POLSKIE TOWARZYSTWO SEMIOTYCZNE

STUDIA SEMIOTYCZNE

Tom XXXI • Numer 2 • 2017

studiasemiotyczne.pts.edu.pl

„Studia Semiotyczne" powstały w 1970 r. z inicjatywy prof. Jerzego Pelca i były przez niego wydawane do 2015 r. jako czasopismo nieperiodyczne. Od 2016 r. czasopismo ukazuje się równolegle w druku i w Internecie jako półrocznik. W „Studiach Semiotycznych" są publikowane artykuły z pogranicza filozofii i semiotyki, w szczególności z zakresu: analitycznej filozofii języka, ogólnej teorii znaku, zastosowań metod semantycznych w filozofii, filozoficznych aspektów lingwistyki, psycholingwistyki i kognitywistyki, semiotycznych aspektów filozofii umysłu, filozoficznych konsekwencji metalogiki i teorii języków formalnych, analizy języka filozofii i argumentacji filozoficznej oraz historii idei semiotycznych i logicznych. Wydawcą „Studiów Semiotycznych" jest Polskie Towarzystwo Semiotyczne. Zasady składania tekstów do Redakcji, prawne i etyczne aspekty publikacji oraz procedura recenzyjna stosowana przez Redakcję są szczegółowo opisane w zakładce „Do Autorów" na stronie:

http://studiasemiotyczne.pts.edu.pl

Studia Semiotyczne (*Semiotic Studies*) is a journal founded in 1970 by Jerzy Pelc, who was its Editor-in-Chief up until 2015. Between 1970 and 2015 *Studia Semiotyczne* was published non-periodically. In December 2015 *Studia Semiotyczne* was formally transformed into a six-monthly published simultaneously in print and on the Internet. The journal publishes papers that fall on the borderline between philosophy and semiotics and in particular, in the fields of analytic philosophy of language, general theory of signs, applications of semantic methods in philosophy, philosophical aspects of linguistics, psycholinguistics and computer science, semiotic aspects of philosophy of mind, philosophical consequences of metalogic and the theory of formal languages, analysis of the language of philosophy and the philosophical argumentation, and history of ideas in semiotics and logic. *Studia Semiotyczne* is published by Polskie Towarzystwo Semiotyczne (The Polish Semiotic Society). The guidelines for submitting manuscripts to the Editors, the legal and ethical aspects of the publication, and the review procedure used by the Editors are described in detail in the "For Authors" section of the website:

http://studiasemiotyczne.pts.edu.pl

Polskie Towarzystwo Semiotyczne

STUDIA SEMIOTYCZNE

Tom XXXI • nr 2

PÓŁROCZNIK

Philosophy of Language: Extending the Field

Warszawa • 2017

Adres redakcji:
Krakowskie Przedmieście 3, 00-047 Warszawa
e-mail: studiasemiotyczne@pts.edu.pl
http://studiasemiotyczne.pts.edu.pl/

ISSN 0137-6608

Redakcja merytoryczna, skład elektroniczny, korekta językowa w języku polskim, zamieszczanie plików PDF na stronie internetowej „Studiów Semiotycznych”, koordynowanie prac i opieka techniczna zespołu redakcyjnego nad procedurą recenzyjną w systemie *peer review* w formule otwartego dostępu, realizacja prac redakcyjnych mających na celu zagwarantowanie oryginalności publikowanych artykułów naukowych są finansowane w ramach umowy 614/P-DUN/2016 ze środków Ministra Nauki i Szkolnictwa Wyższego przeznaczonych na działalność upowszechniającą naukę.

Ministerstwo Nauki
i Szkolnictwa Wyższego

CONTENTS

SPIS TREŚCI

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 5–11
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.01

EDITORIAL

John Searle identified in *Speech Acts* (1969) the following questions as forming the subject matter of the philosophy of language: "How do words relate to the world? (…) How do words stand for things? What is the difference between a meaningful string of words and a meaningless one? What is it for something to be true? or false?" (Searle 1969, p. 3). These questions are echoed by Michael Morris in his *Introduction to the Philosophy of Language*: "What is language? What is it for words to have meaning? What is the meaning of words?" (Morris 2007, p. 1),[1] and they are reflected in the division into three parts applied in the second edition of the monumental Blackwell *Companion to the Philosophy of Language* (Hale, Wright and Miller, eds. 2017). The respective parts of the *Companion* deal with meaning and theories of meaning; language, truth, and reality; and reference, identity, and necessity. Martin Davies observed in *The Blackwell Guide to the Philosophy of Language* (2006) that the "foundational questions in philosophy of language concern the nature of meaning, understanding, and communication" (Davies 2006, p. 29), which basically means that "philosophers are interested in three broad aspects of language: syntax, semantics and pragmatics" (Martinich 2009, p. 1). This last remark implicitly stresses the overall importance of the semiotic approach to the discussed field, since, as observed by Umberto Eco in the introductory comments to his *Semiotics and the Philosophy of Language* (1984):[2] "[a] general semiotics is nothing else but *a* philosophy of language and

[1] More recently, Chris Daly has extended the list to ten key questions, see Daly (2013, p. 1–11).

[2] For a brief background discussion of these issues and relevant references, see Stalmaszczyk (2015).

(…) the 'good' philosophies of language, from *Cratylus* to *Philosophical Investigations*, are concerned with all the semiotic questions" (p. 4).[3]

Max Kölbel has recently discussed the 'new directions in the philosophy of language', and observed that "much recent work in the philosophy of language has been concerned in one way or another with questions concerning the interaction between the standing meaning of expressions and the context in which they are used" (Kölbel 2012, p. 251). Among the new directions, Kölbel mentions double index semantics, the relations between what is said and implicatures, between unarticulated constituents and compositionality, and contextualism and relativism. The constant importance of the foundational questions notwithstanding, recent studies devoted to philosophy of language document a further broadening of the scope of research.[4] Some of the topics currently discussed and analyzed include a wide range of linguistic phenomena, various speech acts, different aspects of non-literal language, the complex relations between language and cognition, and the interconnections between philosophy of language and other fields (especially linguistics, philosophy of mind, philosophy of literature, and argumentation theory).

Themes discussed in this issue of *Semiotic Studies* testify to the recent extension of the field of philosophy of language, they offer a fresh look at some old puzzles and problems, and they include modifications to the theory of mental files, intentional identity and coordination, meaning holism and semantic minimalism, an intensional semantics for generative grammar, the ambiguous semantics of 'ought', metalinguistic value disagreement, a scalar approach to moral adjectives, the liar paradox perceived from the Wittgensteinian perspective, and linguistic relativity in relation to analytic philosophy.

As observed by François Récanati, the idea of a mental file or 'dossier' was introduced by several philosophers in connection with the referential use of definite descriptions or with identity statements (see Récanati 2012, p. vii). For Récanati, mental files are the

[3] For an early discussion of semiotics, considered as the general theory of the mind and language, see the work of Jerzy Pelc, especially Pelc (1971).

[4] This tendency is very clearly seen in companions, handbooks, and guides to philosophy of language, see García-Carpintero and Kölbel (eds.) (2012), Russell and Graff Fara (eds.) (2012), Odrowąż-Sypniewska (ed.) (2016), Hale, Wright, and Miller (eds.) (2017), to mention just four major recent publications.

vehicles of singular thoughts, or "cognitive structures which store information about entities. They are entries in the mental encyclopedia, that is, concepts" (Récanati 2016, p. vii). In the opening contribution, Mieszko Tałasiewicz sketches a modified model of mental files, connected with the debate between singularism and descriptivism. He discusses triggering mechanisms for opening files, and introduces a bipartite structure of a file. This bipartite structure combines an objectual part, encompassing what traditionally has been associated with the notion of a mental file, serving the purpose of storing information about the referent of the file, and a metadata part, serving the purpose of storing information about the file itself. Tałasiewicz demonstrates how such a structure can account for cognitive discernibility of files containing identical objectual information and grounded with the same acquaintance relations.

Hsiang-Yun Chen focuses on intentional identity and coordination. She observes that though the concept of intentional identity has aroused considerable interest since Geach's classic short paper (Geach 1967), its real import is still not fully appreciated. In her contribution she draws on three sets of data (such as intersubjective intentional identity, intrasubjective intentional identity, and cross-speaker anaphora), and provides a unified analysis of coordination that is the key to a proper understanding of intentional identity.

Filip Kawczyński attempts to rejuvenate the theory of meaning holism. In this contribution he assumes the meaning holism principle, according to which the meaning of a single expression depends on the meanings of all other expressions in a given linguistic system. He further observes that, in recent years, the philosophical reflection on language has often concentrated on the problem of the influence of context upon semantic content, and that contextualism and minimalism constitute two dominating approaches to the issue. Kawczyński offers a fresh look at the debate and demonstrates that meaning holism is compatible with minimalism (and hence far more distant from contextualism than usually assumed).

Adriano Marques da Silva devotes his contribution to the relation between Chomskyan generative grammar and semantics. He claims that in order to account for the explanatory role of syntax in the generative program it is necessary to review certain foundational assumptions commonly accepted in formal semantics. In order to do

so, he applies the intensional approach to semantics, developed over the years by Paul Pietroski (most recently in Pietroski 2017, and comprehensively in Pietroski 2005). The methodological background to the discussion is provided by the heuristics of the scientific research programme rooted in the work of Imre Lakatos.

Joanna Klimczyk is concerned with the semantics of 'ought' and the ambiguity connected with its agenetive and non-agentive senses. This contribution reviews research interconnecting linguistic semantics, deontic logic and logic of agency, and discusses more comprehensively and critically the approach advocated by Mark Schroeder (2011). The author proposes a coherent philosophical study of the meaning of 'ought' which takes into account different aspects of agentivity and authorship of approriate actions.

Erich Rast offers a fresh look at metalinguistic value disagreement. He distinguishes two meanings of general terms and value predicates: core meanings represent the lowest common denominators between speakers and they are primarily based on the needs to coordinate behaviour, on the other hand, the noumenal meanings of general terms or value predicates are intended to capture crucial aspects of reality. According to Rast, metalinguistic value disputes (similarly to other disputes about other theoretical terms) are about noumenal meanings on the basis of shared core meanings.

Federico Faroldi and Andrés Soria Ruiz analyze the scale structure of moral adjectives (such as 'good', 'bad', '(un)ethical', 'cruel', '(im) moral', 'virtuous', etc.). They provide an overview of the relevant literature in semantics (including formal semantics), pragmatics, and meta-ethics, and discuss how and whether moral adjectives fit a semantics for gradable adjectives. They also test whether moral adjectives are relative or absolute adjectives. The preliminary results suggest that moral adjectives do not fall neatly under either category, but rather they are multidimensional, relative-standard adjectives.

Jakub Gomułka and Jan Wawrzyniak offer a new analysis of the liar paradox, based on the Wittgensteinian approach to semantic and logical paradoxes. Their main aim is to point out that the liar sentence is only seemingly intelligible. In order to do so, they present the traditional solutions of the paradox and analyze their shortcomings, and claim that the liar sentence is mere nonsense: such sentences do not have any role in any language game or linguistic practice, hence they

are completely useless. The authors also discuss several possible objections to their approach.

In the final contribution, Filippo Batisti discusses linguistic relativity in relation to analytic philosophy. He first presents a brief history of linguistic relativity (distinct waves, associated with Romantic Philosophy, Sapir and Whorf, Berlin and Kay, the research conducted by John Lucy, and Analytic Philosophy, respectively). Next, following several recent accounts, he assumes that language and cognition are conceived as intrinsically social phenomena, and hence argues that relativistic effects should be investigated in social realms, and that, within a multidisciplinary approach, analytic philosophy could help with this task. Batisti also proposes an appropriate definition of the very concept of linguistic relativity, which stresses the fact that some forms of linguistic relativity involve domains that exceed individual experience, such as patterns of language-mediated social interaction, or the by-products of social reality, which is created and accessible only through language.

According to the often quoted metaphor formulated by Scott Soames "philosophy of language is (…) the midwife of the scientific study of language, and language use" (Soames 2010, p. 1); contributions to this issue clearly demonstrate that recent developments in philosophy of language provide appropriate background and tools for the study of language, knowledge, thought, and mind.

ACKNOWLEDGMENTS

Papers in this volume were read during the Fifth International Conference on Philosophy of Language and Linguistics (*PhiLang2017*) in May 2017. The conference series is organized by the Department of English and General Linguistics, University of Łódź. I wish to thank all the participants of the conference for creating a stimulating atmosphere, the reviewers of the papers for their insightful comments, the authors for participating in the project, Tadeusz Ciecierski and Andrzej Biłat for the invitation to edit this volume and their encouragement, and Krystyna Komorowska for her excellent editorial support.

REFERENCES

Daly, C. (2013). *Philosophy of Language. An Introduction*. London, etc.: Bloomsbury.

Davies, M. (2006). Foundational Issues in the Philosophy of Language. In M. Devitt and R. Hanley (Eds.), *The Blackwell Guide to the Philosophy of Language* (pp. 19–40). Oxford: Blackwell.

Devitt, M., Hanley, R. (Eds.) (2006). *The Blackwell Guide to the Philosophy of Language*. Oxford: Blackwell.

Eco, U. (1984). *Semiotics and the Philosophy of Language*. Bloomington: Indiana University Press.

García-Carpintero, M., Kölbel, M. (Eds.) (2012). *The Continuum Companion to the Philosophy of Language*. London: Continuum.

Geach, P. T. (1967). Intentional Identity. *Journal of Philosophy*, *64*(20), 627–632.

Hale, B., Wright, C. (Eds.) (1997). *A Companion to the Philosophy of Language*. Oxford: Blackwell Publishers.

Hale, B., Wright, C., Miller, A. (Eds.) (2017). *A Companion to the Philosophy of Language*. Second Edition. Oxford: Willey-Blackwell.

Kölbel, M. (2012). New Directions in the Philosophy of Language. In García-Carpintero, M. & Kölbel, M. (Eds.), *The Continuum Companion to the Philosophy of Language* (pp. 251–265). London: Continuum.

Lepore, E., Smith, B. C. (Eds.) (2006). *The Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press.

Martinich, A. P. (2009). General Introduction. In A. P. Martinich (Ed.), *Philosophy of Language. Volume 1* (pp. 1–18). London and New York: Oxford University Press.

Morris, M. (2007). *An Introduction to the Philosophy of Language*. Cambridge: Cambridge University Press.

Odrowąż-Sypniewska, J. (Ed.) (2016). *Przewodnik po filozofii języka* [*Companion to Philosophy of Language*]. Kraków: WAM.

Pelc, J. (1971). *Studies in Functional Logical Semiotics of Natural Language* (Janua linguaruarum 90). The Hague: Mouton.

Pietroski, P. (2005). *Events and Semantic Architecture*. Oxford: Oxford University Press.

Pietroski, P. (2017). Semantic Internalism. In J. McGilvray (Ed.), *The Cambridge Companion to Chomsky* (pp. 196–216). Cambridge: Cambridge University Press.

Récanati, F. (2012). *Mental Files*. Oxford: Oxford University Press.

Récanati, F. (2016). *Mental Files in Flux*. Oxford: Oxford University Press.

Russell, G., Graff Fara, D. (Eds.) (2012). *The Routledge Companion to Philosophy of Language*. New York – London: Routledge.

Schroeder, M. (2011). *Ought*, Agents and Actions. *Philosophical Review*, *120*(1), 1–41.

Searle, J. (1969). *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Soames, S. (2010). *Philosophy of Language*. Princeton and Oxford: Princeton University Press.

