

A Skeptical Approach to the Audibility of Semantic Properties

Elvira Di Bona

University of Turin

Abstract

The issue of whether we can auditorily perceive meanings (or semantic properties) expressed in a language we understand has been approached through arguments based either on theoretical reasoning or the discussion of psychological effects. I am skeptical about the use of either type of argument. In this paper, I will first explain the limitations of the standard theoretical argument: the phenomenal contrast method. As for psychological phenomena, I will discuss semantic satiation and the Stroop effect. I will summarize why semantic satiation has already been dismissed and, based on said reasoning, will evaluate the Stroop effect, recently brought up in favor of the perceivability of semantic properties. I will show that, just as the experience of semantic satiation does not exhibit the features required by perceptual experience to be characterized as such, the experience of the Stroop effect also lacks these features. Therefore, neither should be used to show that we can perceive meanings. As a consequence, we have not yet produced either a sound theoretical argument or any useful discussion of such psychological phenomena to account for the audibility of the semantic properties of a language we understand.

Keywords: Semantic properties, Meanings, Language understanding, Stroop effect, Method of contrast, Adaptation, Rich vs thin view, Speech sound.

1. Introduction

Questions on the nature of speech sound and how to characterize the content of perceptual experience have been recently linked by philosophers discussing whether the understanding of spoken language is a perceptual experience. The debate centers on whether the experience of hearing a sentence in a known language is perceptual, just like hearing paradigmatic, audible properties, such as pitch or loudness, is perceptual. If this is the case, the properties of which we are auditorily aware are most probably semantic properties, i.e. properties that express linguistic understanding (Peacocke 1992; Strawson 1994/2010; McDowell 1998a, 1998b; Siegel 2006). If this is not the case, then the properties of which we are perceptually aware are located at a lower-level than semantic properties, such as morphosyntactic properties (Voltolini 2020), phonological properties (O'Callaghan 2011) or even the audible properties of pitch and loudness. If this is the

case, semantic properties are thus perceived indirectly and are merely inferred from lower-level ones. The discussion on the perceivability of meaning has developed within the debate on the content of perception, in which philosophers stating that we perceive meanings support the “rich” or “thick” view, whereas those who do not are in favor of the “thin” or “sparse” view. This said, philosophers who reject the rich view of semantic perceptibility may still allow for other higher-level perceptual properties; philosophers in favor of the thin view, instead, may still reject the idea that other higher-level properties are perceivable.

In this paper, I take a skeptical approach to the rich view of language understanding since, to my knowledge, the arguments in support of this view fail to show that semantic properties (or meanings) are audible. I will justify my skeptical approach by exposing the limitations of published arguments and suggesting a new discussion of the Stroop effect. In Section 2, I will explain the limitations of the method of phenomenal contrast, a method commonly used to support the rich view. In Section 4, I will discuss the phenomenon of the Stroop effect, which has recently been used to support the perceivability of written language (Brogaard 2020). I will evaluate this phenomenon when it occurs in spoken language and show that it cannot successfully be used to prove any auditory awareness of semantic properties. The discussion in Section 4 will draw on material developed in Section 3, which describes what features make the experience of a phenomenon a robust perceptual experience. Given that the experience of semantic satiation lacks such features, I will show that it cannot be used to support the perceivability of meaning. The experience of the Stroop effect lacks these features as well, which means that it cannot be used to argue for the rich view either. We still lack, indeed, convincing arguments and reliable phenomena to support the rich view of language understanding, which justifies my skeptical approach to the audibility of semantic properties.

