argument is restricted to the discussion of natural kind terms whereas my argument is based on general terms and words in general. In fact, this was the same kind of question I had in mind about the debate of *meaning* in the field: *How come could we generalize Putnam's theory of meaning based on such restricted data to the general discussion of meaning?* Putnam does not have much to talk about except for the discussion of natural kind terms when it comes to the debate of *meaning*. Nevertheless, his main claim always takes the strongest position about semantics: he does not say "The meaning of natural kind terms is not in the head", but "Meaning is not in the head". For one thing, I want to point out that it is Putnam that wanted to generalize the discussion of natural kind terms to the discussion of general semantics. For another, my proposal about natural language semantics treats natural kind terms as well as general terms (plus words of various categories in general) consistently. It is clear that mentalistic semantics treats all words in the head in terms of their meaning and reference; hence, my objection to Putnam does not lose its ground by the fact that his reasoning was based on natural kind terms. The reviewer raised a similar question about Kripke's discussion of proper nouns. I will not pursue this issue here, since this paper does not mean to be an attack on Kripke; but I believe some directions to pursue would be apparent to interested readers.

corresponding outside entities. That is, we cannot avoid the primary visual experience of an apple when we see the apple before our eyes. I call these primary automatic psychological experiences phenomenological references of linguistic expressions following the tradition in philosophy.

In the current version of mentalistic semantics, however, words refer to phenomenological references, and thus words as linguistic signs have arbitrary relationships with psychological experiences.⁶ In Putnam's externalism and in most philosophical and formal semantic traditions, however, words refer to entities in the physical outside world. Words as linguistic signs have arbitrary relationships with the physical outside world.

In Putnam's criticism against mentalism, only the meaning of an expression is taken to be mentalistic, and the reference of an expression still remains in the world. In mentalistic semantics, however, both the meaning and the reference of a word are inside the head. To support this idea, I provide a number of arguments from psychology against externalism and for the mentalistic view of language in section 2. After that, I introduce Pustejovsky's (1995) generative lexicon (=GL) theory and Jackendoff's (1983, 1990, 2002) conceptual semantics as plausible models of mentalistic semantics in section 3. Finally, sections 4 and 5 discuss interesting consequences of the theory with respect to classical puzzles: Putnam's (1975) puzzle about *gold* in section 4, and Frege/Russell's puzzles about meaning and reference in section 5.

2. Evidence for Phenomenological References

2.1 Emotion: Evidence from social psychology

In the theory of phenomenological references, linguistic expressions refer to psychological experiences. Words like *chair*, *desk*, and *book* refer to visual experiences of a chair, a desk, and a book in our mind; i.e. the meanings of *chair*, *desk*, and *book* pick out the visual images of a chair, a desk, and a book in our head as references. The word *run* refers to visual experiences in our mind, where an object is moving from one position to another in a reasonably fast manner; *walk* refers to visual experiences in our mind, where an object is moving from one position to another using its feet in a reasonably slow manner; and words like *sad* refer to some psychological state of our mind.

Study on emotion in social psychology provides strong empirical support for the theory of phenomenological references. How many different types of emotion do we have/feel? How do we express our emotion physically? Phenomenologically, we have hundreds of emotions: feelings of HAPPINESS, ANGER, SADNESS, FEAR, LONELINESS, GLOOMINESS, DREARINESS, SYMPATHY, FERVOR, PASSION, EXCITEMENT, etc. Physically or behaviorally, however, we only have several ways to express the hundreds of phenomenological feelings. Ekman's (1972) classical study on emotional expressions of human face shows that there are six reliable – or universal – facial expressions of emotion. Physiological studies yield

⁶ I follow Saussure's (1958) thesis of *arbitrary relationships* between linguistic signs and their correspondences (i.e. references whether physical or phenomenological).

similar results: brain studies show that there are at best two physiological expressions in the brain level for hundreds of phenomenological feelings. Studies with humans and other animals show that the amygdala, a relay station between diverse areas in the brain and the thalamus, is fundamentally related with emotion, and actually the physiological expression of the amygdala is either a positive or a negative response in neural activities (Mullan & Penfield 1958; Devinsky & Bear 1984; Davis 1992; LeDoux 1994; Bechara et al. 1995).

Fortunately, however, we have hundreds of linguistic expressions for hundreds of emotional feelings. If linguistic expressions referred to physical entities, i.e. behavioral or physiological expressions, then it would be almost impossible to see how hundreds of linguistic expressions refer to six facial expressions or two physiological expressions in the brain. On the other hand, if linguistic expressions referred to phenomenological experiences, we could easily understand how hundreds of linguistic expressions refer to hundreds of phenomenological experiences.