Stalmaszczyk, P. (2015). Kant and the Platypus and Umberto Eco (and the Philosophy of language). In A. Gałkowski (Ed.), *Potęga intelektu. Umberto Eco: Recepcja i reminiscencje w Polsce / The Power of Intellect. Umberto Eco: Reception and Reminiscences in Poland* (pp. 409–424). Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Piotr Stalmaszczyk
Chair of English and General Linguistics
University of Łódź
piotr.stamaszczyk@uni.lodz.pl

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 13–35
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.02
Rozprawa

Mieszko Tałasiewicz*

MENTAL FILES. TRIGGERING MECHANISMS, METADATA AND 'DISCERNIBILITY OF IDENTICALS'

Summary: This paper initially follows the final part of the debate between singularism and descriptivism to the point of convergence, and discusses the notion of acquaintanceless singular thought (such that there is a compromise available between hitherto competing parties). Then a sketch of a mental files model is presented. Firstly, the triggering mechanisms for opening files are discussed. Two kinds of discourse situations, acquaintance-situations (A-situations) and decoding-situations (D-situations), are identified and different triggering mechanisms are postulated for each. Secondly, a bipartite structure of a file is introduced, combining an objectual part, encompassing what traditionally has been associated with the notion of a mental file, serving the purpose of storing information about the referent of the file, and a *metadata* part, serving the purpose of storing information about *the file* itself. Being capable of encoding a variety of types of mental files, this structure is then employed to illustrate how singularity/descriptivity of the files can be manipulated (as in the case of descriptive names) and how we can account for the cognitive discernibility of files containing identical objectual information and grounded with the same acquaintance relations.

Keywords: François Récanati, singularism/descriptivism, acquaintanceless thoughts, difference solo numero, metadata

* University of Warsaw, Institute of Philosophy. E-mail: m.talasiewicz@uw.edu.pl

In 2012, François Récanati described his Mental Files framework (MF) with the explicit aim of defending Singularism against Descriptivism. However, it is arguable that MF could bridge the gap and *unite* Singularism and Descriptivism.

To characterise the notion of semantic singularity (SEMS),we might say, after Récanati, that 'whenever a thought has singular truth-conditions, the following schema holds:

(SEMS) There is an object *x* such that the thought is true with respect to an arbitrary possible world *w* if and only if, in *w*, ... *x* ...' (Récanati 2012, p. 15).[1]

Semantic descriptivism (SEMD) would express itself in schemata which embed the idea of identifying an object *x* as something satisfying a predicate *F* in the truth-conditions of a thought about object *x*. It is widely held that the closest approximation to singularism a descriptivist can obtain (in the form of the so-called Rigidified Two-Dimensional Descriptivism) is something like:

(SEMD) There is an object *x* **such that *F*(*x*)**, such that the thought is true with respect to an arbitrary possible world *w* if and only if, in *w*, ... *x* ...

These semantic schemata, however, are not quite adequate tools for expressing what is at stake in the singularism-descriptivism debate. Actual descriptivists, among whom I would count Strawson and Searle, have explicitly endorsed SEMS, not SEMD:

"In general, the contribution that a name makes to the truth conditions of statements is simply that it is used to refer to an object" (Searle 1983, p. 258).

"In an appropriate setting the name, as used, will act as an ideal or Russellian proper name" (Strawson 1974, p. 47).

Their descriptivism is epistemological rather than semantic in nature.[2] As Searle would put it, "The issue is most emphatically not

[1] (SEMS) label is mine, not Récanati's.

[2] For a complete view it might be worthwhile to consider metaphysical versions of singularism and descriptivism (or better 'particularism' and 'anti-particularism'). As an instance of metaphysical anti-particularism, Récanati takes the view that objects are bundles of properties (Récanati 2012, p. 4). Actually, I am not sure whether this should count as a metaphysical analogy of descriptivism at all. This is a stance within the controversy on what constitutes an object: is there

about whether proper names must be exhaustively analysed in completely general terms" (Searle 1983, p. 232). He would not endorse the thesis that all thoughts are general in the sense of *not* conforming to the singular schema, but would rather take a stance about what the epistemic conditions are in which the schema – the singular schema – can hold at all. What does it mean for a subject to have a singular thought, a thought as such that 'there is an object *x* such that the thought is true with respect to an arbitrary possible world *w* if and only if, in *w*, ... *x* ...'? Epistemological singularism (EPS) and epistemological descriptivism (EPD) stress different points here:

(EPS) The subject must *be in some external relation* with that object;

(EPD) The subject must *have some internal content* identifying that object (not necessarily descriptive or even verbal at all).

There is no contradiction though. Both views can be held together and can be true together: there is nothing incoherent in Strawson's view, according to which "any reflective speaker will acknowledge that he cannot have *meant* a particular item by the use of a name on a particular occasion unless he had *some* identifying *knowledge* of that item; and he could not (in general) have acquired such knowledge save by a *causal* route originating in some fact about the particular concerned" (Strawson 1974, p. 49).[3]

---

a sort of substratum needed, or would a mere bunch of properties do? Suppose we decide that a mere bunch is enough. It does not follow that we do not have singular objects. On the contrary, we have objects the easier way – singular, well defined, causally powerful objects, although identified as bundles of properties. Such objects might enter into direct acquaintance with us and thus ground the singularist epistemology.

A better example of a metaphysical view that is somehow analogous to descriptivism would be Searle's or Putnam's view that 'there is not a ready-made world': "Objects are not given to us prior to our system of representation; what counts as one object or the same object is a function of how we divide up the world. The world does not come to us already divided up into objects; we have to divide it; and how we divide it is up to our system of representation, and in that sense is up to us, even though the system is biologically, culturally, and linguistically shaped" (Searle 1983, p. 231).

[3] Récanati will eventually admit the coherence between singularism and descriptivism thus framed: "The thinker stands in a different relation to a primary content than the relation he or she stands in to secondary content. Modulo this distinction between two grasping relations. 2-D Relational Descriptivism can be saved" (Récanati 2013, p. 230).

Admittedly, there used to be a problem for descriptivists: to specify where exactly that identifying content is to be placed/stored/processed – if not in the truth-conditions of a proposition. Perhaps it was this problem that led some philosophers to maintain that there is a connection between semantic and epistemological theses and that SEMD is all an epistemological descriptivist can have in semantics, regardless of what he or she would want or declare.[4] Descriptivists were well aware of this weakness in their stance and worked hard to overcome it. Early attempts were connected with the so-called presupposition accounts, on which I will not comment here. Later, from the seventies on, they started to elaborate on a much more promising account, which is now called 'Mental Files'.

MF (as hereby further known) addresses the problem most directly and comprehensively: the required internal content is truly separated from the truth-conditions of a proposition and stored in the files. From this angle, MF turns out to be a backup for descriptivism rather than singularism: it allows for the relieving of truth-conditional semantics from the burden of encoding the required identifying content, thus for having epistemological descriptivism united with genuine semantic singularism. Both sides can get what they are after: semantically singular thoughts and mediating content for every thought (descriptive and singular alike). No wonder early developments of this framework were due to descriptivists seeking adequate concepts to express their stance properly, notably Strawson (1974).

Yet Récanati's intuition that MF would facilitate singularism would have its merits, too. One of the problems is whether epistemological singularity (EPS) – consistent, as we have seen, with epistemological 'descriptivity' but stronger than it – is really a necessary condition for the availability of semantic singularity (SEMS). Or, in other words, whether there are acquaintanceless singular thoughts.[5]

---

[4] Arguably, SEMD entails SEMS: it is a stronger condition. The phrase '(SEMD) is all a descriptivist can have' means that he or she cannot build the semantics of singular expressions on weaker grounds (thanks to an anonymous referee for bringing this point to my attention).

[5] Nota bene: this has nothing to do with the singularism/descriptivism debate: the possibility of a singular thought about an object without actual acquaintance with that object is equally doubtful for singularists and descriptivists. That it is

In principle Récanati's answer is no (let us call it 'Acquaintance Thesis' (AT)):

(AT) '1. The subject cannot entertain a singular thought about an object *a* without possessing, and exercising, a mental file whose referent is *a*.

2. To possess and exercise a mental file whose referent is *a* the subject must stand in some acquaintance relation to *a*' (Récanati 2012, p. 155).

However, there is a recurrent intuition that in some cases – e.g. in the case of the so-called descriptive names – there is at least some air of singularity in thoughts containing such names despite the absence of actual acquaintance with the names' referents.[6] How can we account for it? Récanati, as it turns out, would not take AT at face value, but announce that "this is a normative claim" (*ibidem*, p. 156).

Normativity has many faces though. What exactly does it mean in this case? There is a moral/legal sense of normativity, for instance. Something may happen against the (moral) law. In this sense one can possess and exercise a mental file without an acquaintance, just like one can exercise someone's property without the owner's consent, only it is immoral or illegal. Quite obviously it is not the intended notion of normativity.

Another sense is connected with definitions. In this case norms determine certain identity conditions of some social or linguistic entities by specifying what is required for an object to belong to a certain kind (or to satisfy a certain sortal predicate). According to this notion of normativity, what does not conform to a norm does not exist, under a given sortal. In this sense – if AT is right – we cannot possess and exercise a mental file without an acquaintance (perhaps something else appears in our heads instead, but not a mental file).

---

doubtful for singularists is no surprise. But it is also doubtful for descriptivists. As I have stressed, descriptivists can discriminate between singular and general thoughts and they can adopt acquaintance as a criterion for this distinction, if they consider this the right thing to do. Thus, the problem of acquaintanceless singular thought is a problem of adequate expression of the notion of singular thought rather than a controversy between singularism and descriptivism. This is an internal problem of singularism, a problem of a relation between semantic and epistemological versions of this stance.

[6] See e.g. Reimer 2004; Jeshion 2004; Kanterian 2009.

This time it is not at all clear that it is not the intended notion of normativity. Récanati approvingly quotes Zeno Vendler, saying that "the fact that a tool can be misused does not alter the function of the tool" (*ibidem,* p. 158), but this stance has its limitations, namely the tool must be identifiable independently of the role it actually plays. A rake remains a rake when used as a weapon in a neighbourhood quarrel only because we can provide a definition of a rake, independent of the function of raking (e.g. an ostensive one – a rake is something like *this*). Récanati does not give us a normal definition, but he defines mental files through meaning postulates: 'Mental files are based on what Lewis calls "acquaintance relations", "The characteristic feature of the relations on which mental files are based, and which determine their reference, is that they are *epistemically rewarding*" (*ibidem,* p. 34–35). Actually, these sound very much like the norms specifying conditions of identity and if something does not fulfil these norms it excludes itself from under the sortal 'mental file'.[7]

We can move on only if we decide to weaken the meaning postulates for 'mental files' so that they would express the notion that "a mental file serves as a storehouse of information that the subject takes (consciously or not) to be about the same object" (or something like that). The stronger phrases, stating that mental files are based upon acquaintance relations, are no longer meaning postulates, so it is possible now to deny factual acquaintance without the file no longer being a file.

Yet, in the 2012 exposition of Récanati's account, without the acquaintance, we can have at most what he would call a 'singular thought-vehicle'. He would insist that "opening a mental file itself is *not* sufficient to entertain a singular thought (in the sense of thought-content)" (Récanati 2012, p. 164). And in such a case "no semantically evaluable thought is expressed" (*ibidem*, p. 160).

However, there are good reasons to believe that some sort of thought-content should be allowed in acquaintanceless cases. As Tim Crane noticed, it is hard to maintain that Le Verrier, when thinking about Vulcan, "was not thinking anything, merely airing an empty 'vehicle'" (Crane 2011, p. 39). Besides, it is not easy to tell which cases are acquaintance-less and which acquaintance-full: "There are

[7] Similar worries are expressed in Pagin 2013.

many distinctions we can make between kinds of knowledge, and the thoughts we have about things do vary depending on the ways we know these things. However, I am skeptical that any specific way of knowing something lines up systematically with thinking about a particular object as such" (Crane 2011, p. 29). This borderline becomes even more blurred when we consider that there are many kinds of objects we can think about in a singular mode, such that "it is at least controversial whether all these 'things' [...] enter into any serious causal relations at all, much less the special (ER) ones that Récanati needs to sustain 'acquaintance'" (Hansen, Rey 2016, p. 428).[8]

Récanati in 2012 tried to evade such charges by saying that even if there is no singular thought-content, "this does not mean that the user of the singular term is not thinking anything: there are other thoughts in the vicinity, which the subject is arguably entertaining" (*ibidem*, p. 160). He would namely ascribe *'derived, metarepresentational'* function to these thoughts (*ibidem*, p. 177), as is the case in modelling someone's propositional attitudes (with the use of the so-called 'vicarious files'). It does not seem convincing though: "people don't normally regard words and rainbows as involving deliberate pretense or metarepresentation" (Hansen, Rey 2016, p. 430). And what is more important, the *purpose* of keeping such a strict connection between singular thought-content and actual acquaintance, while allowing for acquaintanceless singular thought-vehicles, seems more and more evasive.

> Singular vehicles […] are merely *taken* to provide singular reference by those who entertain them […]; at best, they are *treated* as providing singular reference (we may imagine a cautious scientist who is not sure about the existence of the entity she is naming). In each case, theirs is only an *appearance* of singularity and it is not clear how one could go from an appearance of singular thought to singular thought proper, in any interesting semantic sense. So if entertaining a singular vehicle comes down to entertaining a *seemingly singular thought* (which is really not a singular thought in any interesting semantic sense), we do not see how this notion could be of help. (Coliva, Belleri 2013, p. 110)

For "what matters is not that the thought happens to refer to just one thing, but that it has a specific *cognitive role*. Singularity is a matter

[8] Among such an object we might count species, performances, ceremonies, marriages, contracts, companies, stores, clubs, galaxies, black holes, the sky, the wind, the rain, the tide, ocean waves, shadows, reflections, rainbows etc.

of the cognitive – that is, the psychological or phenomenological – role of the thought" (Crane 2011, p. 25).[9]

Eventually, Récanati in 2013 eases his constraint and admits that acquaintanceless thought may have singular *content* albeit only a primary one, as opposed to a secondary (in a two-dimensional sense), which is still lacking (Récanati 2013, p. 208).[10] Accordingly, he reinterprets his distinction of primary and derived functions of the files.[11] While in 2012 primary functions are representational and derived functions are metarepresentational (Récanati 2012, p. 177), in 2013 Récanati says that "'primary' here must be understood in a new, evolutionary sense" (Récanati 2013, p. 211–212): primary functions are 'evolutionary basic' and derived functions are those which have evolved as distinct from them – which no longer carries anything like the presumption of their being 'metarepresentational'.

---

[9] There is a general worry about grounding semantic issues on the swampy grounds of metaphysics. We are accustomed to thinking that when we substitute the name 'Neptune' for a description 'the perturber of the orbit of Uranus', we only *presume* there is some object causally responsible for our observations (sense data) – ready to be so named – and we are accustomed to thinking that in such a case eventually we may be wrong, as in the case of Vulcan. But, actually, from a general enough point of view, things are quite analogous when it comes to naming directly perceived objects: if Pedro Calderon de la Barca were right, no thought can be truly singular, for 'Life is a dream'.

Such is the moral from the skeptic's challenge. Perhaps we cannot tell the external world from a highly sophisticated set of internal data. But the right answer is that we do not need to bother. We shall act as if the world existed and we shall *talk* as if the things we are talking about existed, too. And that is why we do not need to distinguish cognitively singular thoughts (singular vehicles) from 'really' singular thoughts (accompanied by actual acquaintance), while no one but God knows which are which. What is important is which thoughts are *meant* to be singular and which are *meant* to be descriptive. And these are matters of vehicles, not contents.

[10] There is a question about the truth value of such thoughts. In 2012 (p. 164) Récanati would hold that "if reference is not achieved, no singular truth-condition is determined and the thought cannot be evaluated as true or false." But this is just one option of many. Some accounts of truth-conditions may render the proposition in question false, some others as lacking truth value, and yet others perhaps even as true under certain additional conditions (such as the condition that the descriptive content of the predicate is wholly contained in the body of the non-referential file associated with the subject of the sentence – as in the case of 'Sherlock Holmes lived in London').

[11] Mind the ambiguity of the word 'primary' here: primary content vs. secondary content and primary function vs. derived function.

This, I presume, would end the debate. The compromise seems satisfactory to both sides. Along with the unification of singularism and descriptivism and the agreement about acquaintanceless singular thoughts we have quite a bit of consensus in the whole area, as acknowledged by Récanati himself: "I do not think I ever argued for the absolute untenability of the other approaches [...] I believe that, in the end, there is a true convergence" (Récanati 2013, p. 228).