2. Phenomenal Contrast

The method of contrast is based on comparing two experiences—the target experience and the contrasting experience—which are alike in all respects but one. This difference is signaled by a shift in phenomenology which emerges when we compare these experiences. This method seeks to better explain that difference. The phenomenological difference between the two experiences is highlighted by a clear indicator, something which features in the subject’s contrasting experience and which the target experience lacks. This is usually a striking, subjective feeling. Once the phenomenal difference is signaled by this feeling, the task is to explain it. If the phenomenal difference is better explained in terms of the perceivability of a high-level property, then the method has been used to confirm the rich view of perception. Whereas, if the explanation of the phenomenal difference using high-level properties fails to convince, because alternative explanations for that difference in terms of lower-level properties or cognitive properties work better, then the rich view of perception cannot be supported by using the method. That is because the alternative explanation based on lower-level properties or cognitive properties might support either the thin view of perception or a cognitive point of view.

Let us consider an application of the method to explain how it works. Susanna Siegel argued for the rich view of visual perception by using the phenomenal contrast method to show that a high-level property of a natural kind, i.e. a pine tree, is perceivable (Siegel 2006, 2010). In the following passage, she writes:

Suppose you have never seen a pine tree before and are hired to cut down all the pine trees in a grove containing trees of many different sorts. Someone points out to you which trees are pine trees. Some weeks pass, and your disposition to distinguish the pine trees from the others improves. Eventually, you can spot the pine trees immediately: they become visually salient to you. Like the recognitional disposition you gain, the salience of the trees emerges gradually. Gaining this recognitional disposition is reflected in a phenomenological difference between the visual experiences had before and those had after the recognitional disposition was fully developed (Siegel 2010: 100).

Siegel asks the readers to imagine looking at a pine tree for the first time in their life, without knowing that that is a pine tree. They will have a certain visual experience. She then asks to imagine looking at the same pine tree again after some time, when they have learnt to recognize pine trees. There will be an overall phenomenal difference between the first and the second visual experience, which is determined by the new recognitional disposition acquired. The difference is signaled by the appearance of a salient feeling. In the pine tree example, the before/after difference is signaled by the feeling of familiarity the perceiver experiences when looking at the pine tree and recognizing it as such. Siegel describes the appearance of this feeling as follows:

[h]ow the tree looks before and after you become disposed to recognize pine trees is exactly the same; that is, it looks to have certain color and shape properties. But the moment you recognize the tree, you experience a feeling of familiarity, and this feeling accounts for the phenomenological change before and after you gain the disposition. So, on this suggestion the way the tree looks stays the same, before and after you become disposed to recognize it; but the phenomenology of “taking” the tree to be familiar contributes to the phenomenal change accompanying E2 [the contrasting experience] (Siegel 2010: 104).

Once the phenomenal difference is individuated by the striking feeling that signals it, we must review the plausible competing explanations for that difference, and infer the best explanation. Siegel’s best explanation states that the high-level property of being a pine tree is what the viewer perceives in the contrasting experience. The target experience lacks this property, instead. Therefore, this explanation accounts for the rich view of visual perception. Of course, Siegel also evaluates the competing explanations (*ibid.*: 103-104), disregarding those based on low-level properties (the different arrangements of colors and sizes), perceptual judgment, belief, hunch or intuition, and the entertaining of a proposition arising in the mind.

To sum up, the method has two steps: first, to describe the contrasting pair which is supposed to be clearly construed by an overall phenomenological difference; second, to assess the possible explanations of the before/after difference which then, through an inference to the best explanation, leads to the ultimate explanation of the phenomenal difference. The method does not seem unquestionably reliable since there are challenges that concern both steps.

Doubts over the first step stem from the fact that the phenomenal difference is signaled in a highly arbitrary and subjective way. In the case of the pine tree, how can we ensure that we all experience a sense of familiarity upon finally recognizing the pine tree? Is it not possible that someone recognizing a pine tree might experience wonder instead of familiarity, or even no feeling at all? Moreover, the onset of the striking feeling does not show ipso facto that there is any phenomenal difference between the two experiences, since this might be due to an unexplained, subjective change in one's mental state. If the onset of this feeling is so highly arbitrary, the phenomenal difference is too weak to allow any discussion of the nature of the difference. If this is true, the method is baseless.