An influential theory of emotion in social psychology was proposed by James-Lange (James 1890), and supported by numerous subsequent researches (Cannon 1927; Landis & Hunt 1932; Ax 1953; Schwartz, Weinberger, & Singer 1981; Mescuita & Frijda 1992; Sinha & Parsons 1996). An elaboration and modification of the James-Lange theory is found in Schachter & Singer (1962). These studies show how we reach a certain emotional status. Situational factors arouse a very limited number of physical responses. We get certain phenomenological feelings through the interpretation of the situation, and the attribution of the physiological responses to complex psychological statuses. That is, our physiological/physical response when we jog for fun is exactly the same as our physiological/physical response when we are chased by a fierce dog. It is our interpretation of the situation and the attribution of the physical response that distinguish between EXCITEMENT and FEAR. In short, there are no physical correspondents that linguistic expressions like *excited* and *fearful* can refer to; rather, emotional expressions in language must refer to psychological properties of the mind. Hence, we have independent support for the theory of phenomenological references from social psychology.

Emotion words are particularly difficult for the theory of physical references. As we have already learned from the literature of social psychology, there are at most two or six physical states or expressions that correspond to hundreds of psychological states of mind. It is almost impossible to explain how hundreds of linguistic expressions, i.e. words, refer to only two or six physical states in a possible outside world.

In the classical view of physical references, words refer to physical entities in a possible world. The theory sounds fairly reasonable for concrete objects in a physical world like a chair, a desk, and a book; e.g. the word *chair* refers to a chair in front of you. When it comes to references of verbs and adjectives, the theory sounds a bit strange: present-day formal semanticists take the set of all running objects in a world for the extension of an English verb *run*, the set of all ordered pairs in relation of love for the extension of a verb *love*, and the set of all sad entities for the extension of an adjective *sad*.

Emotion words, therefore, pick out sets of entities that go through such emo-

tional states. What is not clear is how words like *sad* refer to sad people, i.e. the set of all sad entities in a possible world, rather than to the sad feeling. The theory of physical references could not find a way to directly refer to psychological states; so it has devised a way to indirectly refer to entities that go through such psychological states. An undesirable consequence is that the word *sad* has the same reference with the phrase *sad people* or *sad entities*; likewise, the word *run* has the same reference with another word *runners*. My point is that emotion adjectives do not refer to entities that go through such emotions; they refer to emotional or psychological states of the mind.⁷

Study of language is an empirical science. It does not belong to an *a priori* domain of formal logic; rather, it studies something that is spoken here and now, and something that is constantly changing. We cannot talk about a natural language without thinking about the empirical coverage of the theory. The empirical coverage of the theory of physical references is restricted to concrete objects. The set-theoretic expansion of the idea into verbs, adjectives, etc. does not adequately explain speakers' semantic intuition of the words.

2.2 Veridicality

Confronted with a formidable question "Is there anything indubitable and certain for its own sake?", some philosophers have tried to reduce *what is indubitable and certain* to our own sense data. Descartes's *cogito ergo sum* was a finding of *indubitable self* based on the indubitable fact that he himself is perceiving/sensing that he is thinking. Russell (1919) defined *logically proper names* as what was certain by our own sense data. Surprisingly, Kripke (1972) did not question Russell's belief in the certainty of our own sense data in his discussion of proper names.

On the other hand, if we look into the vast literature in psychology, we get instantly shocked to see that our sense data do not provide any certainty about the physical outside world. There is dissociation between the physical outside world and the sense data. What is more striking is that our psychological experiences can be something totally unexpected by the sense data.⁸ In this subsection, I present the dissociation between the sense data and our psychological experiences as independent support for the theory of phenomenological references.

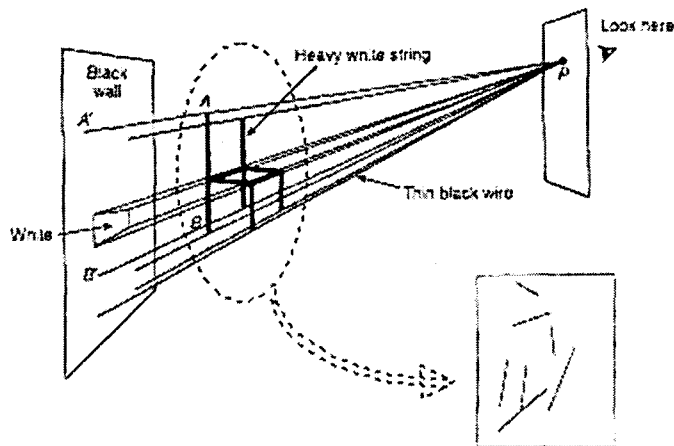
I begin my discussion by presenting two different predictions made by the theory of physical references and the theory of phenomenological references. There is something in the physical outside world. Someone is standing before the physical entity, and says "Oh, it is a chair!". According to the theory of physical references (i.e. externalism), the physical entity in the outside world *must* be a chair, since the word *chair* in "Oh, it is a chair!" picks out a chair from the physical outside world. The theory of phenomenological references, however, makes a different prediction: the physical entity before the person does not have to be a chair. What