Once we relieve MF of fighting unnecessary battles against a misconstrued enemy, we can move on to raise an array of interesting questions about MF and about the use of MF in semantics.[12] These are quite distinct topics: how to model semantic phenomena using mental files and how to model mental files themselves.

It is widely held – among people entertaining the whole idea of mental files – that mental files are mental particulars. Mental files are commonly regarded as real cognitive entities, relatively independent of semantic issues. For if they were just illustrations of what is going on in semantics, they could not play an explanatory role for semantic phenomena, under the charge of circularity or question-begging. As James Pryor would put it: "I'm assuming that for some explanatory purposes we want to model facts about […] thinking" (Pryor 2016, p. 321)**.**

A caveat is needed perhaps: mental files are *theoretical objects* rather than empirical ones, much like quarks or strings are theoretical objects of physics: we cannot actually see them, but assuming their existence are out there and having certain properties explains the data within certain theories. Accordingly, assuming the existence of mental

[12] It should be noted that semantics here is taken in a quite broad sense in which it might be understood as a theory of utterance meaning or intended meaning, making use of cognitively real representations and reflecting mental processing of natural language utterances. As with any theory, it must be focused on general mechanisms and role-models rather than particular neuronal activity of flesh and blood speakers and hearers, yet it is arguably less detached from a human cognitive system than some old Frege-style antipsychological objective semantics, devoid of any cognitive aspects whatsoever. Such a broad sense is employed e.g. in Katarzyna Jaszczolt's Default Semantics, where compositionality is meant to be working on the level of merger representations, which are entities unifying elements of syntax, lexicon, world-knowledge, cultural and social stereotypes, inferential patterns and discourse situation. Such semantics "brings truth-conditional methods closer to cognitive, conceptual analyses" (Jaszczolt 2010).

files *in our minds* and their having certain properties might explain certain phenomena in semantics, while assuming their existence and their having some other properties would help to explain certain phenomena in psychology. Perhaps assuming their existence and their having different properties would contribute to explanations in neurophysiology (of which I am ignorant). These different theoretical perspectives can draw significantly different pictures of what mental files are. Outcomes of one discipline might back up or constrain the pictures in another – which is a customary gain to be expected from interdisciplinary research – yet there is no need to expect, until a universal theory of everything is reached, that these pictures shall be at all points commensurable. And it definitely cannot be demanded that a semanticist shall not speak about mental files (as cognitive entities) unless psychology or neurophysiology have fully established the 'real' nature of the files. Semantics is a part of cognitive science – on a par with philosophy, psychology, neurophysiology, informatics etc. – and is fully entitled to postulate (tentatively, as always) certain cognitive entities as having such and such properties.

Eventually, there is a sort of ambiguity in the use of the notion of mental files (as always when theoretical objects are concerned). In one sense, mental files are real entities in our heads, something we have only limited access to and something we try to roughly and tentatively model, from different perspectives. In another sense, mental files are proposed models of these entities, capturing some of the perspectives on these objects but not necessarily all. In this sense it is reasonable to ask if certain perspectives, or certain aspects of our mental processes, can be modelled better in a different way, for instance in a framework of so-called mental graphs.[13] Those different models are not competing, though, as long as they model different aspects or give the same explanations in areas where they overlap.

In what follows, I will present a sketch of a mental-file-style (as opposed to mental-graph-style) model of cognitive entities called 'mental files' and discuss some of its consequences.

As cognitive entities, mental files need to be formed or activated by some triggering cognitive mechanisms. Récanati seems to maintain that it is a matter of a *conscious decision* by the thinker to open a mental

[13] Pryor 2016.

file, a decision that could be justified and reasonable – or not: "The only *reason* to open a file in [acquaintanceless] cases is that the user expects to stand in the appropriate relation to the referent" (Récanati 2012, p. 158); "the (expected) existence of an information link is what *justifies* opening a file" (*ibidem*, p. 167). However, as soon as we realise that mental files are meant to function, *inter alia,* as a part of tracking systems for visually detected objects (Murez, Récanati 2016, p. 266) we will be bound to reject the idea that mental files as such belong solely to the sphere of the subject's *conscious* reflection. There are not only *reasons* to open files, there are also *causes*. Files open – or do not – whether we want them to or not. By careful examination of our semantics we can *discover* the Revolutions of the Files, their capacities or tendencies for merging, linking, sharing, etc., but we do not have full access to *manipulating* the files, not directly. Arguably, Récanati is bound to take such an anti-psychologist stance himself, as it is an important part in his argumentation against the circularity objection: "[clustering] It may be entirely a matter of subpersonal binding of information. Thus in the case of proto-files at least it is the cognitive system, not the subject, that takes the pieces of information to concern the same object and cluster them within a file" (Récanati 2012, p. 98).

Definitely, one of the most prominent triggering mechanisms for opening the files would be entering into an acquaintance relation with an external object. However, since effectively an acquaintance with an object is *not* necessary to open *de facto* a mental file, we face an interesting problem: what else, if not solely entering into an acquaintance relation, is a triggering mechanism for opening mental files?

First, let us note that in fact we may find ourselves in two different kinds of discourse situations: situations where the object is given and the use of a word is related to it (acquaintance situations, A-situations), and situations where the word is given and the corresponding object has to be identified (decoding situations, D-situations). I have elaborated this issue a bit more in Tałasiewicz (2010).[14] It may seem similar to the speaker-hearer distinction but it is not. Indeed, usually speakers are in A-situations and hearers are in D-situations. But all combinations are possible. Both speaker and hearer can be in an A-situation

[14] In that paper 'D' in 'D-situations' stands for 'descriptive' – which I consider no longer appropriate, as we can decode singular terms in D-situations.

(when they share actual acquaintance with the object being referred to) and both can be in a D-situation (when they both lack such acquaintance). Moreover, in certain circumstances – usually artificially prepared – speakers can be in D-situations and hearers in A-situations. This is the case e.g. in 'Blind Date' type shows, where the speaker speaks about someone he or she cannot see while the audience can see the person very well. Such situations are odd and sometimes funny, but they can and do happen.

What is essential is that there are different cognitive mechanisms involved in language processing in respective situations, especially different triggering mechanisms. When we get, say, a visual stimulus, which usually means that we are entering into an acquaintance relation with some object, and are about to react linguistically to it, we open or activate a mental file for this object. But when we just hear an utterance, without any trace of personal acquaintance with the objects the utterance is about, we do open the files for them, too, but due to a completely different triggering mechanism.

A promising way to search for such a mechanism is to accommodate some of Strawson's ideas according to which mental files open not only with the prospect of a singular thought, however such thought is conceived, but along with any nominal use whatsoever (Strawson 1974, p. 35–60). According to Strawson, a 'nominal use' is a syntactic notion rather than semantic. The very same expression, e.g. some definite description, might be used nominally, as in 'The tallest mountain in Europe is more than 4000 meters high', where it is the subject of the sentence, or ascriptively/predicatively as in 'Mont Blanc is the tallest mountain in Europe', where it is a part of a predicate.[15] Arguably, a nominal use – a referring use in Strawson's terminology[16]– can be

[15] This old distinction is accommodated in some quite new developments in MF, for instance by Josef Perner and his colleagues: "Files capture the predicative structure of language and thought: the distinction between what one is thinking/talking of [...] (*individuating* information) and what one thinks/says about it [...] (*predicative* information)" (Perner, Huemer, Leahy 2015, p. 78–79).

[16] Note that Strawson's ascriptive/predicative use, sometimes even called by him 'attributive' (Strawson 1950, p. 13), is something entirely different from Donnellan's much more recognised attributive use. Donnellan alludes to Strawson's notion in the following passage: "There are some uses of definite descriptions which carry neither any hint of a referential use nor any presupposition or implication that something fits the description. In general, it seems, these are recogniz-

modelled as belonging to the category of names as construed in classical Categorial Grammar.[17] What triggers the opening of a file would be thus just a grammatical processing of the sentence conducted by the 'syntactic module' of our cognitive apparatus aimed at making space for further information, the need for which the module sort of 'anticipates' from 'the frame of the sentence'.

That would require further differentiation among the files in order to account for undeniable differences between singular and general semantics. A syntactic module would produce mental files for general thoughts as well as for singular ones, just as general expressions as well as singular ones might appear in a subject position in sentences. It would be convenient to call the files themselves general and singular, respectively. And it would be reasonable to stipulate that the difference between singular thought and descriptive thought will be somehow represented in the structure of respective files.

Another matter convenient for representation in the structure of the file, is information about the file. Normally, the metaphor of a file evokes a picture of a catalogue card divided into two sections: the heading or the label of the file and the body of the file, where the information about the referent of the file is stored. According to my present proposal, the files should contain information about objects as well as information about *themselves* (metadata). Thus, they should consist of *two* parts, both having a heading and a body of information. A heading – as in a real card file in a library catalogue – serves the purpose of identifying a given body of information (we can have, in principle, many different bodies containing the same information under different headings). The heading of the metadata

---

able from the sentence frame in which the description occurs. These uses will not interest us" (Donnellan 1966, p. 363). Donnellan's referential/attributive distinction is a refinement of Strawson's referring use only. Thus, both referential and attributive uses in Donnellan's sense belong to what I call 'nominal use'.

[17] More on that in Tałasiewicz (2014)**.** Interestingly, through Categorial Grammar the MF framework would get related to a Husserlian notion of 'intentional object' and thus help to clear up the entanglements of the theory of intentionality and exorcise the spectre of Meinong from the debate. Récanati himself took some preliminary steps in this direction in his discussion of the medieval witchcraft of Peter T. Geach (Récanati 2012, p. 204–205; Geach 1967). Following Récanati, I will leave this point without elaboration, as 'orthogonal' to the main line of argument in the paper.

part of the file is a sort of label for the whole file, an image carried through a visual stimulus or a particular wording of a heard sentence that triggered the opening of the file (depending on the triggering mechanism involved). The body of the metadata part of the file would contain some information about opening circumstances, linking and merging history that establishes the internal structure of one's system of the files, indexing/embedding for reporting propositional attitudes, which requires representing other people's files in one's mind, etc. The body of the objectual part of the file is the main storage compartment; it contains all information one has about the object represented by the file, while the heading of the objectual part represents only the reference-fixing information about the file (Fig. 1). Only the content of this heading – not the whole body of information – enters the truth-conditions schema for the thought employing the file.

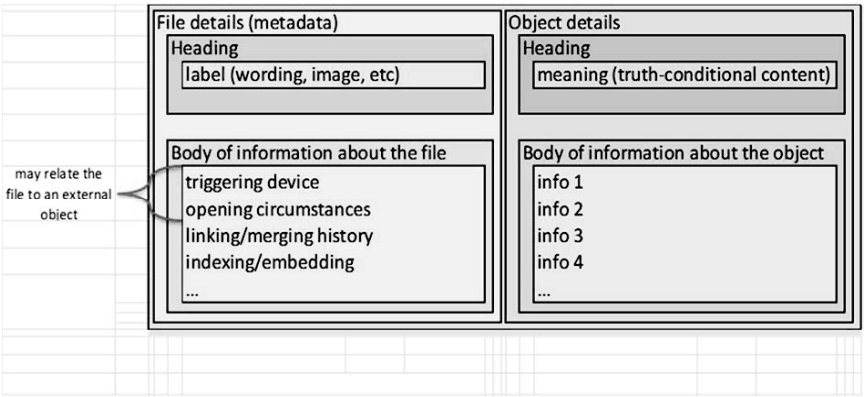

Fig.1. The structure of a file.

The important thing about singular thoughts (Figs. 2, 3) is that no identifying content about the object of reference enters the truth-conditional scheme. We will represent this by leaving the objectual heading of singular files just empty. 'No' means 'no'.[18] But note that

[18] Arguably this is true only about the so-called Spelke-objects: objects individuated by cognitive mechanisms analysing three-dimensional patterns of surface motions. According to Elisabeth Spelke such processes "are not overlaid and obscured by processes for recognizing objects of a multitude of kinds" (Spelke 1990, p. 30) and are crucial in early infant development. However, although such mechanisms are not limited to infancy (as Spelke mentions, "development enriches

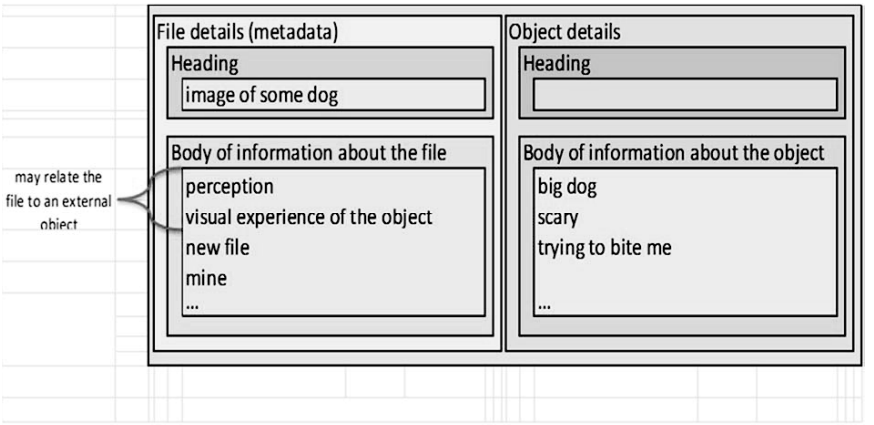

Fig. 2. Singular file triggered by perception.

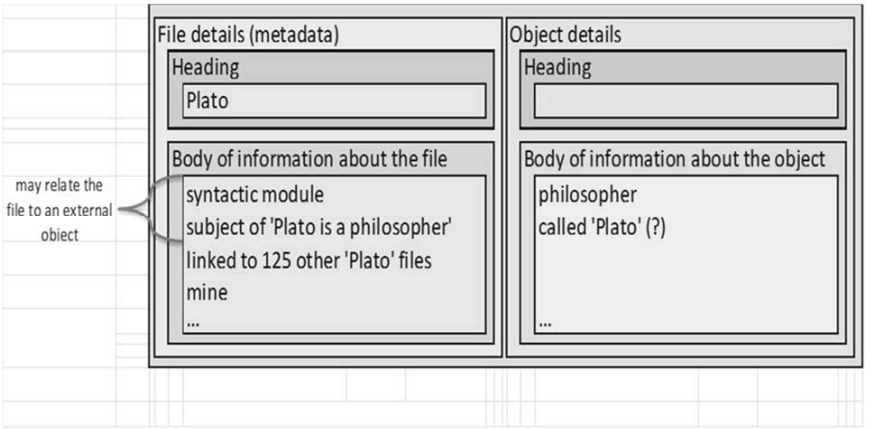

Fig. 3. Singular file triggered by syntactic module.

object perception without fundamentally changing it" so that adults may use such mechanisms too, among others), more sophisticated individuation of objects has to be connected with some categorisation, so that "a mental file for an observed object always incorporates information about as-what an object under discussion is being individuated" (Perner, Leahy 2016, p. 497). If so, it is reasonable to follow Perner's and Leahy's suggestion that the headings in non-Spelke singular files are not empty, but rather contain a sortal expression, under which the referents of the files are to be individuated. Such files are still different from descriptive files, whose objectual headings contain not just sortals but uniquely identifying descriptions. For the sake of simplicity, I will refrain from showing this variation on figures in the text [thanks to a referee for drawing my attention to this issue].

this would not dissolve the identity of the file itself, safely preserved by the metadata part of it (including the history of opening which may – or may not, on which I will comment below – relate the file causally to the object of reference).[19]

Let us compare this with two different kinds of descriptive files, for plain and rigidified descriptions respectively (Figs. 4, 5).