Perhaps the second step of the method of contrast is more reliable? The two-fold nature of this step involves assessing plausible explanations for the phenomenal difference and choosing the best one. There are usually a number of plausible explanations (Siegel discusses low-level properties, perceptual judgment, belief, etc.) but it is far from obvious that they cover all possibilities. Something may be left out because the choice to discuss specific explanations is somehow arbitrary, and ill-founded. Unless one finds reasons to justify the range of specific explanations to be taken into account, there is no firm basis for any inference to the best explanation.

Thus, the method of contrast appears unreliable when evaluating whether a property is part of the content of perception. I am skeptical about the reliability of the method when applied to the specific case study of the high-level property of a natural kind but, obviously, the challenges also apply to contrast cases designed to show that we can perceive semantic properties, namely contrast cases of hearing speech in an unknown foreign language first and then hearing it again when we have learned the language.

3. Auditory Adaptation, Semantic Satiation, and Perceptual Experience

Aristotle was the first philosopher to describe a case of perceptual adaptation in his "waterfall illusion". After looking at a waterfall, when the viewer looks at a stationary object, like a tree nearby, it will appear to be moving upwards although it is not. This suggests that the property of motion can be part of visual experience. The phenomenon of perceptual adaptation has been used to support the rich view of visual perception when focusing on properties like numerosity, or the specific emotions expressed by human faces (Fish 2013; Block 2014). This can be justified by saying that, when a subject adapts to a certain property such as motion, this property appears in the content of visual perception as an after-effect in the form of a robust direct acquaintance. The experience of the after-effect cannot be reduced to the experience of lower-level properties, and occurs without involving perceptual judgements.

With similar reasoning, auditory adaptation has been used to support the rich view of auditory perception and to show that the high-level property of gender voice is audible (Di Bona 2017). Moreover, when evaluating whether semantic properties are also perceivable, it has been suggested that the phenomenon of semantic satiation—a specific form of auditory adaptation—shows that they are perceivable (Nes 2016; Brogaard 2020). Semantic satiation is the phenomenon in which a word rapidly repeated aloud many times, heard repeatedly, or read silently for a prolonged

length of time, loses its meaning and is perceived by the listener or reader as meaningless (Pilotti and Antrobus 1997).¹ This is a form of perceptual adaptation since the meaningless word “appears” only after one adapts to the stimulus (a spoken or read word), and it is experienced as an after-effect. Those who support the perceptibility of semantic properties on the basis of semantic satiation usually point to semantic satiation being a form of perceptual adaptation, and assert that all forms of adaptation show by default that the property we are adapting to is perceivable. At first glance this seems to be true, but upon closer inspection it is clear that the effect of semantic satiation, despite being a form of adaptation, cannot be used to show that we can hear semantic properties (Di Bona 2020). A strategy has thus been advanced on the assumption that the experience of a phenomenon must have specific characteristics for it to be a reliable perceptual phenomenon. Therefore, given that the experience of semantic satiation lacks these characteristics, it is unlikely that it is a perceptual experience.² The fact that it is a form of adaptation does not show ipso facto that semantic properties are audible.

The features which make the experience of a certain phenomenon reliably perceivable (auditorily) as a high-level property are:

- 1) The perceptual experience cannot be reduced to the experience of lower-level properties;
 - 2) It has epistemic immediacy;
 - 3) It does not involve perceptual judgements.
- 1) When examining any argument in support of the rich view by discussing the perceptual experience of high-level properties (such as gender or semantic properties expressed by spoken language), we must ensure that the phenomena through which we perceive those properties cannot be explained by our experience of lower-level properties. In auditory adaptation to gender properties, researchers have studied whether subjects adapted to the low-level property of pitch rather than gender (Di Bona 2017: 2640). They found that pitch does not appear as an after-effect and were therefore unable to state that the phenomenon could be interpreted as support for the thin view. Rather, the phenomenon of semantic satiation does not seem to be a purely semantic one. Apparently one can habituate to an earlier lexical level of processing known as “verbal transformations” (Balota and Black 1997). For example, after many repetitions of the spoken term “royalty”, instead of losing its meaning for a listener, the latter may experience auditory illusions and instead hear “loyalty”, “realty” or “specialty”. Listeners may also hear “royalty” again when habituated to other lexical units (Pilotti and Kurshid 2004). Martina and Voltolini (2017) and Voltolini (2020) suggest that one might habituate to morphosyntactic components, since there may be cases of semantic satiation in which repeatedly hearing a certain expression induces the subject to hear a distortion of the word, regardless of whether the expressions involved are meaningful. As we cannot