⁷ One anonymous reviewer points out that recent developments in formal semantics have introduced events and states as the denotations of verbs and adjectives; e.g. *run* denotes a set of running events, and *sad* denotes a set of sad states. One conceptual problem with this view is that the judgment for sad states, for instance, is made with reference to the sad feeling in our mind. It is not clear how a psychological state like the sad feeling can be treated without reference to our mental state except for a totally formal treatment of *state* as a semantic type.

⁸ In this sense, psychological experiences are *emergent* properties of the mind.

the word *chair* in “Oh, it is a chair!” picks out is a mental image of a chair, i.e. a phenomenological experience of a chair. The role of the physical entity in the outside world is to arouse that mental experience, and thus the entity is not necessarily a real chair.

The latter, but not the former, prediction is confirmed by Ames’s (1955) famous chair-demonstration. In the demonstration (Figure 1), what is *really* out there in the physical world is just a collection of jumbled strings in the 3D space. Those strings are not meant to make any comprehensible shape, nor are they meaningful in their configurations. Nevertheless, what you see right from a certain visual angle is exactly the shape of a chair. Then, subjects all say “Wow! It’s a chair!”. One question arises: What does the word *chair* refer to when subjects say “Wow! It’s a chair!”? It is crystal-clear that the *chair* in “Wow! It’s a chair!” does not refer to any real chair in the physical outside world. Rather, the *chair* in “Wow! It’s a chair!” in Ames’s demonstration refers to the phenomenological experience of the collection of jumbled strings in the physical world, which happens to be equivalent to the phenomenological experience of a real chair.⁹



[Figure 1] Chair demonstration (from Hershenon 1999, 84)

In fact, the vision literature in perception psychology provides ample evidence for the mismatch between the physical outside world and our own sense data, and between the sense data and our psychological experiences. To borrow the techni-

⁹ One anonymous reviewer wonders if the illusive perception could really support the theory of phenomenological references, since once the subject realizes the nature of illusion, (s)he would *correct* the mental experience. For instance, once the subject learned the nature of illusion in Ames’s chair-demonstration, (s)he would realize that it is not a chair. This prediction is, however, not correct. Strikingly enough, explicit learning of the nature of illusion does not change our mental experience at all. After we explicitly learn that Ames’s chair is not a chair, it still looks like a chair invariably; after we explicitly learn that Ames’s trapezoidal window is not a rectangular window, it still looks like a rectangular window invariably. Hence, the illusive perception provides powerful evidence for the theory of phenomenological references.

cal jargon from the field, there are mismatches between *distal stimuli* and *proximal stimuli*, and between the *proximal stimuli* and our perception. Transactionalists' demonstrations show these mismatches well. Besides the chair demonstration, Ames (1955) shows that a trapezoidal window in the physical outside world is not distinguishable from a rectangular window in its retinal image, and that strikingly enough the ambiguous retinal images are invariably perceived as a rectangular window. Ittelson's (1968) demonstration with a distorted room, too, shows that a distorted room is perceived as a rectangular room (cf. Ames 1955). Sometimes, natural settings provide exactly the same effect. In Cecwu Island, south of the Korean peninsula, there is a hill called magic hill or magic road – there is a paved road these days. Visually, we have no doubt that it is a hill, and that the road is an up-slope. But the fact is that the hill is not an up-slope, but a down-slope. It is a visual illusion due to surrounding environments that makes a down-slope look like an up-slope. Hence, when we stop our vehicle at the (visual) foot of the hill, and put the transmission into the neutral position, the vehicle starts to go upward! But, what is really happening is that the vehicle is going downward. What is apparent from the nature's demonstration is simple: we cannot and should not trust our own sense data and our perception. Our sense data are in many cases not veridical.

I will mention a couple more examples from the thousands of experiments and demonstrations in the psychology literature that report the non-veridical nature of our perception. Ittelson (1960) shows that there is a family of equivalent configurations for one perceptual outcome. That is, countless figures of countless shapes and sizes in the physical outside world produce exactly the same perceptual outcome. If the word *rectangle* referred to a rectangle in the physical outside world, and if *trapezoid* referred to a trapezoid in the physical outside world, then we would have no reliable basis to pick out a rectangle for the word *rectangle*, and to pick out a trapezoid for the word *trapezoid*, since our perception of a rectangle and a trapezoid in reality comes from members of a family of equivalent configurations.