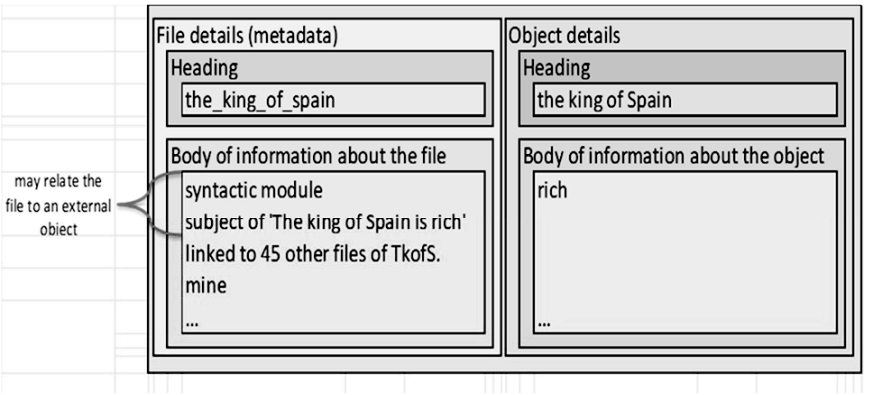

Fig. 4. Descriptive file for plain description.

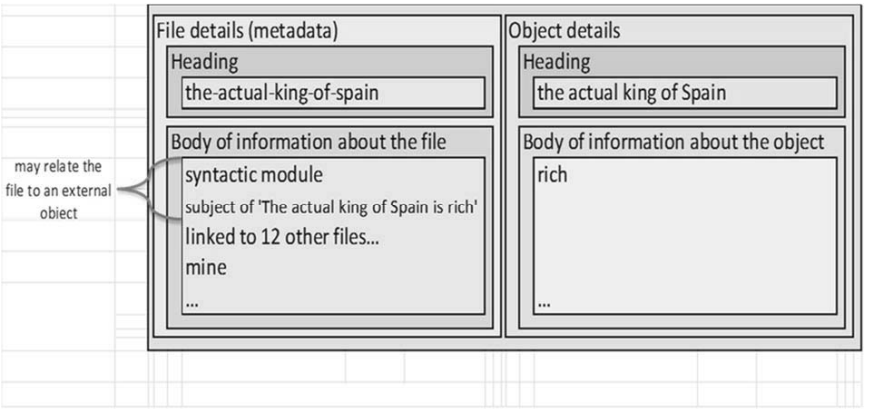

Fig. 5. Descriptive file for rigidified description.

[19] A triggering device plus opening circumstances would constitute what we might call a 'information channel' as in Azzouni (2011).

As we can see, Récanati and others were right to say that a so-called rigidified description is still a description – mere rigidifying would not make the thought singular. There is still a non-empty heading in the objectual part of the file.[20]

Now, we can have a look at how this bipartite model of mental files can help us to account for certain semantic phenomena. First, let us go back for a moment to the problem of singularity.

Technically we can fit singular files without acquaintance into our model very easily. Singular files are those files which were opened as such, with an empty objectual heading. A triggering mechanism decides whether the file is singular or not and produces accordingly

[20] However, there is something misleading in Récanati's account of this. He seems to maintain that the sign of descriptive character of a phrase is the possibility to accompany it with 'whatever it is' or something in such gist. He says: "a rigid use need not be referential: it may be attributive. Thus I may say: The actual *F*, whoever he is, is *G*" (Récanati 2012, p. 18). Well, I may say: Jean-Luc Mélenchon, whoever he is, is... well... is G. This would not make the name 'Jean-Luc Mélenchon' attributive. It just makes explicit my ignorance of who Jean-Luc Mélenchon is. I think there is a deeper issue here, namely a peculiar double-mindedness with respect to what counts as the actual world. The rigidifying of a description may come in two aspects: formal and ontological. Formally, rigidifying consists in choosing one element from the class of worlds, dubbing it w_0 or so, and relating the description to this element. We do not need to know anything about this w_0, it might be a mere abstraction for us. We know nothing about whether this element in the class of worlds is *really* our world. It has no ontological criterion of identity; it is identified only as 'the world we've chosen for w_0'. In fact, we can 'quantify over worlds playing the role of the actual world', with respect to which Martin Davies and Lloyd Humberstone introduced the operator 'Fixedly' (Davies 2006, p. 143). In cases where 'actual world' is a role that many actors can play, it makes sense to say: 'in an actual world, whichever it is,' and the statement has the air of a general thought, indeed. But there is another interpretation, an ontological one, of the 'actual' operator according to which 'actual world' is not *a role* but a metaphysical *reality*. In such cases there is nothing *descriptifying* in adding 'whichever it is', any more than there is in the phrase 'you, whoever you are'. There are some affinities between the distinction between the ontological and formal notion of actuality introduced here and the distinction between deep and superficial contingency discussed by Gareth Evans. Deep contingency is a metaphysical matter: 'whether a statement is deeply contingent depends upon what makes it true', whereas superficial contingency is a formal matter: "whether a statement is superficially contingent depends upon how it embeds inside the scope of modal operators" (Evans 2006, p. 179). Let us postpone closer elaboration of these affinities to another occasion, though. All we need here is to elicit some general intuition of this sort of distinction.

an appropriate structure of the file. Normally we expect that the mechanism would be activated by some epistemically rewarding relation with the object, such that we might take it as acquaintance, but this is not a condition *sine qua non*. Both syntactic module and perception alike, as triggering mechanisms, may produce some errors as to whether there is a relevant relation substantiating the opening of a singular file. Suppose, for instance, that I hear an utterance from behind the wall: 'Be silent, Michael is sleeping' and I open a file for this Michael. It is possible, that in fact there is a woman behind the wall, talking to me, and passing to me, through this act of communication, her causal connection to the said Michael (causal-historical singularity). But it might well be that there is a monkey behind the wall, tapping randomly on the keyboard of a speech-synthesiser and producing in this way the utterance. The monkey cannot pass to me any acquaintance with Michael because there is none. There is no Michael at all in that situation. Yet, cognitively, nothing is different for me (provided I do not know which is the case). I open a mental file for Michael anyway.

There are plenty of such possibilities for error: descriptions mistakenly taken as names,[21] discourse anaphora mistakenly taken as a demonstrative,[22] or hallucinations (fake visual triggering). And since there can be acquaintanceless singular files due to some errors, we can deliberately open such files if there is a need for them, as in the case of descriptive names. All we need is to have a regular descriptive file to start with. We can duplicate this file, change headings, and link the new one to the old one (Fig. 6).

Another matter that might be nicely accounted for in our model is the question of whether there can be any difference in cognitive significance between two mental files without any semantic difference between them. The problem was posed in Pagin (2013), who

[21] '[MrTumnus] "And you – you have come from the wild woods of the west?"
[Lucy] "I – I got in through the wardrobe in the spare room."
[MrTumnus] "Daughter of Eve from the far land of Spare Room where eternal summer reigns around the bright city of War Drobe, how would it be if you came and had tea with me?"'

C.S. Lewis, *The Chronicles of Narnia*

[22] "This man is the king" [as a part of "At the court there are many people. Sometimes you may meet a man wearing a crown. This man is the king. Bow to him"].

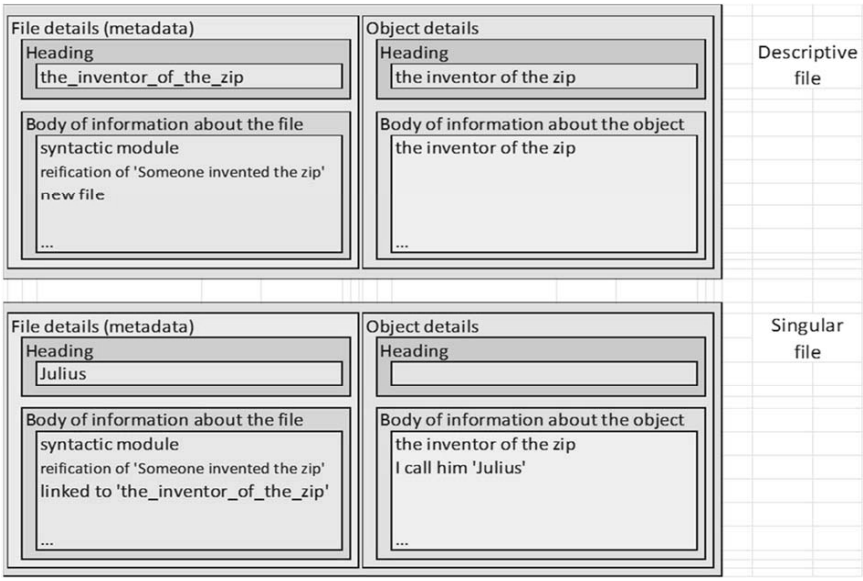

Fig. 6. Files for descriptive names.

maintained that mental file theorists would have problems accounting for such differences. He gave the following scenario as an example: "Suppose *X* takes herself to see two moths flying around in her kitchen. She opens a file for each, alpha and beta, thinking of them as '*A*' and '*B*' respectively […]. In the case the subject in fact is mistaken, and there is only *one* moth causing the opening of both files, there does not seem to be any difference between acquaintance relations of alpha and beta" (Pagin 2013, p. 140). Récanati replied that: "two files can differ in cognitive significance *solo numero*" (Récanati 2013, p. 214). However, as Aidan Gray argued, maintaining that merely a numerical difference affects cognitive significance of the files is hardly a plausible option:

> If names *a* and *b* are of the same generic type, they are both articulated with some sign*s*. This means that a hearer who is competent with both names will need to disambiguate occurrences of *s* as being articulations of either *a* or *b*. A hearer who associates no different information with names *a* and *b* will have no reason to interpret an occurrence of *s* as an articulation of *a* rather than as an articulation of *b* […]. [D]isambiguation requires differential information. (Gray 2016, p. 354)

The present account would resolve this dispute in the following way: two files can be identical in their objectual parts, which contain information about objects and contribute to the truth conditions of the thoughts based upon the files, yet they are different not solely numerically, but also in their metadata parts. For the channel of information of the file – contrary to what Pagin says (p. 137) – is not constituted merely by acquaintance relations (which are identical in the moth scenario) but rather by the triggering mechanism and opening circumstances of the file – and the latter are different in the moth scenario: the subject would not have opened two separate files for moth *A* and moth *B* if she had not seen a moth (at least) twice in two different circumstances.[23] Thus we can have two files based upon acquaintance with the same object, containing the same information about the object, and yet cognitively different due to differences in information about *the files themselves*.

There are many more issues that the present account may facilitate approaching, but we have to postpone their discussion to another occasion. Let us take stock now of what we have obtained so far.

First, we have followed the final part of the debate between singularism and descriptivism, witnessed a sort of 'convergence', as Récanati described it, and reached the point at which there is not much left for serious quarrel. Then we have examined the problem of acquaintanceless singular thought. This debate is quite convergent too, and at this present stage seems to be a promising candidate for a good compromise: a thought without acquaintance might be singular in the sense of vehicle *and* in the sense of primary content while lacking secondary, referential content. Since the notion of primary content regarding files, not expressions, is a bit metaphorical, there are still some doubts, as expressed by Pagin (2013, p. 142–144), about how exactly we shall understand this singular acquaintanceless primary content. Récanati's statement that "primary content must be cashed out in terms of functions or roles" (2013, p. 231) would not resolve these doubts.

At this point I have proposed a sketch of a model of mental files designed to handle some of their important features. First, since there

[23] Thus Pagin is right that it is difficult to have distinct acquaintance relations without distinct relations (Pagin 2013, p. 141) but we do not need distinct relations to have distinct information channels.

can be acquaintanceless files and a conscious decision is not always available to initiate the opening of the files, we need some triggering mechanisms for opening files. I have distinguished between two kinds of discourse situations: acquaintance-situations (A-situations) and decoding-situations (D-situations) and posited a syntactic module as the triggering mechanism for opening the files. Then a bipartite structure of a file has been introduced, combining an objectual part (consisting in an objectual heading, encoding the truth-conditional import of the file, and an objectual body of information or misinformation about the referent of the file) and a metadata part (consisting of the main heading of the file and a body of meta information about the file itself, notably about the triggering mechanism responsible for opening the file and about the circumstances of opening the file, as well as linking/merging history, indexing etc.). This structure is capable of encoding a variety of types of mental files, which is then employed to illustrate how singularity/descriptivity of the files can be manipulated (as in the case of descriptive names) and how we can account for cognitively distinct files containing identical objectual information and grounded with the same acquaintance relations.

REFERENCES

Azzouni, J. (2011). Singular Thoughts (Object-Directed Thoughts). *Proceedings of the Aristotelian Society Supplementary*, *LXXXXV,* 45–61.

Coliva, A., Belleri, D. (2013). Some Observations on Fracois Récanati's Mental Files. *Disputatio*, *V*(36), 107–118.

Crane, T. (2011). The Singularity of Singular Thought. *Proceedings of the Aristotelian Society Supplementary*, *LXXXXV*, 21–43.

Davies, M. (2006). Reference, Contingency, and the Two-Dimensional Framework. In M. García-Carpintero and J. Maciá (Eds.), *Two-Dimensional Semantics* (pp. 141–175). New York: Oxford University Press.

Donnellan, K. S. (1966). Reference and Definite Descriptions. In P. Ludlow (Ed.), *Readings in the Philosophy of Language* (pp. 361–381). Cambridge MA: The MIT Press 1997.

Evans, G. (2006). Comment on 'Two Notions of Necessity'. In M. Garcia-Carpintero and J. Maciá (Eds.), *Two-Dimensional Semantics* (pp. 141–175). New York: Oxford University Press.

Geach, P. T. (1967). Intentional Identity. *Journal of Philosophy*, *74*, 627–632.

Gray, A. (2016). Minimal Descriptivism. *Review of Philosophy and Psychology*, *7*, 343–364.

Hansen, C., Rey, G. (2016). Files and Singular Thoughts Without Objects of Acquaintance: The Prospects of Récanati's (and Others') 'Actualism'. *Review of Philosophy and Psychology*, *7*, 421–436.

Jaszczolt, K. (2010). Default Semantics. In B. Heine and H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*, Oxford handbooks in linguistics. Oxford–New York: Oxford University Press.

Jeshion, R. (2004). Descriptive Descriptive Names. In M. Reimer and A. Bezuidenhout (Eds.), *Descriptions and Beyond*. Oxford–New York: Oxford University Press, pp.591–612.

Kanterian, E. (2009). Puzzles About Descriptive Names. *Linguistics and Philosophy*, *32*, 409–428.

Murez, M., Récanati, F. (2016). Mental Files: an Introduction. *Review of Philosophy and Psychology*, *7*, 265–281.

Pagin, P. (2013). The Cognitive Significance of Mental Files. *Disputatio*, *5*(36), 133–145.

Perner, J., Huemer, M., Leahy, B. (2015). Mental Files and Belief: A Cognitive Theory of how Children Represent Belief and Its Intensionality. *Cognition*, *145*, 77–88.

Perner, J., Leahy, B. (2016). Mental Files in Development: Dual Naming, False Belief, Identity and Intensionality. *Review of Philosophy and Psychology*, 7, 491–508.

Pryor, J. (2016). Mental Graphs. *Review of Philosophy and Psychology*, 7, 309–341.

Récanati, F. (2012). *Mental Files*. Oxford: Oxford University Press.

Récanati, F. (2013). Mental Files: Replies to my Critics. *Disputatio*, *V*(36), 207–242.

Reimer, M. (2004). Descriptively Introduced Names. In M. Reimer and A. Bezuidenhout (Eds.), *Descriptions and Beyond* (pp. 613–629). Oxford–New York: Oxford University Press.

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge–New York: Cambridge University Press.

Spelke, E. (1990). Principles of Object Perception. *Cognitive Science*, *14*, 29–56.

Strawson, P. F. (1950). On Referring. In P. Ludlow (Ed.), *Readings in the Philosophy of Language* (pp. 335–359). Cambridge MA: The MIT Press.

Strawson, P. F. (1974). *Subject and Predicate in Logic and Grammar*. University paperbacks; 538. London: [New York]: Methuen; distributed by Harper & Row, Barnes & Noble Import Division.

Tałasiewicz, M. (2010). Asymmetrical Semantics. In P. Stalmaszczyk (Ed.), *Philosophy of Language and Linguistics* (pp. 329–345). Frankfurt a. M.: OntosVerlag.