¹ Since the phenomenon of satiation may also occur with meaningless words, in this paper I will use the expressions “semantic satiation” and “satiation” interchangeably.

² Of course, one might question that the characteristics I take into consideration make a phenomenon perceptively reliable. I am providing, indeed, necessary but not sufficient conditions for a phenomenon to be perceptively reliable. I am only claiming that these characteristics seem to commonly characterize perceptual experience, and are believed to make an experience perceptual also by other philosophers, such as O’Callaghan (2011) and Masrour (2011).

rule out explanations of semantic satiation in terms of lower-level properties, we cannot exclude that the semantic properties we seem to experience when satiated can be reduced to lower-level properties. Therefore, the effect of semantic satiation cannot support the thesis that we can hear meanings.

- 2) Perceptual experience is typically isolated from doxastic influence, meaning that one's beliefs and desires do not strongly influence or affect perception. For example, we have a perceptual experience of the color red because, when we see red, and even if we are told we are seeing blue, we cannot help but see red regardless. The belief that the color in front of us is not red does not affect our perception of red. This isolation from beliefs and desires means that perception gives us an immediate epistemic grasp of the outer world. A phenomenon which allows such epistemic immediacy is likely quite stable and reliable as a basis for forming perceptual beliefs. Epistemic immediacy is due to inferential isolation. The gender properties which emerge in our auditory experience as an after-effect of auditory adaptation last several minutes. This guarantees that this specific case of adaptation is persistent; it is not mediated by beliefs or desires and provides a good basis for forming perceptual beliefs and desires (Di Bona 2017). The phenomenon of semantic satiation, however, is not robust in all circumstances. There are variations in the intensity of this phenomenon due to aging which suggest that it is perhaps less consistent over a lifetime than one might think. Therefore, we cannot be sure that semantic satiation provides the necessary perceptual stability to support immediate epistemic access to the semantic properties which then can successfully help to form justified perceptual beliefs (Di Bona 2020: 232).
- 3) The phenomenon under examination works in favor of the rich view only if it can be explained in terms of the perception of high-level properties. Therefore, it is crucial to rule out the explanation using lower-level properties and to consider it possible to exclude any explanation in cognitive terms, including one that describes the experience of auditory adaptation as a concept-based experience which constitutes perceptual judgements. This cognitive explanation was successfully excluded in the case of habituation to gender voices (Di Bona 2017: 2640) but it cannot be totally ruled out in semantic satiation. Indeed, to test the cognitive explanation of semantic satiation, researchers have studied whether a word loses its meaning when it is heard and when it is presented in different ways, namely displayed on a screen or read aloud by the subjects themselves. Their assumption is that a word is connected to a certain concept which can be conveyed through different modalities. If the word still loses its meaning despite being presented differently, this may indicate adaptation to the attached concept. Pilotti and Khurshid (2004: 1010) examined the effects of semantic satiation presented in different modalities (auditory, visual plus auditory, visual) and verified that the meaningfulness of those words was attenuated. This leads to the conclusion that in semantic satiation, subjects can adapt to different concepts representing the same meanings. These concepts can generate perceptual judgments, thus semantic satiation cannot be explained by appealing to semantic properties.

In the next section, I will discuss the Stroop effect to see whether it meets the three criteria above. This will enable us to conclude that it does not serve the cause of the rich view of spoken language understanding.