Nevertheless, surprisingly enough, we pick out a rectangle for the word *rectangle*, and a trapezoid for the word *trapezoid* in most situations correctly and reliably. Transactionalists like Ames and Ittelson provide many interesting pieces of explanation for how humans overcome this apparent ambiguity due to a family of equivalent configurations. The theoretical details are not relevant for the purpose of this paper. What is important is that words cannot and do not refer to physical entities; rather, words *must* refer to phenomenological references.

The moon illusion – the moon looks bigger and brighter on the horizon than on the *zenith* – is probably the oldest observation which shows the mismatch between the physical outside world and our perception (Taylor & Boring 1942; Kaufman & Rock 1962; Solihkah & Orbach 1969; Hershenson 1982, 1989; Gogel & Mertz 1989). Gestalt psychologists' organization principles also provide ample evidence that our perception is not veridical to the actual physical outside world (Wertheimer 1912, 1923; Köhler 1929; Koffka 1935). For instance, when two light bulbs are placed in front of a subject in the dark room, and they are flashed asynchronously (i.e. at different time points), what (s)he see is *apparent movement* of the light. It is not simply an illusion of movement from one position to the other. Rather,

subjects actually see a gradual movement of light from one position to the other. Wertheimer's (1912) analysis of apparent movement shows that human perception is subject to some gestalt organization principles. Again, what is important is the non-veridical perception: our sense data do not provide us with true images of the physical outside world.

These studies consistently show that we cannot access directly and veridically to the physical outside world. Nevertheless, linguistic expressions in a natural language are reliably used by competent speakers. Since our sense data cannot and do not provide veridical information about the physical outside world, linguistic expressions cannot reliably pick out entities from the physical outside world. On the other hand, if words refer to phenomenological experiences in our head, we will be able to pick out correct references for words most of the time. The mismatch between the physical outside world and our perception is handled and overcome by other cognitive principles like transactionalists' encyclopedic knowledge base and gestalt psychologists' organization principles. I have made my point sufficiently clear: words do not refer to things in the world; words refer to phenomenological experiences in the head.

3. Toward Mentalistic Semantic Theories

Putnam's argument against mentalism/psychologism assumes a strange version of mentalism, where meanings are in the head and references are in the physical outside world. On the other hand, in the mentalistic view of semantics advocated in this paper, both meanings and references are in the head. Before I discuss subsequent problems and puzzles of the latter view, I will, in this section, sketch two plausible models of mentalistic semantic theories: Pustejovsky's (1995, 1998, 2000) generative lexicon theory and Jackendoff's (1983, 1990, 1997, 2002) conceptual semantics.

3.1 Pustejovsky

Inspired by Moravcsik's (1975, 1981, 1990) interpretation of Aristotle's modes of expressions (*aitiae*), Pustejovsky (1995, 1998, 2000) has developed the idea of the generative lexicon (=GL), an influential theory in lexical semantics. Aside from all the technical details of the GL machinery, the theory tries to incorporate a competent speaker's general knowledge of the world into the meaning of a word. For instance, the meaning of water is not just H₂O, as Putnam believes, but a collection of our detailed knowledge about water. That is, such pieces of world knowledge as "how water came into being," "the use/purpose/proper function of water," etc. collectively define water in the mental lexicon.

According to Moravcsik (1975, 1981, 1990), there are four modes of expressions (*aitiae*) in the Aristotelian concept of meaning; namely, Agentive (factors involved in bringing it about), Formal (what distinguishes it within a larger type system), Telic (the purpose and proper function), and Constitutive (the relation between an object and its parts). Pustejovsky takes a view that the meaning of a word can be decomposed into these four Aristotelian modes of expressions, the qualia structure to use his term. In the GL theory, every word or a lexical item has the qualia

structure in the decomposition of meaning, and the qualia structure incorporates a competent speaker's knowledge about the world.

One great advantage of the theory is its empirical coverage. Pustejovsky, using a minimum number of linguistic machinery like co-composition, coercion, and selective binding, succeeds in showing how a number of problems and puzzles in natural language semantics are solved by the GL theory. For instance, when we say *John began a book*, the *book* is never construed as a physical object called *book*; instead, the sentence means either 'John began to read a book' or 'John began to write a book'. This problem, according to Pustejovsky, is expected if we decompose the meaning of *book* into the qualia structure. That is, the Telic quale of *book* is READ(e,x,y) and the Agentive quale of *book* is WRITE(e,x,y). The interaction between the event-taking predicate *begin* and its complement *book* is defined by GL machinery called type coercion and co-composition, and then the appropriate meaning of *begin a book* is easily predicted.