Tałasiewicz, M. (2014). Categorial Grammar and the Foundations of the Philosophy of Language. In P. Stalmaszczyk (Ed.), *Philosophy of Language and Linguistics. The Legacy of Frege, Russell and Wittgenstein*, Philosophische Analyse/ Philosophical Analysis. Berlin, New York: de Gruyter.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 35–51
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.03
Rozprawa

Hsiang-Yun Chen*

INTENTIONAL IDENTITY AND COORDINATION

Summary: The concept of intentional identity has aroused considerable interests since Geach (1967). I argue, however, that the real import of intentional identity is still not duly appreciated. Drawing on three sets of close-knit data – intersubjective and intrasubjective intentional identity, along with cross-speaker anaphora, I submit coordination as the key to its proper understanding and propose a set of success conditions thereof.

Keywords: anaphora, attitude ascription, externalism, intentional identity

1 INTRODUCTION

According to Geach (1967), the sentence "Hob believes a witch blighted Bobs mare, and Nob believes that she killed Cobs sow" is true given the relevant story. It is not prima facie obvious, however, how to analyze the Hob-Nob sentence using standard semantic tools: altering the scope of the existential quantifier results in either a *de re* or a *de dicto* reading; the former leads to a dubious ontological commitment, while the latter misconstrues Hob's and Nob's respective beliefs. As a result, the puzzle is commonly reckoned as a logical conundrum.

Nevertheless, intentional identity has a wide coverage. As Geach himself notes explicitly, "[w]e have intentional identity when a number of people, or one person on different occasions, have attitudes with

* Academia Sinica. E-mail: hsiangyun@gmail.com

a common focus, whether or not there actually is something at that focus" (Geach 1967, p. 627). Cases where the common focus is purely intentional are but one variant of the many permutations of intentional identity. Given Edelberg's (2006) helpful distinction between inter-subjective and intrasubjective intentional identity, several long-standing problems, including Kripke's puzzles about belief, fall within the scope of intentional identity. On the other hand, less investigated, but closely-related, is the phenomenon of cross-speaker anaphora, where "two or more agents discuss and exchange information about a subject they have agreed upon, when actually there need not be a real thing they are talking about" (Dekker, van Rooy 1997, p. 3).

In this paper, I argue that the real import of intentional identity has been under-estimated and that coordination is the key to its proper understanding. Drawing on the three sorts of data mentioned above, I propose a unified set of correctness conditions of coordination that is anti-descriptivist and externalist in spirit.

In what follows, I first examine Geach's intentional identity, together with some representative responses, and submit a preliminary general moral concerning coordination (Section 2). Next, I extend the discussion to intra-subjective intentional identity, which includes some well-known philosophical double visions and parasitic attitudes (Maier 2015), and then zoom in on cross-speaker anaphora and consider its contrast to the other two types of intentional identity (Section 3). I close by highlighting the advantages of addressing intentional identity in this holistic approach (Section 4)

2 GEACH'S HOB-NOB SENTENCE & COORDINATION

2.1 GEACH'S INTENTIONAL IDENTITY

Consider the following scenario:

Hob and Nob are residents of a town where witch superstitions are rampant. Hob and Nob live on opposite sides of the town; they have never encountered or heard of each other at all. They each read the local newspaper *Village Voice*'s story that "A witch has been terrorizing the town." Hob and Nob independently came to the conclusion that this is the cause of his friends' livestock problem. Hob and Nob have no particular witch in mind. As a matter of fact, there are no witches.

Given the scenario, (1) seems true:

(1) Hob thinks a witch blighted Bob's mare, and Nob thinks she killed Cob's sow.[1]

Traditional resources, however, fail to provide the truth conditions of the Geach sentence that seem so natural.

Suppose we analyze (1) by having the *existential quantifier* take wide scope relative to *belief*, as illustrated in (2):

(2) $\exists x(witch(x) \wedge BEL(h, B(x)) \wedge BEL(n, K(x)))$[2]

(2) is problematic for two reasons. First, the truth of (2) entails the existence of something that satisfies all the relevant properties. But no real individual can meet that requirement. Accepting (2) would, hence, force upon us a problematic ontological commitment. Worse still, (2) does not even provide an accurate description of Hob's and Nob's mental lives, for neither Hob nor Nob has some particular witch in mind, yet the wide scope analysis entails specificity.

A second approach to intentional identity is descriptivism. The descriptivist program involves two critical moves. First, the descriptivist argues that the correct way to analyze the first conjunct of (1) is to let the existential quantifier take narrow scope; that is, the *de dicto* reading is the way to go.

(3) a. (*de re*) $\exists x(BEL\ (h, witch(x) \wedge B(x)))$
b. (*de dicto*) $BEL(h, \exists x(witch(x) \wedge B(x)))$

Since "a witch" and "she" are supposed to be about the same individual, "she" is anaphoric to "a witch" and the former refers to whomever the latter refers to. In order to capture the anaphoric relation, the pronoun "she" in the second conjunct is then treated as a variable,

[1] Geach's original example is this: (G) Hob thinks that a witch blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's sow. For simplicity's sake, the Hob-Nob sentence discussed in the literature is typically the one listed here, a close variant of (G).

[2] An alternative is (1*): $\exists x(BEL\ (h, witch(x) \wedge B(x)) \wedge BEL(n, witch(x) \wedge K(x)))$. (2*) faces the same difficulties as (2).

bound by the same existential quantifier that binds "a witch." Thus we have (4):

(4) $BEL[h, \exists x(witch\ (x) \wedge B(x) \wedge BEL\ n, K(x)))]$

But (4) is false, based on the set-up introduced at the beginning. Hob has no belief whatsoever about Nob; he does not even know that this person lives in the village.

So, the descriptivist's second move is to convert the problematic anaphora into something the conventional theory can better manage, i.e. to reconstruct the anaphora as a definite description. Thus, we arrive at (5a) and the corresponding (5b):

(5) a. Hob believes a witch blighted Bob's mare, and Nob believes *the witch that blighted Bob's mare* killed Cob's sow.
b. $BEL[h, \exists x(witch\ (x) \wedge B(x))] \wedge BEL\ [n, \exists x(witch\ (x) \wedge B(x) \wedge \forall y\ (witch\ (y) \wedge B(y) \rightarrow y = x) \wedge K(x))]$

Again this is incorrect, for nothing in the original story guarantees that Nob knows anything about Bob's mare. An alternative descriptive reconstruction is (6), yet it is still flawed because Hob and what Hob believes need not be part of Nob's mental life:

(6) a. Hob believes a witch blighted Bob's mare, and Nob believes *the witch that Hob believes blighted Bob's mare* killed Cob's sow
b. $BEL[h, \exists x(witch\ (x) \wedge B(x))] \wedge BEL[n, \exists x\ (BEL(h, witch(x) \wedge B(x)) \wedge \forall y\ (BEL(h, witch\ (y) \wedge B(y)) \rightarrow y = x) \wedge K(x))]$

Though other descriptive paraphrases are certainly available, they all eventually run into similar troubles. What is problematic about (4) is that we cannot write into Hob's belief state any information regarding Nob; what is wrong with (5), (6) and the like is that we cannot include in Nob's mental state anything about Hob.

It might appear that descriptivism is curbed due to the lack of appropriate descriptions that accurately characterize the agent's mental life, but an even more serious problem is that when agents do associate the same or similar descriptions to a certain individual, intentional identity does not necessarily follow. Consider the case brought out by Pagin (2014):

(7) a. Hob believes: the tallest witch in the world has brought about a storm.
b. Nob believes: the tallest witch in the world has gold teeth (Pagin 2014, p. 99).

(8) Hob believes that a witch has brought about a storm and Nob believes that she has gold teeth.

At first sight, the truth of (7a) and (7b) implies the truth of (8). Nevertheless, there are at least two reasons to be skeptical. For one thing, if Hob and Nob live in different communities and come to their respective beliefs independently, then (8) would have been false even if (7a) and (7b) were true.[3]

For another, the scenario that renders (7a) and (7b) true can be something like the following: Hob believes that his next door neighbor, Freya, is the tallest witch, and Hob believes that Freya has brought about a storm. Nob believes that his next door neighbor, Ingrid, is the tallest witch, and Nob believes that Ingrid has gold teeth. Freya and Ingrid are not identical, and so while Hob and Nob employ exactly the same identifying description – the tallest witch, they do not intend the same individual. Hence, the prospect of descriptivism as a solution to intentional identity is doubly jeopardized: not only is it difficult to come up with descriptions that suitably portray the agent's inner life, but shared descriptions are simply insufficient to ensure the agents are thinking about the same thing.

2.2 COORDINATION

Since the challenge presented by intentional identity is something that our conventional semantic tools cannot handle, the problem is commonly deemed a logical conundrum. Typical responses often resort to a broader domain of quantification. One line of thought is to explain the truth of (1) in terms of exotic objects, such as mythical objects, abstract objects, merely possible objects, or non-existents objects (e.g. Parsons 1974, Saarinen 1982, Salmon 1998, 2002, Priest

[3] Burge (1983, p. 95–96) expresses something similar. This is in contrast to Crane's (2013) account of intentional identity, according to which similarity is paramount.

2005). The other prominent route is to quantify over intentions, particularly shared intentions among different agents (e.g. Cohen 1968, Geach 1981, Edelberg 1992, Asher 1987).[4]

For example, take Edelberg's (1992) analysis, according to which (1) is represented as follows:

(9) $\exists\alpha\exists\beta[BEL(h, B(\alpha))\wedge BEL(n, K(\beta)) \wedge \alpha \approx \beta]$

Here α and β are variables ranging over belief objects or "person-bound" mental images and "≈" stands for the counterpart relation. So (9) states that Hob and Nob each has a certain mental representation (the belief objects) and that there is some relation between them.

Yet Edelberg's analysis and the like[5] are vulnerable to the following criticism. First, the exact nature of the belief object (or mental representation) seems hazy. If an account make recourse to such things, it must explain what it means for an agent to believe, or more generally, to entertain thoughts about the said entity. Second, the counterpart relation is crucial, but what does it take for one agent's belief object to be a counterpart of another agent's? To this question, Edelberg (1992) claims that belief objects of different agents are counterparts if and only if they play a "similar explanatory role" in the agent's belief system. Still, it remains rather unclear what that amounts to. As our earlier discussion of (7) suggests, two agents may construe their respective mental representation in very similar fashion, such that the belief objects in question play corresponding roles in each agent's mental life. For instance, we can explain Hob's and Nob's behavior towards their neighbors in terms of the thought they each associated with their belief object. While the belief objects appear to be counterparts, intentional identity does not hold, simply because Freya and Ingrid are not identical. A third problem concerns the derivation of the logical

[4] To be sure, there are others who claim the apparent true reading of the Hob-Nob sentence is but an illusion; there is in fact no Geachian reading (e.g. Braun 2012). Yet others argue that it is the assumption that natural language quantifiers are ontologically-loaded that renders the semantics of the Hob-Nob sentence difficult (e.g. Azzouni 2012; cf. Crane 2013).

[5] According to Asher 1986, intentional identity reports are true if and only if the links between discourse markers in the relevant agents' discourse representation structures (DRSs) obtain a certain specification. Here the nature of the links is reminiscent of Edelberg's counterpart relation.

form. In the surface structure of (1), the noun phrase "a witch" lies in the scope of the attitude verb "believe"; however, in (9), it is the existential quantification (over belief objects) that takes the wider scope. What justifies this mismatch? If we are dissatisfied with the wide scope analysis illustrated in (2) or (2*), should we not also worry about (9)?

It seems that we are caught between a rock and a hard place. The dilemma is that we cannot rely on descriptivism and the conventional semantic apparatus, but expanding the domain of quantification seems risky.

The way out, I believe, is to remind ourselves of the generality of intentional identity and take a cue from a modified Hob-Nob scenario:

(10) a. Hob believes that Hesperus is very hot.
b. Nob believes that Phosphorus is very bright (Pagin 2014, p. 96).

Suppose (10a) and (10b) faithfully report Hob's and Nob's beliefs. Suppose further that Hob and Nob do not know each other, and neither of them knows that Hesperus is Phosphorus. It does not really matter whether Hob and Nob attach different descriptive content to (their mental representation of) the same heavenly body. As "Hesperus" and "Phosphorus" are co-referring, the truth of (10a) and (10b) is enough to make the Geach-style (11) true:

(11) Hob believes a heavenly body is very hot and Nob believes it is very bright (Pagin 2014, p. 96).

Actual identity implies intentional identity.[6] Here the agents do in fact have attitudes towards a common focus, and there is a real entity at that focus. Thus, we may elaborate on Geach's informal definition of intentional identity: "[w]e have intentional identity when a number of people, or one person on different occasions, have attitudes with a common focus, whether or not there actually is something at that focus, *and whether or not the people involved realize this*."

One might think that this case is in sharp contrast to the original Hob-Nob sentence where the common focus is merely *intentional*.

[6] I am grateful to Peter Pagin for making this point to me when I presented an earlier version of this paper at PhiLang 2017 in Lodz.

When there is an actual entity, the task is relatively straightforward, of course; what is baffling in the Geachian scenario is that there is nothing at the common focus. We cannot talk about intentional identity, or any kind of identity, if there is nothing whatsoever about which we can make the identity judgement.

However, it is not the case that there are no clear criteria for the identity claims. The examples examined so far demonstrate that even in cases of *merely intentional* identity, people do naturally make typically unanimous judgments. Most agree that the problem is not how people, laypeople and experts alike, do in fact make such judgements, but how theorists explain the reasons behind the judgements people actually make. These judgements are not arbitrary, and so long as we take them seriously, its is not difficult to see that the determining factor is always something factual.

To follow Geach's way of speaking, this means when there is something at the common focus, that very thing is the ultimate measure of intentional identity; when there is nothing at that focus, intentional identity is, and must be, grounded in something that we know of the agents and/or of their environment as a matter of fact.[7] For Geach's Hob-Nob sentence, intentional identity can "only make sense if the agent's attitudes are coordinated together, whether by means of communication or some other mechanism, in such a way that the two agents can be said to have the 'same' individual in mind'" (Asher 1987, p. 127). The source of coordination that validates a true reading of the Hob-Nob sentence is the fact that there is this newspaper article based on which Hob and Nob form their beliefs. The so-called counterpart relation should be understood in an externalist light as a necessary link between individual agents and the factual, causal mechanism. When the wide scope, *de re* reading is not available, we can opt for a realist *de origine* interpretation in the sense of Dekker and van Rooy 1997 and Zimmermann 1999.[8] That is, counterparts are not counterparts unless

[7] This is known as the common source condition. See Moltmann (2006).

[8] In his discussion of the epistemic role of discourse referents, Zimmermann notices what he calls the intentional puzzle; he critically assesses a number of potential solutions, and argues that the "*de origine* solution," though not without its own drawbacks, is the most satisfying. According to the *de origine* solution, a discourse referent represents a source of the informational content of its information state.

they are caused by the same source. Therefore, should it be that Hob and Nob come to their respective beliefs about a witch by reading two different newspapers that are informationally independent, since their beliefs are not grounded in the common source, there is no intentional identity.[9]

This realist, externalist stance entails that not all coordination matters. For example, just as multiple agents can associate potentially drastically different descriptions to their intended object (e.g. (11)), an agent may also hold dissimilar, changing attitudes toward the same thing. Coordination of descriptions is beside the point. Besides, an agent's self-awareness of whether intentional identity obtains is irrelevant (e.g. (1)); it is in this sense that the agent's belief system need not coordinate with the reality.

So, despite the many controversies that Geach's intentional identity has triggered, the general morals are straightforward. What is surprising, however, is that the lessons have not been fully appreciated in the literature. As "[o]ne leitmotif in the philosophy of language and mind of the past fifty years has been its anti-descriptivism [,]" (Récanati, Murez 2016, p. 267) here we have another instance against descriptivism. Moreover, the anti-descriptivist move is motivated by externalism: ultimately, it is the external fact that decides whether intentional identity is sustained or fails. Meanwhile, not all expansions of the quantification domain are the same. While variables ranging over nonexistent or mythical objects do appear suspicious, it is not at all impossible to maintain a realist view towards intentions, mental representations, or possibilities, especially when they can be well-individuated.[10]

[9] Zimmermann considers an example from Edelberg (1992) where two groups of astronomers independently observed a peculiar motion of super-clusters of galaxies and, despite calling it by different names, came up with basically the same explanation. As it turns out, the cause that both groups identified does not exist at all; the peculiar motions of the various super-clusters are each caused by independent factors. Edelberg's story is supposed to show that intentional identity does not require different agents' informational or intentional states to be coordinated through communicative endeavor; Zimmermann further argues that if we have a Twin Earth (Putnam 1975) version of the same story, then because the two groups observed in fact distinct phenomena, there is no intentional identity.