4. The Stroop Effect

In a recent paper, Brogaard (2020) defends the non-inferential view of speech comprehension, in which we comprehend speech by perceiving or grasping semantic properties or meanings directly, rather than by inferring them from linguistic principles or lower-level, perceivable properties. She uses “speech” in a broad sense to refer to spoken language and to written messages, symbols, and Braille. She supports her view by discussing certain phenomena, including the Stroop effect. This effect is meant to show that we can directly experience meanings conveyed through written language. I will evaluate this phenomenon when it occurs in spoken language and argue that it does not support the theory that we can perceive semantic properties through hearing.

The Stroop effect was first studied in visual science by John Ridley Stroop in 1935. It is a form of interference which occurs when a subject is asked to state the ink color of a printed word which names another color. For example, subjects are asked to identify the color of the word “green”, which is printed in red, not simply to read out the word “green”. It takes longer to name the ink color in this situation than when the word is printed in a congruent color, namely green. The common explanation is that reading and understanding a word is more automatic than analyzing its appearance or color, so the interference of it being printed in an incongruent color creates a processing delay for the subject. Brogaard affirms that:

The effect [...] appears to indicate that the grasp of meanings occurs automatically as a result of sensory processing, which points to the non-inferential view of meaning comprehension (Brogaard 2020: 17).

Brogaard concludes that the effect shows that grasping the meaning of a word is more automated than color naming, which occurs as a “result of sensory processing”. This leads directly to the claim that semantic properties, or meanings, are visually perceivable and are not inferred from lower-level properties; this is precisely what the rich view of perception claims.

Even though most research on the Stroop effect has focused on vision, there are studies investigating the Stroop effect on hearing, in particular when dealing with pitch (Cohen and Martin 1975), loudness (Morgan and Brandt 1989) and gender words (Green and Barber 1981, 1983; Gregg and Purdy 2007). I will discuss the auditory Stroop effect because my objective here is to investigate the understanding of spoken terms, not written terms. The stimuli here are terms spoken by specific voices. They utter meanings that may be congruent or incongruent with the pitch, loudness, or the gender of the voice. Therefore, when a voice utters terms like “high” or “low”, pronounced at a high or low pitch, the Stroop effect for pitch is being studied. Obviously, when the word “high” is pronounced at a high pitch the stimuli are congruent; they are incongruent when the word “high” is pronounced at a low pitch. To study the effect of loudness, the spoken words are “loud” and “soft”, pronounced either loudly or softly. Concerning the Stroop effect on gender words, Green and Barber (1981, 1983) asked participants to identify a speaker’s gender while the speaker said “man” and “girl”. Subjects had to recognize a male voice incongruently saying “girl” and congruently saying “man”, and a female voice incongruently saying “man” and congruently saying “girl”. These studies on the auditory Stroop effect confirm the results in visual Stroop studies, namely that it takes longer to identify the sex of the speaker, or the pitch and loudness of a tone,

when sex, pitch and loudness stimuli interfere with simultaneous and incongruent stimuli. Let us now evaluate whether the experience of the auditory Stroop effect can be characterized by the features required by a perceptual experience to be perceptual, namely 1) non-reducibility to the experience of lower-level properties, 2) epistemic immediacy and 3) the exclusion of perceptual judgements.