I do not want to go any further to introduce the GL theory in this paper. I just want to emphasize that the GL machinery has achieved a broad range of empirical coverage in explaining natural language semantics, and that the GL theory is a plausible theory of mentalistic semantics in which meanings and references are in the head.¹⁰

3.2 Jackendoff

Jackendoff (1983, 1990, 1997, 2002) has explicitly maintained mentalistic view of semantics for a long time. He shares the Chomskyan (1959, 1965) belief that language is a psychological and biological object, and explicitly claims that both meanings and references are in the head. He assumes, following Chomsky (1959, 1965), that humans are born with an innate language faculty, which provides sufficient machinery for acquiring syntax, semantics, and phonology of a natural language in the linguistic community.

The semantic module of the language faculty provides us with some handful of conceptual primitives and categories, with which we organize the world into meaningful linguistic units. Primitive categories include THING, EVENT, STATE, ACTION, etc., and these categories are organized by primitive functions like GO, BE, CAUSE, etc. For instance, a linguistic expression *run* is decomposed into a conceptual structure of some THING's GOing through a certain PATH with some SPEED/Spatial Representation. In other words, humans conceptualize the running activity with the innately given primitives like THING, GO, PATH, etc. into a conceptual structure like [_{Event} GO(Thing, Path); Spatial Representation], and lexicalize this conceptualization into a linguistic expression *run*.

The meaning of *run*, therefore, is the conceptual structure in our head, and it is not the set of all running objects in the world as formal semanticists believe. One great benefit of conceptual semantics is its explanatory power of how humans acquire language and esp. meaning, and how humans use concepts in a creative manner. If meanings were external to humans as is advocated in externalism, there

¹⁰ Pustejovsky does not make an explicit assumption that both meanings and references are in the head. But his mentalistic position is obvious in his writings (also p.c.), and I find no conflicts between GL assumptions and my proposal about mentalistic semantics.

would be little room for the creative manipulation of meaning, an essential property of human language. Again, I do not want to go into details of Jackendoff's conceptual semantics, but I want to emphasize that in conceptual semantics, the meaning of a word is a mental object organized by primitive categories and functions given to us from our birth, and that conceptual semantics is also a plausible theory of mentalistic semantics in which both meanings and references are in the head.

4. Putnam's Puzzle about *Gold*: Under-Determination of References

I have already shown that Putnam's twin-earth thought experiment does not provide a strong argument against mentalistic semantics once we adopt the assumption that both meanings and references are in the head. The mentalistic stance, however, makes an undesirable prediction about the fake *gold* and the genuine *gold* discussed in length by Putnam (1975): since both the fake gold and the genuine gold give us the same phenomenological experience, both a lump of fake gold and a lump of genuine gold lying before us in the physical world are exactly the same gold. In this section, I will show that the puzzle about fake vs. genuine gold does not make a counter-example at all once we couch our mentalism in the GL theory or conceptual semantics.

According to Putnam (1975), a linguistic expression *gold* means *Au*, as water means H_2O to Earthians. On the other hand, in the GL theory, the meaning of *gold* is decomposed into the qualia structure according to a competent speaker's ontological knowledge of gold. In this sense, the meaning of a word is in the epistemological domain in mentalistic semantics. The qualia structure of *gold* may reasonably include the Agentive quale formally specifying that gold is an ore obtained by mining, the Telic quale formally specifying that gold has monetary value in society that can be used for exchange with other valuable goods, etc. The fact that the atomic composition of gold is *Au* is actually the last thing that can be considered as part of the meaning of *gold*.¹¹ Linguistic knowledge is acquired and used by lay people prior to any formal education. A competent English speaker acquires the word *gold* and uses it before (s)he gets into chemistry class at school. Hence, what an ordinary English speaker knows about *gold* does not in general include gold's atomic composition.¹²

Each person's knowledge about gold may differ, and thus the meaning of *gold* can vary slightly from person to person in the GL theory. Most philosophers find this consequence annoying. They want to take the meaning of a word as an absolute intensional function that never varies. But the assumption of the absolute

¹¹ One anonymous reviewer questions how the qualia structure incorporates the atomic composition of gold as part of the meaning of *gold*, and how we could account for the distinction between fake gold and genuine gold. The first question is easy: the atomic composition of gold is expressed in the constitutive quale. Next, Putnam already provided an answer to the second question: lay people cannot distinguish between fake gold and genuine gold; only experts can distinguish between these two; and therefore, the atomic composition of gold does not appear in the meaning of *gold* in lay people's mind.