[10] An anonymous referee raised the worry that it is quite a challenge to provide a theory of the said individuation. I do not deny the difficulty in providing such a theory, but I think there are works in the literature that show some promise.

A realist extension of the usual domain along these lines is not only innocent but necessary. After all, we are committed to entities that can best explain the data in a systematic way. For the purpose of semantic theorizing, a wide scope analysis, when carefully qualified, can be part of our best analysis.

3 INTRASUBJECTIVE CASES & CROSS-SPEAKER ANAPHORA

3.1 INTRASUBJECTIVE INTENTIONAL IDENTITY

So far, I have focused on Geach's intentional identity and diverged to its close variant only briefly; but intentional identity has a very wide coverage. For instance, consider:

(12) Le Verrier believed that Vulcan is located between the Sun and Mercury, and many others believed they had seen it.

As it turns out, Vulcan does not exist, but that does not change the fact that there can be intentional identity regarding Vulcan. Crucially, scientific progress often relies on entertaining thoughts and investigating ideas that involve entities that may or may not exist; breakthroughs in technology frequently hinge on people working together to bring a previously non-existent object into being.[11]

What's more, as Edelberg (2006) correctly points out, intentional identity has both intersubjective and intrasubjective versions. My goal in this section is to lay out some representative examples of intrasubjective intentional identity and re-evaluate the realist, externalist conditions of coordination sketched in the previous section.

To begin, the philosophical literature is well stocked with an assortment of cases where an agent associates not just different, but

---

See, for example, Asher 1987, Kamp, van Genabith, and Reyle 2011, Pagin 2014, and Maier 2015. Recent discussions on mental files (Récanati 2012, 2014, 2016) also shed light here: we can distinguish the information stored in a file, its *contents*, from the file itself. Files are not individuated based on their contents, but the causal relation that brings them about. The realist, externalist view of mental representations that I have in mind is comparable to what Récanati says of mental files.

[11] For example, a team of scientists and engineers worked for years, with a common focus of course, before AlphaGo became known to the public.

sometimes inconsistent predicates to the same thing without knowing it. This list includes Frege's puzzle, Quine's discussion of Ralph and Ortcutt, and of course, Kripke's puzzle about belief. For simplicity's sake, I center on Kripke's examples of Pierre and Paderewski:

(13) Pierre thinks London is pretty, and he thinks it is not pretty (Kripke 1979).[12]

(14) Peter thinks Paderewski has musical talent, and he thinks he doesn't have musical talent (Kripke 1979).[13]

These philosophical double visions provide support for our anti-descriptivist, externalist analysis. Obviously, the agent in question has attitudes about the same thing, despite failing to realize this himself. Since coordination works in a factual fashion and does not depend on the agent's inner awareness, intentional identity holds. To be sure, if Peter later found out that what he previously thought of as two people are one and the same, he would have to coordinate the content of his belief system accordingly. While this higher level coordination is demanded by rationality, it is not required for intentional identity.

On the other hand, there are cases of intrasubjective intentional identity that do not seem so philosophically baffling:[14]

---

[12] Here is a brief summary of The Pierre Puzzle: Pierre is a normal French speaker living in France. He learns, in French, the name "Londres" as the name for London. He accepts, in French, many claims about the city, including that it is beautiful. So, in French, he says "Londres est jolie." Under unfortunate circumstances, Pierre is later moved to and confined in a rather unattractive part of London. He manages to pick up the local language through interaction with his neighbors, who speak no French. Pierre acquires "London" as the name for London, and thinks of it as not very pretty.

[13] The Paderewski Puzzle is the monolingual version of Kripke's puzzle about belief. Peter learns the name "Paderewski" as the name of a famous pianist. He later learns of someone called "Paderewski" and this person was a Polish national leader and Prime Minister. Since Peter doubts the musical abilities of politicians, he concludes that these are two different people who happen to share the same name.

[14] The following examples have been debated extensively in the literature of philosophical semantics and the intersection of semantics and pragmatics. Following Roberts (1996), they are often referred to as modal subordination, or intensional subordination per Moltmann (2006). As it turns out, it is extremely

(15) Bill believed that Fred had been beating his wife and he hoped that Fred would stop beating her (Karttunen 1973, ex (42)).

(16) Bill believes he saw a fish and wishes that he had caught it. (McKinsey 1986)

(17) Alice fears there's a squirrel in her kitchen cabinets; she hopes to catch it alive and turn it outside (Roberts 1996).

(18) Grandmom thinks a snake is in the barn, and she wants to shoot it (Edelberg 2006).

On the face of it, (15) through (18) are rather pedestrian. None of them involves any mythical or non-existent creatures. These examples strike one as unremarkable precisely because we use such talk on a daily basis; they exemplify how we use folk psychology to explain people's thought and behavior. For instance, the truth of (17) can explain why Alice acts in an awkward way in the kitchen; the truth of (18) can explain why an eighty-year old lady is taking a gun to the barn. But, notice that it is conceivable there is in fact no squirrel in the kitchen, in which case Alice's intended object does not exist. The truth of (17) is independent of whether "a squirrel" denotes something, just as Geach's Hob-Nob sentence can be true while "a witch" is empty.

There are also cases where the apparent descriptions do not cohere:

(19) John thought he heard a woman's voice but suspected that it was not a woman's voice.

(20) Arya wants to shoot and kill a werewolf, but she fears that she will only hurt it.[15]

---

difficult to delineate the semantics for these sentences in a compositional way. Part of the problem has to do with the lack of a satisfactory theory of presupposition; other issues include the complex hierarchy between attitudes. Note that not all combinations of attitudes are felicitous: a. John tries to catch a unicorn and wishes to eat it. b. #John wishes to catch a unicorn and tries to eat it.

[15] The intended reading is where the indefinite "a werewolf" takes the narrow scope; that is, Arya's desire is non-specific. Arya has a general desire to shoot and kill one werewolf or another, and a subsequent fear about the same thing.

In (19), the agent appears to hold conflicting attitudes toward the same thing: on the one hand, John believed that he heard a woman's voice, but on the hand doubted that it is. The clash in (20) is subtler, however Arya apparently has two attitudes – a desire and a fear – toward the same thing. In terms of possible world semantics, in those of Arya's desire-worlds where she shoots and kills a werewolf, none of them is such that she just hurts it. So the specific content embedded in the scope of each attitude just does not match.

Again, these two sorts of data concerning intrasubjective intentional identity validate our criteria of coordination. First, they demonstrate the inadequacy of the descriptivist approach to intentional identity. Note that (15) through (20) are attitude reports of rational agents; even with tricky examples like (19) and (20), one would still confidently accept their truth without having to conclude that the agent's mind is confused or disturbed. Coordination in these cases does not require the agents to entertain thoughts about their intended objects in descriptions that are entirely consistent. This is so especially for *counterfactual* attitudes, such as wishing, pretending, and imagining, which are typically inconsistent with what the agents believe.[16] Note further, that all the cases of intrasubjective intentional identity examined here, including both Kripke's puzzles and the various cases of modal subordination, are anti-descriptivism. Whether or not the agent in question is aware of the identity of her intended objects, descriptivism fails to provide the desired explanation.

Furthermore, it is worth emphasizing that in (15) through (20), the agent's second attitude is always referentially dependent on their first attitude: Alice's hope to catch a squirrel is based on her fear that there is one in her kitchen, and Arya's fear that she only hurts a werewolf stems from her desire to kill one. Indeed, the referential dependency found in these phenomena of parasitic attitudes (Maier 2016) entails the need to specify a tracking device in our theory of mental representation, typically along the lines of a referential reconstruction of natural language terms as relating to specific entities (e.g. discourse referents) in the agent's information state.[17] As stated earlier, a well-

[16] For more details, see Ninan (2008) and Maier (2016).

[17] See, for instance, Asher 1986, 1987, Dekker and van Rooy 1997, Kamp et. al. 2011, and Maier 2015, 2016.

structured theory of mental representation is in line with the realist stance, so the referential or *de re* element in the formal analysis is above reproach.

3.2 CROSS-SPEAKER ANAPHORA

Now I want to contrast intentional identity with the closely related phenomenon of cross-speaker anaphora, or what Dekker and van Rooy (1997) call "Hob-Nob situations." According to Dekker and van Rooy, these situations are "cases where two or more agents discuss and exchange information about a subject they have agreed upon, when actually there need not be a real thing which they are talking about" (Dekker, van Rooy 1997, p. 3). Defined as such, its resemblance to Geach's intentional identity is hard to miss.

Consider the following:

(21) A: The man drinking a martini looks happy.
B: He is not drinking a martini.

(22) A: A man jumped out of the crowd and fell in front of the horses.
B: He didn't jump, he was pushed (Strawson 1952).

(23) A: A man is sleeping over there on a park bench.
B: It is not a man, it is a woman and she is not asleep, she is just sunbathing.
Besides, it is not a park bench. (Dekker, van Rooy 1997, p. 4).

In (21), we are reminded of Donnellan's "the man drinking martini;" (22) is a classic example from Strawson, and (23) its reinforcement. In each of the above examples, the second speaker does not agree with and corrects the predicative content the first speaker employs. Once again, the anaphoric pronouns and their antecedents are described in conflicting ways; as (23) shows, the second speaker may object to every piece of descriptive information that the first speaker mentions. Furthermore, it could be the case that both speakers are mistaken about the predication, or that there is nothing at their common focus, as in the case when both agents are hallucinating.

The situations exemplified by (21) through (23) are not odd or uncommon. They too are cases of intentional identity. Moreover, these

Hob-Nob situations support our externalist analysis of coordination, despite an important asymmetry.

Let me first highlight a number of key features of cross-speaker anaphora. To begin, the information exchange in these scenarios is intelligible because, regardless of the conflicting predication different speakers attributed to their common intended object, the referential intention remains stable. This is both an endorsement of the anti-descriptivist stance and a verification of the realist commitment. Again, when there is something at the common focus in these Hob-Nob situations, that very thing is the ultimate anchor for the multiple speakers' reference; it also serves as the measure of intentional identity. On the other hand, in cases where nothing exists at the common focus, the lack of an anchor means we must trace the agents' referential intention to an external, factual common source. Take (23) for example, we can imagine that both speakers are under the influence of drugs (or alcohol), and it is due to this common factor that they take themselves to be conversing about the same entity.

Crucially, however, cases of cross-speaker anaphora are special in that the second (or non-first) speaker bears the responsibility to preserve the first speaker's referential intention. While this responsibility is asymmetrical, we do not find any such thing among the array of intentional identity we have so far investigated. In other words, in-person communication places a unique demand on coordination: the second speaker is required to coordinate with the first in their referential intention. This reference-preserving intention is absent in the other cases of intentional identity.

4 CONCLUDING REMARKS

Philosophers have a long history of bemusement by reference and attitudes, and intentional identity is the perfect testimony to this bewilderment. Entertaining thoughts about entities and talking about them is such a familiar and fundamental part of our lives that puzzles thereof are both extremely intriguing and frustrating.

The ubiquity of intentional identity implies that foundational questions about human communication fall within the scope of its proper study. Drawing on the traditional problem of intentional identity, cases of intrasubjective intentional identity, and the deeply

connected phenomenon of cross-speaker anaphora, I propose a unified analysis of coordination that is key to the proper understanding of intentional identity. This understanding of intentional identity is holistic: it places the problem Geach first identified in a broader context and connects it to other interesting data whose inter- relatedness is under-investigated. The anti-descriptivism is validated because we can cluster diverse information and entertain distinct or even clashing attitudes toward the same entity in language as well as in thought. If reference is exhausted by the predicative information, there is no justification for how intentional identity can ever be established. In addition, the proposed condition of coordination makes no recourse to obscurity and sustains the externalist, realist spirit. While the analysis respects the fact that delineation of mental states often requires stipulating referential devices in linguistic and mental representations, finally, all the stipulations, be they pragmatic or not, must be grounded in external, factual terms.

REFERENCES

Asher, N. (1986). Belief in Discourse Representation Theory. *Journal of Philosophical Logic*, *15*, 127–189.

Asher, N. (1987). A Typology for Attitude Verbs and Their Anaphoric Properties. *Linguistics and Philosophy*, 10(2), 125–197.

Azzouni, J. (2010). *Talking about Nothing: Numbers, Hallucinations, and Fictions*. Oxford: Oxford University Press.

Braun, D. (2012). Hob, Nob, and Mythical Witches. In W. Kabasenche, M. O'Rourke and M. Slater (Eds.), *Reference and Referring* (pp. 149–187). Vol. 10. of Topics in Contemporary Philosophy. Cambridge, MA: MIT Press.

Burge, T. (1983). Russell's Problem and Intentional Identity. In J. E. Tomberlin (Ed.), *Agent, Language, and the Structure of the World: Essays Presented to Hector- Neri Castaneda with his Replies* (pp. 79–110). Indianapolis, Ind.: Ridgeview Publishing.

Cohen, L. J. (1968). Geach's Problem about Intentional Identity. *Journal of Philosophy*, *65*, 329–335.

Crane, T. (2012). *The Objects of Thought*. Oxford: Oxford University Press.

Dekker, P. and van Rooy, R. (1997). Intentional Identity and Information Exchange. In *Second Tbilisi Symposium on Language, Logic, and Information*.

Edelberg, W. (1992). Intentional Identity and the Attitudes. *Linguistics and Philosophy*, *15*(6), 561–596.

Edelberg, W. (2006). Intersubjective Intentional Identity. *Journal of Philosophy*, *103*(10), 481–502.

Geach, P. T. (1981). *Logic Matters* (2nd edn). Oxford: Basil Blackwell.

Geach, P. T. (1967). Intentional Identity. *Journal of Philosophy*, *64*(20), 627–632.

Kamp, H., van Genabith J., Reyle, U. (2011). Discourse Representation Theory. In D. M. Gabbay and F. Guenthner (Eds.), *Handbook of Philosophical Logic* (Vol. 15). Springer.

Kripke, S. (1979). A Puzzle about Belief. In A. Margalit (Ed.), *Meaning and Use* (pp. 239–283). Dordrecht: Reidel.

Maier, E. (2015). Parasitic Attitudes. *Linguistics and Philosophy*, *38*, 205–236.

Maier, E. (2016). Referential Dependencies Between Conflicting Attitudes. *Journal of Philosophical Logic 46*(2), 141–167.

Moltmann, F. (2006). Unbound Anaphoric Pronouns: E-type, Dynamic, and Structured Propositions Approaches. *Syntheses Volume*, *153*(2), 199–260.

Murez, M., Récanati, F. (2016). Mental Files: an Introduction. *Review of Philosophy and Psychology*, *7*, 265–281.

Pagin, P. (2014). Intersubjective Intentional Identity. In M. García-Carpintero and G. Martí (Eds.), *Empty Representations: Reference Non-Existence* (pp. 91–113). Oxford University Press.

Parson, T. (1974). A Prolegomenon to Meinongian Semantics. *The Journal of Philosophy*, *71*(16), 561–580.

Priest, G. (2005). *Towards Non-being: The Logic and Metaphysics of Intentionality*. Oxford: Oxford University Press.

Putnam, H. (1975). The Meaning of 'Meaning'. In *Mind, Language and Reality. Philosophical Papers*. Vol. 2 (pp. 215–271). Cambridge: *Cambridge* University Press.

Récanati, F. (2012). *Mental Files*. Oxford University Press.

Récanati, F. (2014). Empty Singular Terms in the Mental File Framework. In M. García-Carpintero and G. Martí (Eds.), *Empty Representations* (pp. 165–182). Oxford: Oxford University Press.

Récanati, F. (2016). *Mental Files in Flux*. Oxford: Oxford University Press.