- 1) Interference from the Stroop effect slows the identification of the gender of the voice (or its pitch or loudness) but it does not indicate whether semantic properties can be perceived as such. The effect shows that grasping meaning interferes and captures the attention but does not prove that semantic properties arise as genuine perceivable properties. The effect shows something “in negative”, namely that grasping meaning “prevents” (or slows down) the identification of the gender, pitch, or loudness of the voice. However, it does not positively show that meaning is perceived as such. Consequently, we cannot rule out the possibility that when semantic properties are finally processed, they are preceded by multiple lower-level properties (such as phonological, lexical or morphosyntactic properties) which form the basis of semantic properties. This is sufficient to claim that the non-reducibility of semantic properties to lower-level ones cannot be ensured.
- 2) The meaning of the spoken term, e.g. “man”, distracts the listener from identifying the gender of the speaker when spoken by a female voice. We would expect it to show that the meaning of the spoken term is experienced automatically and that the term is understood before the gender of the voice is identified. Nevertheless, the fact that the meaning of a spoken term, e.g. “man” is recognized before the speaker’s identity is determined does not show that we acquire meanings automatically and sensorily. It merely shows that we recognize meaning *before* we process the data about the sex of the speaker and tells us little about the speed or automaticity with which we grasp the meaning itself. The Stroop effect is simply based on a comparison of two skills and shows which is faster. Although grasping meaning is faster than recognizing the gender of a voice, grasping meaning still takes place through an underlying process. The explanation of the Stroop effect does not rule out this possibility. If it does not show that meaning is experienced automatically, but merely that this occurs faster than the subject can identify the speaker’s gender, we obviously cannot be sure that experiencing meaning gives epistemic immediacy to semantic properties. Rather the contrary is true: if we cannot rule out the mediation of an underlying process when acquiring meaning, there is no room to justify epistemic immediacy to meanings. Nevertheless, let us assume that the Stroop effect actually shows automatic perception of semantic properties, and check whether automatically experiencing the meaning of spoken terms provides epistemic immediacy to semantic properties. This can only happen unconsciously. However, if semantic properties have epistemic immediacy without the listener being aware of it, it is unlikely that they can properly introspect such properties and use them to form justified beliefs. Therefore, it appears impossible to grant epistemic immediacy to any semantic properties experienced through the Stroop effect.
- 3) Two arguments suggest that the Stroop effect phenomenon, instead of being explained by the perception of semantic properties, can be characterized by the application of perceptual judgements. The first is that (as in the case of the argument about the involvement of lower-level properties), if the effect has a

“negative” outcome (i.e. it shows that one factor prevents or slows down an event), then we cannot rule out the idea that the Stroop effect is due to the use of perceptual judgments. The negative outcome merely indicates that grasping meanings prevents or slows down the identification of the gender, pitch or loudness of a voice, instead of saying why and how semantic properties are perceived directly and through a sensory process.

Secondly, and more importantly, for the specific Stroop effect involving the recognition of the speaker’s sex when uttering gender-related words, results suggest that perceptual judgements rather than genuine perception of semantic properties are involved in the interference generated by the Stroop effect. Gregg and Purdy (2007) asked participants to identify a speaker’s sex while they said words commonly used to refer to gender or sex-related relationships, such as “male”, “female”, “mother”, “grandfather”. Congruent stimuli words corresponded to the sex of the speaker, while incongruent stimuli words were spoken by a person of the opposite sex, for example “female” or “sister” spoken by a man. As expected, the results of this experiment showed that the fastest reaction time occurred in congruent conditions. Moreover, a smaller Stroop effect occurred for relationship label words (*ibid.*: 552) when compared to gender words, i.e. subjects listening to a male voice uttering words like “mother” or “grandmother” recognized the sex of the speaker faster than when the voice said “female”. While Gregg and Purdy used conflicting gender terms and sex-related relationship words, Most *et al.* (2007) investigated the Stroop effect by opposing gender names with gender-stereotyped words. The stimuli they used were female stereotyped words, such as “cheerleader”, “lipstick”, “makeup” and “pink”; male stereotyped words, such as “baseball”, “football”, “rough”, “soldier” and “tough”; female names, such as “Cindy”, “Jenny”, “Jill”, “Nancy”, “Rachel” and male names, such as “David”, “Henry”, “John”, “Michael” and “Peter”. The study also required subjects to identify speaker’s voices as male or female. It was predicted that subjects would be slower to perform the task when the speaker uttered a word stereotypically connected with the opposite sex. The Stroop effect occurred for names and gender-stereotyped words, but it was substantially greater for names, suggesting that “the more strongly a word is associated with one sex or the other, the more of a Stroop-like effect it will yield” (*ibid.*: 291).