¹² This is why Putnam believes that only experts, but not lay people, have access to the meaning of a word.

meaning of a word is a naive expectation that does not take into consideration the reality that language is constantly changing.¹³ Philosophers who maintain strong externalism tend to see meaning as an invariable fixed monad. Meaning is not decomposable, and it is an impenetrable whole that never changes. This idea is explicitly maintained in Fodor (1983), and harshly criticized by Jackendoff (1990, 2000). Following Jackendoff (1990, 2000), the view that the meaning of a word is an impenetrable monad is problematic in all respects: it cannot account for the creative use of meaning; it cannot account for the first language acquisition; and it is odd to think that there is an impenetrable meaning as a monad for words like *telephone*, *computer*, etc. Meaning is, like every aspect of all natural human languages, constantly changing, and a basis for the change of meaning is the fact that meaning of a word varies from person to person.

In the theory of phenomenological references, the psychological reference of a word is determined by the psychological meaning of the word. Let us consider the case of *gold*. We have a lump of genuine gold before us. Also, we have a lump of fake gold before us. Let us further assume, following Putnam, that these two metallic objects before us give us exactly the same psychological experience. According to the standard assumptions of conceptual semantics, the meaning of *gold* includes featural specifications for its spatial representation. Since the two objects in front of us are satisfying the spatial representation (about color, hue, hardness, etc.) of gold, both objects pass at least one set of criteria for *gold*. According to the GL theory, however, the Telic quale of *gold* must specify its proper function and use in the society. Unfortunately, however, we have no adequate information to determine whether or not the two objects before us satisfy this part of meaning. Hence, when we say *gold* for the two objects before us, we are actually not *definitely* saying that they *must* be gold, but *tentatively* saying that they must be gold if they satisfy other criteria that we cannot test at the moment.

One important consequence follows: in mentalistic semantics, meaning in the head determines reference in the head, but a tentative determination is also possible when we lack of appropriate information for determination. I call this tentative determination *under-determination* that is prevalent in natural language uses. This idea does not undermine the Fregean principle that meaning determines reference. Rather, it supports the Fregean principle from a more realistic perspective. In the original twin-earth argument, Putnam discusses the pre-scientific stage that both earthians and twin-earthians do not have any tools to determine the atomic composition of H₂O and XYZ. The pre-scientific stage is, in my theory, the situation of underdetermination.

We are concerned about a natural human language that is constantly changing as if it had some life of its own. Any expectation that a natural human language should show some logical perfection is a naive hope. We have no pre-theoretical ground to assume that meanings should enable us to pick out all *correct* references despite the lack of information. We have no empirical ground to assume that the Fregean principle cannot be maintained tentatively when we lack of information.

¹³ Berger's (1989) proposal for *reference change* is telling in that he takes the language change into serious consideration in his theory of meaning and reference. Unfortunately, however, he does not pay any attention to the change of meaning.

Pustejovsky (p.c.) holds a soft objection to Putnam's thesis. He says that the meaning of a word is opaque to us as Putnam maintains, but that Putnam's meaning is actually something different from the meaning of mentalistic semantics. He then draws a distinction between meaning and semantic meaning; the former for Putnam, and the latter for mentalistic semantics. I do not want to take this soft objection to Putnam's thesis. I still do not see any empirical motivation for external meaning, nor do I think that the theory of external meaning i.e. non-semantic meaning following Pustejovsky – has any broad empirical coverage over natural language semantics. Furthermore, not many things in the world are reduced to the atomic/chemical decomposition like *Au*, H_2O , and XYZ. For all these reasons, I strongly believe that meaning is not a fixed monad; that meaning can underdetermine reference when we lack of appropriate information; and that meaning, like reference, may change over time.

5. More Puzzles and Consequences of the Theory

The view of mentalistic semantics is not entirely new to this paper. I am just trying to make everything explicit that has been assumed quite implicitly by many linguists, to introduce new arguments for the theory based on old findings in the psychology literature, and to examine some consequences of the theory. Although I do not intend to provide a full examination of the classical puzzles in the philosophy of language in this paper, the theory have direct interesting consequences to some of the old-time puzzles. I will mention a few here briefly.

Most of all, mentalistic semantics based on the theory of phenomenological references eliminates the famous Frege/Russell's puzzles concerning reference. First, Russell's (1919) question of how the statement *The present King of France is bald* can be meaningful with apparent reference to nonexistence is not a problem at all, since the *present King of France* refers to a psychological entity in mentalistic semantics. By the *present King of France*, we just establish a psychological entity in our mind which is currently the King of France. It does not matter whether France has a King or a President. Thanks to Gricean (1975) maxims of language, we take the *present King of France* as a qualitatively and co-operatively informative and adequate statement. My point becomes clear here: the Russellian puzzle of apparent reference to nonexistence matters only when you have adequate world knowledge that France does not have a King at present. To those who have no such knowledge, the compositional meaning of the *present King of France* enables them to establish a phenomenological entity called the present King of France. Even to those who have adequate knowledge about the political system in France, the *present King of France* provides an imaginary character in their head. In this sense, the *present King of France* is analogous with Kripke's (1972) fictional characters or mythical creatures.