Roberts, C. (1996). Anaphora in Intensional Contexts. In *The Handbook and Contemporary Semantic Theory*. Basil Blackwell.

Saarinen, E. (1982). Intentional Identity Interpreted: A Case Study of the Relations among Quantifiers, Pronouns and Propositional Attitudes. In E. Saarinen (Ed.), *Game-Theoretical Semantics* (pp. 245–327). Dordrecht: Reidel.

Salmon, N. (2002). Mythical Objects. In *Metaphysics, Mathematics, and Meaning* (pp. 91–207). Oxford: Oxford University Press.

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

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 53–75
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.04
Rozprawa

Filip Kawczyński*

IS MEANING HOLISM COMPATIBLE WITH SEMANTIC MINIMALISM?

Summary: Meaning Holism and Contextualism are standardly acknowledged to be similar relativistic theories that often lead to similar troubles, in particular to issues concerning instability. On the other hand, the main rival of Contextualism, which is Minimalism, is taken to be resistant to these problems. In effect, it seems inevitable to see Meaning Holism and Minimalism as natural enemies. In my paper, I attempt to reject such a view. My argumentation consists of three main parts. First, I argue that Minimalism does not differ that much from Meaning Holism with respect to the instability issues as it also faces some of them (although in a slightly different way from the case of Holism of Contextualism). Second, I put forward several arguments to show that in fact Minimalism is not incompatible with the two versions of Meaning Holism I distinguish, namely Global Holism and Local Holism. I argue that a meaning holist has to accept some not uncontroversial principles to become an anti-minimalist – and vice versa. Finally, I demonstrate that Minimalism and Meaning Holism can be reconciled. Such a possibility occurs when something I called 'purely semantic processes' is allowed. The role of these processes is, roughly speaking, to protect literal meanings from being affected by strong pragmatic factors.

Keywords: meaning holism, contextualism, minimalism, local holism, global holism, instability, context sensitivity, literal meaning, pragmatic processes

* University of Warsaw, Institute of Philosophy. E-mail: f.kawczynski@uw.edu.pl

INTRODUCTION

The main goal of this paper is to make the initial move in the process of rejuvenating the theory of Meaning Holism.[1] In particular, I am going to figure out what is the location of Meaning Holism (MH) on the map of today's theories in the philosophy of language. In recent years the philosophical reflection of language has been dominated by the problem of context influence upon semantic content. The two main paradigms in this dispute are Contextualism and Minimalism. The common opinion is that MH and Contextualism – as they both rest on the relativistic foundations – stay in some tight theoretical relations and support each other. Consequently, it appears that there is no other option than to consider MH and the chief rival of Contextualism, i.e. Minimalism, as enemies that exclude each other. I am going to argue that this common opinion is wrong[2] and the alleged kinship between MH and Contextualism is in fact much more distant than it is supposed to be. Furthermore, I am going to offer arguments for the claim that MH and Minimalism are in fact compatible. *Summa summarum*, I am going to present MH from a new perspective which, as I believe, makes the theory more attractive.

The crucial point in judging what is the relation between MH and Minimalism is to make clear what the two theories hold. It is also not the easiest point to discuss, as both theories have several formulations which differ significantly. Since in Kawczyński (2018) I presented a wider picture of what MH is, on the one hand, and how the Contextualism-Minimalism debate looks like, on the other, here I will restrain myself to the nuts and bolts of the issue.

[1] The theory slipped to oblivion a few years ago – largely due to the severe criticism offered by Fodor and Lepore (1992).

[2] In Kawczyński (2018) I offer argumentation against such a standpoint and I show that MH is logically independent of any version of Contextualism i.e. although it is compatible with most of them, it neither entails, nor is entailed by any. Discussed in this paper will be the other side of the coin i.e. the relation between MH and Minimalism, which is theoretically independent of the previous analysis although they complement each other.

1. THE THEORIES

1.1 MEANING HOLISM

MH can be characterised in several ways. I guess that when there are doubts concerning the definition of some theory, the first thing a philosopher usually does is checking the appropriate entry of the *Stanford Encyclopedia of Philosophy*; hence, let us look how the theory in question is defined there by Henry Jackman:

> The label "meaning holism" is generally applied to views that treat the meanings of all of the words in a language as interdependent. Meaning holism is typically contrasted with *atomism* about meaning (where each word's meaning is independent of every other word's meaning), and *molecularism* about meaning (where a word's meaning is tied to the meanings of some comparatively small subset of other words in the language […] (Jackman 2017: §1)[3]

What Jackman actually characterises is the holistic *rule* for meaning, not any particular theory of meaning. The rule can be applied to various sets of semantic axioms and thus output different semantic theories. Since I do not want to commit myself to any particular theory of meaning, I am going to consider something I call (*Meaning) Holism as a Principle*:

(H-Principle) Meaning of a single expression depends on meanings of all other expressions in a given linguistic system.

I believe the H-Principle reflects the general idea of the holistic account of language. I would like to emphasise that it is *merely* a principle, not a *theory*, and as such it can be reconciled with various *theories* regarding meaning.[4] Since I want the principle to stay as broad as possible, I am not going to try to make it more precise. Instead, I would like to draw the distinction between two possible versions of MH distinguished with regard to what is defined as the *linguistic system* mentioned in the principle. When applied to whole languages,

[3] For an in-depth analysis of different possible formulations of MH I recommend Pagin (2006) and Peacocke (1997).

[4] Actually, even the word "meaning" appearing in the above formulation of the pronciple is broader than I would like it to be. Depending on what subject matter the principle is applied to, "meaning" can be interpreted as e.g. semantic significance, semantic value, character, content, or even representational properties.

the H-Principle yields the account I call *Global Holism*, according to which meaning of every linguistic item in a given language depends on meanings of all other linguistic items of the language. The thesis of Global Holism could be thus worked out as the claim that *literal meanings* are constituted / defined / formatted[5] in a holistic way.[6] On the other hand, *Local Holism*[7] is the theory which stems from applying the H-Principle to a given part of language, in particular to a single sentence or to a speech act or an utterance.[8] That leads to the view that all words occurring in a sentence, an utterance, a speech act etc. are associated in the way that the meaning of every word depends on the meanings of all other words.[9] Shortly speaking, according to Local Holism *meanings-in-contexts* behave holistically. Simplifying it slightly, it might be said that Global Holism concerns meanings of *types*, while Local Holism regards meanings of *tokens*.[10] In the later sections I examine in what relation Minimalism stands in these two versions of MH.

1.2 MINIMALISM

As we get to define Minimalism it is good to start with mentioning that how the borderline between Minimalism and Contextualism goes is in itself a challenging question without a good answer to date.[11]

---

[5] I use the word "formatted" just as it is used in the debate between Contextualists and Minimalists; see e.g. Récanati (2004, p. 140–141).

[6] Global Holism may be understood as the *metasemantic* account according to which meaning (or more generally: semantic significance) is assigned to linguistic items in a holistic way i.e. in a given language each [assignment of] meaning depends on every other [asssignment of] meaning.

[7] The terms "Global Holism" and "Local Holism" may be found in the literature about holism (e.g. see: Peacocke 1997; Penco 2001) but they have not earned fixed meaning or reference so far.

[8] But it could be a *set* of utterances, speech acts, sentences etc. as well, so for instance, it might be said that the whole monologue or a scientific theory are holistic in the sense provided by the H-Principle.

[9] Analogously to what has been said about Global Holism, Local Holism may be seen as the *semantic* theory which says that semantic values of linguistic items – the values that are their contributions to truth-conditions of relevant sentences – are determined holistically.

[10] Another approximation could be that Global Holism deals with Kaplanian *character*, while Local Holism with *content*.

[11] As pointed by Joanna Odrowąż-Sypniewska (2013, p. 65–70) there are four possible criteria of the division and being classified as a contextualist according to one of them does not guarantee being such classified according to another.

However, basing on the definitions offered by several prominent philosophers taking part in the dispute[12] we can draw the conclusion that in general being a minimalist consists of accepting conjointly the two following theses:

(M1) Intuitive propositional content of a well-formed sentence is never determined by the strong pragmatic effects.

(M2) Class of natural language context-sensitive expressions overlaps with the set of obviously indexical expressions (i.e. than the "Kaplan's set") or is insignificantly bigger than it, i.e. there is *not many* context-sensitive expressions.[13]

For the sake of precision, let me say that the intuitive propositional content is *propositional* because it has truth-value and it is *intuitive* because it is distinguished from the *literal* content – e.g. the notorious sentence "The table is covered with books" may be regarded as expressing the literal russellian proposition (i.e. that there exists exactly one table and it is covered with books) and the intuitive proposition that there is a particular table (not necessarily the only table in the universe) that is covered with books.[14] The best way to explain what the strong pragmatic effects are is to give the floor to Jeffrey King and Jason Stanley who have introduced the notion:

A weak pragmatic effect on what is communicated by an utterance is a case in which context (including speaker intentions) determines interpretation of a lexical item in accord with the standard meaning of that lexical item. A strong pragmatic effect on what is communicated is a contextual effect on what is communicated that is not merely pragmatic in the weak sense. (King, Stanley 2005/2007, p. 140)

---

[12] I rely in particular on: Bach 2005; Borg 2004; 2012, p. 4–5; Cappelen, Lepore 2005, p. 143–145; Récanati 2004; 2010; Stanley 2007.

[13] Such a standpoint excludes from the group of minimalists the philosopher who pronounces himself to be a "radical minimalist", namely Kent Bach, who rejects so called *propositionalism* i.e. the claim that every well-formed sentence expresses a proposition. However, there are good reasons for doing it this way – in particular, I find Récanati's (2010, 12–14) arguments for the thesis that the only reasonable version of minimalism is the one that assumes propositionalism to be pretty convincing. For further discussion on Bach's place in the debate see Odrowąż-Sypniewska (2013, p. 73–74).

[14] For the wider explanation of what the intuitive propositional content is see: Récanati 2004, p. 8–16; Stanley, Szabó 2000/2007, p. 25.

All in all, minimalists can be characterised as those who think that there exists a level of propositional content which is immune to strong pragmatic effects and thereby exists something that is shared by all expressions of a sentence of a given syntactic type; furthermore, minimalists also believe that the number of context-sensitive expressions is remarkably limited.[15,16]

2. ALLEGED INCOMPATIBILITY OF THE THEORIES

2.1 THE INSTABILITY PROBLEMS

I would like to start this attempt of figuring out what relation actually holds between MH and Minimalism by taking a closer look at the possible reasons to think that they are not compatible. I think that the crucial one among them concerns the issue of so called *instability*.

MH has been classified as one of these useless relativistic theories that inevitably lead to the problem-causing *instability*. The phenomena of instability has a lot of faces that have been analysed in various ways (see e.g. Fodor, Lepore 1992; Jackman 2017; Pagin 2006). To make a long story as short as possible: if each meaning in a system depends on every other meaning in the system (as the H-Principle says) and thereby a change of any meaning entails changes in the whole system, then such a system appears unstable. Instability *per se* should not be considered a failure (although it often is), however, when it affects language it seems to give rise to several detailed problems. I would like to focus on two of them.

The first one concerns the *impossibility of genuine disagreement*: if holism is true, then the meaning of each expression *in someone's idiolect*

[15] It is worth emphasising here that the context-sensitivity that M2 is about concerns *any* sensitivity i.e. to both the strong and the weak pragmatic effects, and since in M1 the influence of the strong effects is ruled out, M2 effectively regards sensitivity of literal meanings to the weak pragmatic influences. Due to that, M2 allows us to differentiate minimalism from Stanley's (2007). Indexicalism, according to which most words have encoded in their semantics a requirement for some pragmatical adjustments (i.e. for some weak pragmatic influences).

[16] Within the types-tokens framework it may be said that minimalists believe that in most cases what is contributed to propositions by given tokens are the literal meanings of the relevant types (while contextualists think in many cases what a token contributes is something not identical to the literal meaning of a given type).

depends on all other meanings and in practice it makes it impossible for two people to mean the same by "*p*"; and thus when one expresses it with assertion, while the other one with negation (i.e. "¬*p*"), there is no genuine disagreement between the speakers because they use "*p*" with different meanings.[17] The second problem is somehow wider as it involves the possibility of communication at all. Shortly speaking, if it appears impossible that two speakers ever mean *the same* by their utterances, then either we stay under the illusion that we communicate but in fact we do not (which is dispelled due to empirical observations that we *do* communicate), or our communication should be considered in terms of miracles.[18,19]

These problems, commonly taken to be the arguments against MH, are familiar to anyone who bothers with holism and defenders of MH have advanced several counterarguments to them (for some of them, direct or indirect, see Bilgrami 1998; Block 1986; 1994; Brandom 2000; 1994; Field 1977; Harman 1973; 1993; Jackman 1999; Lormand 1996; Rovane 2013, among others). I have presented these issues briefly not because I am going to offer further counterarguments, but because I want to cast some light on similarities between these particular problems and the arguments appearing in the discussion between supporters of Minimalism and Contextualism. Adherents

[17] Another but closely related problem stems from treating the same person at different times as two theoretical speakers and it consists in the *impossibility of changing one's mind*. If at the moment t_0 I have a belief which I express by uttering "*p*" and at the moment t_1 I express "¬*p*" standardly, we would describe such a situation as changing my mind about *p*. However, just as it is practically impossible that two speakers mean the same by "*p*", I cannot mean the same by "*p*" in t_0 in and t_1 (since if at least one meaning in my idiolect has been modified between those moments, the whole idiolect has changed). Hence, it appears as if I have not changed my attitude towards *p*, but rather have endorsed some *new* belief *q* in t_1 and decided that the best expression of that belief would be uttering "¬*p*".

[18] For the discussion concerning disagreement and communication as troublesome to MH see e.g. Churchland 1993, p. 668–672; Fodor 1987, p. 55–60; Fodor, Lepore 1992, p. 17–22; Fodor, Lepore 2002, misc.

[19] For the sake of clarity my exposition of these issues is somehow simplified and focuses on the side of these problems directly associated with MH. The other side concerns the idea that meanings attached to words are determined by relevant beliefs of speakers. Roughly speaking, a change of a single belief causes the change of meaning of at least one word, and because of MH, it results in change of all the meanings, and thus the change of all the beliefs.

of the former view often accuse contextualists of capturing meaning as something highly vulnerable to change and thus language as something unstable. In particular, it is often pointed as one of the unfortunate consequences of Contextualism that it entails that successful communication seems to be just a *miraculously* happy coincidence (see: Cappelen, Lepore 2005, ch. 8; Récanati 2010, p. 6–10; Stanley 2000; 2002; 2005). Within Contextualism meanings are claimed to change quite rapidly and freely across contexts which according to Minimalists have the effect that for two speakers expressing *the same* meaning by the same words[20] is highly unlikely. On the other hand, the problem concerning the so called *faultless* disagreement,[21] which is taken to be one of the main arguments against Contextualism, in principle boils down to the same issues as those involved in the above-mentioned argument from the impossibility of genuine disagreement.[22]

The resemblance of the problems that MH and Contextualism have to face is probably responsible for why MH and Contextualism are often tarred with the same brush; that, in turn, determines how the relation between MH and Minimalism is seen: if MH has the instability problems and so does Contextualism, then MH and Contextualism are similar, and thereby MH cannot be compatible with Minimalism, because anything that is similar to Contextualism has to go against Minimalism. However, we leave aside these family animosities between Minimalism and Contextualism, and we are going to investigate whether the instability is something that actually settles the question concerning the compatibility of Minimalism and MH.

2.2 INSTABILITY OF MINIMALISM?

Minimalism appears to be the "conservative," noble view that provides a solid ground for communication (and for the analysis of communication as well) as it assumes that "there is a level of content minimally influenced by context" (Cappelen, Lepore 2006, p. 425)

[20] I mean here tokens of the same syntactical type.

[21] The discussion was apparently set up by the famous paper of Max Kölbel (2004).