To gain a deeper understanding of the phenomenon, to further investigate how the meaning of a word influences subjects’ performance, and to see whether it is possible to develop a graded auditory Stroop effect, Gregg and Purdy (*ibid.*: 552) tested reaction times when subjects were presented with words associated with sex to different degrees. The stimuli were relationship label words (e.g. “daughter”, “father”), names (e.g. “Susan”, “Kevin”), jobs (e.g. “nanny”, “butler”), clothing (e.g. “dress”, “tuxedo”), adjectives (e.g. “beautiful”, “handsome”) and nouns (e.g. “lipstick”, “cologne”). The results clearly showed that incongruent words slowed down the recognition of the speaker’s gender in a significant way for all word categories except for adjectives and nouns. This suggests that “the size of the Stroop effect depends on how strongly the stimulus words hold a gender connotation” (*ibid.*). That is, when the task is to identify the sex of a voice, we are more distracted when the voice is uttering words usually associated with the gender of the voice itself. For example, when a male voice says “grandmother”, “Susan”, “nanny” or “dress”, it takes far longer for subjects to recognize the voice as male than

when he says “lipstick” or “beautiful”. The results of these studies by Gregg and Purdy and Most *et al.* seem to suggest that the extent of the Stroop effect strongly depends on the strength of the conceptual connection between a specific gender and a determinate word: it takes longer to identify the sex of a voice when a female voice is saying “Kevin” or “butler” than when that female voice is uttering terms like “cologne”; at the same time, it takes longer to identify the sex of a voice when a male voice utters “male” than when it utters “father” and it takes longer to identify the sex of a voice when it says a gendered-stereotyped word like “football” than a name like “David”.

All of these experiments seem to suggest that we have three levels of mental state: the level of background knowledge (made up of concepts we usually associate with a specific gender), the meaning of a term (that is how we usually characterize it) and vocal quality (the gender of the voice). Clearly, background knowledge influences the meaning of a term to a certain degree and seems to anticipate and actually impact the proper attribution of meaning to a term. These studies show that when identifying a speakers’ sex by listening to the words they are saying, not all words create the Stroop effect to the same extent precisely because background knowledge attributes different meanings to the stimuli words. This determines various degrees of interference. Given that background knowledge is formed of judgements, including perceptual judgements, we can conclude that the experience of the Stroop effect when identifying the sex of someone saying gender words is strongly influenced by perceptual judgements in a subject’s existing gender-related background knowledge. Therefore, the Stroop effect experience is likely to be an experience mediated by perceptual judgements instead of a perceptual experience of semantic properties.

As discussed, we see how the Stroop effect fails to prove that we can perceive semantic properties. This is clear from reasoning that if the experience of a phenomenon lacks specific features, this phenomenon will not determine perceptual experience. Such features have already been used to show that semantic satiation does not equate to perceptual experience. In conclusion and to my knowledge, no psychological effects have yet been suggested which show that we can auditorily perceive meanings. I have also successfully challenged the standard theoretical argument about the method of phenomenal contrast. The remaining options for characterizing the nature of spoken language understanding are either that semantic properties are processed at a purely cognitive level—which seems to be supported by the way schema-driven segregation operates when parsing the auditory scene (Bregman 1990)—that they are experienced at the perceptually-based cognitive level (Martina and Voltolini 2017; Voltolini 2020), or that they can be reduced to lower-level properties—as shown in recent analyses of the perception of meaning (O’Callaghan 2011).³

References

- Balota, D.A. and Black, S. 1997, “Semantic Satiation in Healthy Young and Older Adults”, *Memory & Cognition*, 25, 190-202.

³ I would like to thank two anonymous referees from *Argumenta* for their helpful comments.