Secondly, as Jackendoff (2002) mentions, a sentence S is true if the purported referent of S is in one's knowledge base or the readily available context. Thus, the statement *The present King of France is bald* is true if the psychological reference of the *present King of France* is bald in our head. When the psychological reference of the *present King of France* is not bald, the opposite statement *The present King*

of *France is not bald* is true, instead. To those who have no prior knowledge about France, both statements (*The present King of France is bald* and *The present King of France is not bald*) are meant to be informative according to Gricean maxims. Then, the decision of the truth-value of these sentences is never a question, and in fact meaningless.¹⁴ Both sentences are assumed to be true and informative according to Gricean maxims. Thus, Russell's logical puzzle of how both S and Not-S cannot be true – i.e. the puzzle of excluded middle – does not rise at all in our mind.

Finally, in the theory of phenomenological references, Frege's (1892) sense/reference bifurcation matters only in our psychological experiences. The statement *The morning star is the evening star* is informative to a listener who has no idea of astronomy, since two phenomenological entities surprisingly turn out to be identical. Again, language users do not care about the truth value of the paradoxical statement. Overall, these classical puzzles that have bothered many philosophers and that have given fame and great reputation to some scholars do not matter at all in the mentalistic theory of meaning and reference.

6. Conclusion

I have so far elaborated the theory of phenomenological references, making justifications based on psychological findings, presenting plausible models of the theory from linguistics, and attempting to solve some classical puzzles in terms of the current proposal. The validity and the soundness of a theory depend upon its empirical strength. So I have tried to explicitly show the empirical power of mentalistic semantics that is hard to achieve in a theory based on externalism.

According to the theory of mentalistic semantics, we are living in a phenomenological world. In this phenomenological world are such things as phenomenological rivers, mountains, oceans, our beloved friends, etc. The phenomenological rivers and oceans are filled with phenomenological water, and our phenomenological friends drink the phenomenological water. What linguistic expressions refer to is the things and events in the phenomenological world. Obviously, there is a real world outside us. But linguistic expressions cannot and do not refer to things and events out there. Linguistic expressions refer to the phenomenological world in our mind, and the relationship between the physical outside world and the phenomenological world is determined by psychology, and not by formal logic.

<References>

- Ames, A. Jr. 1955. *An Interpretive Manual for the Demonstrations in the Psychology Research Center*. Princeton Univ. Press, Princeton.
- Ax, A. F. 1953. The Physiological Differentiation of Fear and Anger in Humans. *Psychosomatic Medicine*, 15, 433–442.
- Bechara, A., D. Tranel, H. Damasio, R. Adolphs, C. Rockland, and A. Damasio. 1995.

¹⁴ Formal semanticists always look for the truth value of a sentence. On the contrary, the truth value of a sentence does not matter in conceptual semantics; conceptual semanticists just care about what is happening inside our head when we understand the meaning of a sentence.

- Double Dissociation of Conditioning and Declarative Knowledge Relative to the Amygdala and Hippocampus in Humans. *Science*, 269, 1115–1118.
- Berger, A. 1989. A Theory of Reference Transmission and Reference Change. *Midwest Studies in Philosophy*, XIV, 180–198.
- Cannon, W. B. 1927. The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *American Journal of Psychology*, 39, 106–124.
- Carnap, R. 1956. *Meaning and Necessity*, 2nd edition. Univ. of Chicago Press, Chicago.
- Chomsky, N. 1959. Review of B. F. Skinner's (1957) Verbal Behavior. *Language*, 35.1, 26–58.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
- d. Saussure, F. 1958. *Course in General Linguistics*. McGraw-Hill Book Company, New York. Trans. Wade Baskin.
- Davis, M. 1992. The Role of the Amygdala in Conditioned Fear. In J. P. Aggleton (ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*. Wiley Liss, New York, pp. 255–306.
- Devinsky, O. and D. M. Bear. 1984. Varieties of Aggressive Behavior in Temporal Lobe Epilepsy. *American Journal of Psychiatry*, 141, 651–656.
- Ekman, P. 1971. Universals and Cultural Differences in Facial Expressions of Emotion. In J. Cole (ed.), *Nebraska Symposium on Motivation*, 19. University of Nebraska Press, Lincoln, pp. 207–283.
- Fodor, J. A. 1983. *Modularity of Mind*. MIT Press, Cambridge, MA.
- Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophisches Kritik*, pp. 22–50.
- Gogel, W. C. and D. L. Mertz. 1989. The Contribution of Heuristic Process to the Moon Illusion. In M. Hershenson (ed.), *The Moon Illusion*. Erlbaum, Hillsdale, NJ, pp. 235–258.
- Grice, H. P. 1975. Logic and Conversation. In P. Cole and J. L. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*. Academic Press, New York, pp. 41–58.
- Hershenson, M. 1999. *Visual Space Perception: A Primer*. MIT Press, Cambridge, MA.
- Ittelson, W. 1960. *Visual Space Perception*. Springer, New York.
- Ittelson, W. 1968. *The Ames Demonstration in Perception*. Hafner Publishing Co., New York.
- Jackendoff, R. 1983. *Semantics and Cognition*. MIT Press, Cambridge, MA.
- Jackendoff, R. 1987. *Consciousness and the Computational Mind*. MIT Press, Cambridge, MA.
- Jackendoff, R. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.
- Jackendoff, R. 1997. *The Architecture of the Language Faculty*. MIT Press, Cambridge, MA.

- Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford Univ. Press, Oxford.
- James, W. 1890. *Principles of Psychology*. Henry Holt, New York.
- Kaplan, D. 1989. Demonstratives. In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford Univ. Press, Oxford, pp. 481–563.
- Kaufman, L. and I. Rock. 1962. The Moon Illusion, I. *Science*, 136, 953–961.
- Koffka, K. 1935. *Principles of Gestalt Psychology*. Harcourt Brace Javanovich, New York.
- Köhler, W. 1929. *Gestalt Psychology*. Liveright, New York.
- Kripke, S. 1972. *Naming and Necessity*. Harvard Univ. Press, Cambridge, MA.
- Landis, C. and W. A. Hunt. 1932. Adrenalin and Emotion. *Psychological Review*, 39, 467–485.
- LeDoux, J. F. 1994. Emotion, Memory, and the Brain. *Scientific American*, 270, 50–57.
- Mesquita, B. and N. H. Frijda. 1992. Cultural Variations in Emotion: A Review. *Psychological Bulletin*, 112, 179–204.
- Montague, R. 1973. The Proper Treatment of Quantification in Ordinary English. In K. Hintikka, J. Moravcsik, and P. Suppes (eds.), *Approaches to Natural Language*. Kluwer, Dordrecht, pp. 221–242.
- Moravcsik, J. M. 1975. Aitia as Generative Factor in Aristotle's Philosophy. *Dialogue*, 14.4, 622–636.
- Moravcsik, J. M. 1981. How do Words Get their Meanings. *The Journal of Philosophy*, 78.1, 5–24.
- Moravcsik, J. M. 1990. *Thought and Language*. Routledge, London.
- Mullan, S. and W. Penfield. 1958. Illusions of Comparative Interpretation and Emotion. *Archives of Neurology and Psychiatry*, 80, 269–284.
- Pustejovsky, J. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA.
- Pustejovsky, J. 1998. Generativity and Explanation in Semantics: A Reply to Fodor and Lepore. *Linguistic Inquiry*, 29.2, 289–311.
- Pustejovsky, J. 2000. Syntagmatic Processes. In A. Cruse (ed.), *Handbook of Lexicography*. Mouton De Gruyter, Berlin.
- Putnam, H. 1975. The Meaning of 'Meaning'. In *Mind, Language, and Reality*. Cambridge Univ. Press, Cambridge, pp. 215–271.
- Putnam, H. 1981. *Reason, Truth, and History*. Cambridge Univ. Press, Cambridge.
- Quine, W. V. O. 1951. Two Dogmas of Empiricism. *Philosophical Review*, 60, 20–43.
- Russell, B. 1919. Descriptions. In *Introduction to Mathematical Philosophy*. George Allen and Unwin Ltd, London, pp. 167–180.
- Schachter, S. and J. A. Singer. 1962. Cognitive, Social and Physiological Determinants of Emotional State. *Psychological Review*, 69, 379–399.

- Schwartz, G. E., D. A. Weinberger, and J. A. Singer. 1981. Cardiovascular Differentiation of Happiness, Sadness, Anger, and Fear Following Imagery and Exercise. *Psychosomatic Medicine*, 43, 343-364.
- Sinha, R. and O. A. Parsons. 1996. Multivariate Response Patterning of Fear and Anger. *Cognition and Emotion*, 10, 173-198.
- Solhkhah, N. and J. Orbach. 1969. Determinants of the Magnitude of the Moon Illusion. *Perceptual and Motor Skills*, 29, 87-98.
- Taylor, D. W. and E. G. Boring. 1942. The Moon Illusion as a Function of Binocular Regard. *American Journal of Psychology*, 55, 189-201.
- Wertheimer, M. 1912. Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161-265.
- Wertheimer, M. 1923. Laws of Organization in Perceptual Forms. In W. D. Ellis (ed.), *A Source-Book in Gestalt Psychology*. Routledge & Kegan Paul, London.

Submitted on: April 14, 2004

Accepted on: November 21, 2004