- Block, N. 2014, "Seeing-As in the Light of Vision Science", *Philosophy and Phenomenological Research*, 89, 3, 560-72.
- Bregman, A.S. 1990, *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA: The MIT Press.
- Brogaard, B. 2020, "Seeing and Hearing Meanings: A Non-Inferential Approach to Speech Comprehension", in Chan, T. and Nes, A. (eds.), *Inference and Consciousness*, New York: Routledge, 99-124.
- Cohen, G. and Martin, M. 1975, "Hemisphere Differences in an Auditory Stroop Task", *Perception and Psychophysics*, 17, 79-83.
- Di Bona, E. 2017, "Towards a Rich View of Auditory Experience", *Philosophical Studies*, 184, 2629-43.
- Di Bona, E. 2020, "On Hearing Meanings: Reflections on the Method of Contrast, Adaptational Effects, and Semantic Satiation", *Rivista di Filosofia*, 111, 2, 215-35.
- Fish, W. 2013, "High-level Properties and Visual Experience", *Philosophical Studies*, 162, 43-55.
- Gregg, M.K. and Purdy, K.A. 2007, "Graded Auditory Stroop Effects Generated by Gender Words", *Perceptual and Motor Skills*, 105, 549-55.
- Green, F.J. and Barber, P.J. 1981, "An Auditory Stroop Effect with Judgments of Speaker Gender", *Perception and Psychophysics*, 30, 459-66.
- Green, F.J. and Barber, P.J. 1983, "Interference Effects in an Auditory Stroop Task: Congruence and Correspondence", *Acta Psychologica*, 53, 1067-70.
- Martina, G. and Voltolini, A. 2017, "Perceiving Groupings, Experiencing Meanings", *Rivista di Estetica*, 66, 22-46.
- Masrour, F. 2011, "Is Perceptual Phenomenology Thin?", *Philosophy and Phenomenological Research*, 83, 366-97.
- McDowell, J. 1998a, *Meaning, Knowledge and Reality*, Cambridge, MA: Harvard University Press.
- McDowell, J. 1998b, *Mind, Value and Reality*, Cambridge, MA: Harvard University Press.
- Morgan, A. and Brandt, J.F. 1989, "An Auditory Stroop Effect for Pitch, Loudness, and Time", *Brain and Language*, 36, 592-603.
- Most, B.S., Verbeck Sorber, A., Cunningham, J.G. 2007, "Auditory Stroop Reveals Implicit Gender Associations in Adults and Children", *Journal of Experimental Social Psychology*, 43, 2, 287-94.
- Nes, A. 2016, "On What We Experience When We Hear People Speak", *Phenomenology and Mind*, 10, 58-85.
- O'Callaghan, C. 2011, "Against Hearing Meanings", *The Philosophical Quarterly*, 61, 783-807.
- Peacocke, Ch. 1992, *A Study of Concepts*, Cambridge, MA: The MIT Press.
- Pilotti, M., Antrobus, J.S. and Duff, M. 1997, "The Effect of Presemantic Acoustic Adaptation on Semantic 'Satiation'", *Memory & Cognition*, 25, 305-12.
- Pilotti, M. and Khurshid, A. 2004, "Semantic Satiation Effect in Young and Older Adults", *Perceptual and Motor Skills*, 98, 999-1016.
- Siegel, S. 2006, "How Can We Discover the Contents of Experience?", *Southern Journal of Philosophy*, 45, 127-42.
- Siegel, S. 2010, *The Content of Visual Experience*, New York: Oxford University Press.

- Strawson, G. 1994/2020, *Mental Reality*, Cambridge, MA: The MIT Press.
- Stroop, J.R. 1935, "Studies of Interference in Serial Verbal Reactions", *Journal of Experimental Psychology*, 18, 6, 643-62.
- Voltolini, A. 2020, "Different Kinds of Fusion Experiences", *Review of Philosophy and Psychology*, 11, 203-20.