[22] The contextualist analogue of the *impossibility of changing one's mind* problem would go as follows: if context heavily influences content of my utterances, it appears very unlikely that I would be able to express "*p*" with *the same* meaning in different contexts: in one context with assertion and in another one with negation.

and that content is "fully determined by its syntactic structure and lexical content: the meaning of a sentence is exhausted by the meaning of its parts and their mode of composition" (Borg 2012, p. 4). Shortly speaking, minimalists claim that there are so called *literal meanings* and that literal meanings as they are – without further adjustments – constitute intuitive propositional content.[23] Literal meaning is something that expression *has* i.e. it is its property, which does not change across contexts. Whatever way I may use the word "dog", all of my uses – according to minimalists – have something in common: namely, the literal meaning of the word. And whatever I wish to mean on different occasions by expressing the sentence "The dog is ready," all my uses express the same *minimal proposition*. The minimal proposition consists of the literal meanings of the single words appropriately composed and possibly complemented by the *weak* pragmatic effects (the same minimal proposition would be expressed by e.g. "Der Hund ist bereit" in German or "[Ten] Pies jest gotowy" in Polish etc.). As a result we obtain one pretty stable picture of language and communication. The picture includes the catalogue of stable literal meanings which constitute stable semantic content whereas everything that seems unstable is a matter of implicatures and other *strong* pragmatic phenomena. At this point it appears quite evident that it would be really difficult to reconcile such a stable view with the unstable MH.

Let us investigate, however, whether minimalism is genuinely immune to instability, or it is just *more stable* than its rivals, or maybe just makes the impression of being stable while in fact it is not that stable. Think of the above-mentioned argument concerning the possibility of communication.[24] If a word, say "dog", is supposed to possess a literal

[23] Adherents of Moderate Contextualism or the Wrong Format View also accept literal meanings, however they do not endorse the thesis that literal meanings are essentially involved in the intuitive propositional content, instead they claim respectively that literal meanings sometimes do that or never do that.

[24] Minimalists believe that literal meanings are what lay the foundations of communication and what makes it possible to explain communication. Some of them are quite straightforward in saying that only Minimalism can deal with the issue in question: "Semantic Minimalism, and no other view, can account for how the same content can be expressed, claimed, asserted, questioned, investigated, etc. in radically different contexts. It is the semantic content that enables audiences who find themselves in radically different contexts to understand each other, to agree or disagree, to question and debate with each other. It can serve this function

meaning, then there has to be *the* literal meaning of the word. I believe that the nearest approximations of literal meanings we can arrive at are the appropriate dictionary entires. For instance, according to the "English Oxford Living Dictionaries" the primary meaning of "dog" is:

[dog[1]] a domesticated carnivorous mammal that typically has a long snout, an acute sense of smell, non-retractable claws, and a barking, howling, or whining voice[25]

For the sake of argument let's assume that this is *the* literal meaning of "dog" – that it was somehow extracted from all uses of "dog" – and agree with minimalists that this meaning is stable and all uses of the word express this particular literal meaning (although due to the strong pragmatic effects they can additionally convey some *more* content than the literal meaning). Regardless of how defining such meaning would be even possible without a dose of divine help, it seems really unlikely that accepting such an account would make communication more comprehensible and less miraculous (than in case of Contextualism, for instance). Notice that [dog[1]] involves a dozen or so meanings of other terms: "domesticated," "carnivorous," "mammal," and so on. Now, if to regard communication as not-miraculous one requires speakers to share the same meaning,[26] then it is very unlikely that one will be satisfied with the account of literal meaning we are discussing at the moment. For it seems at least very uncommon that two speakers would share exactly the same literal meaning [dog[1]] and other literal meanings, each of which supposedly involves numerous other meanings (and [dog[1]] probably does not belong to the most complex cases).[27] To look at the

simply because it is the sort of content that is largely immune to contextual variations" (Cappelen, Lepore 2005, p. 152).

[25] See https://en.oxforddictionaries.com/definition/dog. Access: September 2017.

[26] I do not think such an account of communication is right, however, this is the account that grants arguments against MH or Contextualism described above. For a critique of this account see e.g. Block 1986, Churchland 1993, Harman 1973, Rovane 2013.

[27] This may be replied to by saying that to take dictionary entries as literal meanings approximations is attacking a straw man and in fact literal meanings are not complexes of any kind but rather simple sense-entities that human mind has access to and no description is able to actually give an account of them (I guess Fregean senses are the closest to this picture). I agree that such a view is immune to my argument, however, if communication boils down to *grasping non-definable sense-entities* and *sharing* them, I believe that is what most people would call the unbelievably miraculous miracle.

case from a different angle: it appears to be quite a miracle when, without any help of context[28], two people are able to use "dog" with the same meaning. It seems miraculous, because accomplishing this requires that they also share meanings of "domesticated," "carnivorous," "mammal," etc. and apparently many many more.[29]

A minimalist may now reply that she obviously does not assume that *identity* of meanings associated by different people with the same words is necessary for the communication to be successful. Instead, what is enough to understand each other is to share a *sufficient part* of the literal meanings. Such a strategy, however, cannot succeed as it faces the notorious problem of defining what is the "sufficient part."[30] Without having that defined it is impossible to explain satisfactorily how we succeed in communicating, so it still does not allow to anything more enlightening than taking the successful communication to be a very lucky coincidence.

Another possible response of a minimalist is to claim that in their theory the possibility of successful communication stems from the fact that in cases of misunderstandings speakers can always refer to the literal meanings which play then the role of the "highest authority". However, it still looks like a miracle that we usually get along *without* referring to the literal meanings (neither verbally, nor *mentally*). Moreover, in principle, the remedy for a misunderstanding would be referring to *any* meaning that all speakers accept as the one expressed by the words used, and that meaning does not have to be the literal meaning (if Alice agreed with Humpty-Dumpty that "glory" means "there's a nice knock-down argument for you" they would understand each other without a problem whenever speaking of *glories*).[31]

---

[28] Such a help for most expressions is banned by the minimalist credo.

[29] From this point of view it looks as if Contextualism provides a more efficient way of analysing communication – roughly speaking, if two people do not share (in the strict sense concerning identity) the meaning of "dog", context may help them to adjust their meanings in a way that would enable them to succeed in communication.

[30] *Nota bene*, context looks like a good candidate for the auxiliary in defining that, doesn't it?

[31] I am not going to present an analogous argumentation for the instability within Minimalism with regard to the *possibility of genuine disagreement*, since I find it follows directly on from what I have said about the *miracle of communication* issue – it is enough to imagine a case of two people arguing about the truth-value

So the upshot is that Minimalism – although for other reasons and in a slightly different way than e.g. Contextualism or MH – also confronts the problem of instability. Hence, it is incorrect to say that [in]stability is what proves Minimalism to be incompatible with MH.

3. MINIMALISM AND MEANING HOLISM RECONCILED

In this section I am going to answer separately the question whether Minimalism is compatible with Global Holism, on the one hand, and whether Minimalism is compatible with Local Holism, on the other.

3.1 MINIMALISM AND GLOBAL HOLISM

Let me start with the attempt to figure out where exactly the alleged incompatibility of Minimalism and Global Holism (hereafter: "GH") may be discovered. To be incompatible with Minimalism, GH would have to entail the rejection of either M1 or M2 (or both). First, think of M2 which boils down to the claim that there are *not-many* context-sensitive expressions. At first glance it seems that any significant substantial connection between GH and M2 cannot exist, as the H-principle neither includes nor entails any claim concerning the number of context-sensitive expressions. As far as I am concerned this observation suffices to settle that GH does not entail the rejection of M2, yet I can imagine someone saying that the fact that in the GH credo there is nothing about the number of context-sensitive expressions[32] justifies only the conclusion that within GH the quantity of context-sensitive expressions is not strictly specified (i.e. GH does not assume that there is many of them, not many, few etc.). Hence, one may say, GH still can be ascribed with the view that *all* expressions (to put it plainly: literal meanings of all expressions) are context-sensitive.[33] Let's make the

of "This is a dog". As I tried to show, in the minimalist framework it is extremely difficult for them to use the word "dog" with the same meaning so – along the lines of the *disagreement argument* against Contextualism – they do not genuinely disagree.

[32] Although GH as such boils down to the *general* claim concerning *all* literal meanings (namely that all literal meanings are formatted in a holistic way), yet it does not say anything about these meanings as appearing in contexts.

[33] I ignore here the idea that GH entails that *no* expressions are context-sensitive, since I cannot think of any – even the most extravagant – reasons to believe

effort to take this at face value for a moment. A global holist believes that literal meanings are constituted with regard to how the system they belong to is set up – in other words: with regard to how all other meanings are organised. Thus, it may be said that at the level of constituting literal meanings every meaning is *vulnerable to change* and the change occurs accordingly to the behaviour of the rest of language: if the literal meaning of "animal" changes, this will cause a change in the meaning of e.g. "dog", as well as in every other literal meaning in the language. That is what GH is about – defining / constituting / formatting / etc. of literal meanings – and as such GH does not bother with what happens to literal meanings when they appear in particular contexts.

To endorse the anti-minimalistic view that all expressions are context-sensitive a global holist has to commit themselves to something I call "the principle of *essential changeability*". According to the principle if a meaning undergoes changes at one level it changes at every other level as well.[34] Applied to the case of GH the principle says that if a meaning changes at the level of defining/constituting/formatting, it changes also in contexts. Although there is nothing in the H-principle itself that bans a global holist from accepting the principle of essential changeability, there is also nothing that would force them to endorse it. Concisely speaking, it is not inconsistent for a global holist to accept the minimalist thesis that only some limited group of expressions is sensitive to changes (caused by the strong pragmatic effects) in contexts,[35] while the rest – after being beforehand holistically defined – behaves *stably* in contexts (e.g. the literal meaning of "dog" is holistically formatted as [dog^{l}] and may be said to be invulnerable to any further changes in contexts). In other words, a global holist can accept the claim that in most cases the meanings introduced by tokens to propositions are identical to literal meanings of the relevant types.

---

that it could be the case. The only possible way I can imagine a global holist to commit themselves to the thesis that there is no context-sensitivity in language is some unjustified confession of faith that it is so.

[34] In Kawczyński (2018) I refer to this principle to argue against the view that GH entails or is entailed by Contextualism.

[35] Moreover what GH offers is the general line of explaining *why* these expressions are context-sensitive – namely, due to the fact that language as something that is used to refer to *occasions*, has to contain some occasional parts and these are those parts that we call context-sensitive.

To sum up, GH – when not combined with the principle of essential changeability – does not entail the rejection of M2. GH and Minimalism are compatible with respect to that point.

What about M1? As a matter of fact, what has already been said is enough to leave no doubts that GH does not entail the rejection of M1 either. According to M1 these are the literal meanings of words that enter intuitive propositional content of a sentence used in a context. To paraphrase succinctly what I have explained above: GH is concerned with how literal meanings are constituted, and not with how they function later in particular contexts. That enables global holists to choose between various views concerning the latter issue (i.e. the functioning), among which is the minimalist one, expressed in M1. To entirely eliminate the possibility of endorsing Minimalism, global holists would have to accept the principle of essential changeability – which as pointed out earlier may be an option for a holist, but does not have to be.

To complete the picture of the relations between Minimalism and GH it should be asked whether Minimalism entails the rejection of GH.[36] I think that the answer for this one is quite straightforward: minimalists are not committed to any particular account of the origin of literal meanings and at the same time it is hard to recognise any obstacle for them to endorse the holistic view. What a minimalist maintains is that for the majority of expressions their literal meanings are what constitute the intuitive propositional content – *where those meanings come from* is above their worries. To put it roughly, you can think that literal meanings are *invulnerable to changes* in contexts but before occurring in the particular contexts they are *vulnerable to changes* indeed – e.g. as during their constitution-processes they are holistically adjusted accordingly to the changes of other meanings in a given language.[37]

[36] Actually, the claim that Minimalism does not exclude GH follows (by the rules of logic) from the already-proved conclusion that GH does not entail the rejection of either M1 nor M2. Nevertheless, I am going to offer also some additional independent reasons supporting that claim.

[37] For example, the literal meaning of "dog" may be shaped to be [dog^1] as a result of some holistic processes taking place throughout the whole language, but when someone uses the word in a context, it always expresses that previously-holistically-formatted literal meaning [dog^1].

Again, essentiality of the vulnerability to change is activated when it comes to hindering the reconciliation between Minimalism and GH. This time, however, we should speak of "the principle of *essential unchangeability*": if a meaning cannot undergo changes at one level it cannot undergo changes at any other level as well. Thereby, if a minimalist – i.e. someone, who states that meanings do not undergo changes in contexts – accepts this principle, she cannot accept the holistic view, according to which literal meanings can change at the level of language as a whole. Instead, in such a case she is obliged to accept that literal meanings are unchangeable atoms and each of them is independent of any other. Although this kind of Minimalism is imaginable, it is but one version of the theory, one which assumes much more than the conjunction of M1 and M2. As a matter of fact, such an account can be seen as Minimalism enriched by the explicit *declaration* of endorsing anti-holistic, atomistic semantics.

To conclude then, the arguments I have offered show that Minimalism and GH do not exclude each other, are possible to reconcile and what is more – the theoretical cost of doing so is low for both sides. In addition, the rejection of GH by a minimalist boils down to enriching their theory with rather controversial claims for which it is hard to find a justification that would be independent of the *decision* of avoiding the holistic framework. All in all, a minimalist can reject GH, but if they do it should be considered an extension of Minimalism and not the consequence of its credo.

3.1 MINIMALISM AND LOCAL HOLISM

To discuss the apparent [in]compatibility of Minimalism and Local Holism we must provide the interpretation of the latter theory which would make it possible that Local Holism and Minimalism were on a collision course. Namely, since the minimalist credo is formulated in terms of propositions which are expressed by single sentences, I am also going to consider LH as concerning single sentences, i.e. as the view according to which:

(LH) Meanings of all simple expressions appearing in a sentence, which is uttered in a context, are formatted in a holistic way.[38]

[38] Or in other words: *for all meanings involved in a sentence uttered in a context: meaning of a simple expression depends on meanings of all other expressions*.

Minimalism and LH are thus captured as having the same interests – namely, how literal meanings change in contexts – so the conflict between the theories is genuinely possible.

The point that calls for some clarification is what the processes allowed by the theories in question to modify meanings in contexts are. It is perfectly clear within Minimalism: only the weak pragmatic effects are accepted to modify the literal meanings that eventually become components of propositions. Within LH, on the other hand, it is not precisely defined what the holistic processes determining meanings are. And since the answer to this question will be decisive in judging if LH and Minimalism are compatible, I am going to take a wider look at the issue.

The following nomenclature will be helpful: "STR," "WEAK," "SEM" stand for being respectively a *strong pragmatic* / *weak pragmatic* / *purely semantic*[39] process. The scope of variables in the following statements encompasses all [holistic] processes that take part in formatting literal meanings which eventually become elements of propositions.

(1) Homogenous interpretations. The first three accounts of the nature of the holistic processes are *homogenous* as they assume that all holistic processes are of the same given kind.

(1a) If $\forall x$ STR(x), then LH $\rightarrow \neg$M1.[40]

In this interpretation it is assumed that all holistic processes which affect the literal meanings (to eventually make them constitute a proposition) are the strong pragmatic effects. It is quite explicit that LH interpreted this way is fundamentally incompatible with Minimalism.[41]

[39] I take purely semantic processes to be such that they take part in meaning formatting in contexts without any call to contextual factors. I analyse their nature more broadly at the end of this section.

[40] Whether LH in such a form entails also the rejection of M2 depends on how the difference between the strong and the weak pragmatics is defined. If being sensitive to the strong effects *includes* being sensitive to the weak ones, then LH indeed entails the rejection of M2. If, on the other hand, the strong pragmatics is defined as separate from the weak pragmatics, then LH in the version now discussed assumes that all (which can be assumed to be more than *many*) expressions are sensitive to the former, while not to the latter type of processes and this account does not stand in contradiction to M2.

[41] As a matter of fact such a holism becomes nothing more or less than the Wrong Format View which is the polar opposition of Minimalism (in the field of the theories that accept literal meanings at all).

If all expressions go through holistic processes and these processes are the strong pragmatics, then there is no place in a proposition for literal meanings in their original form – and that is obviously against Minimalism.

(1b) If $\forall x$WEAK(x), then LH $\rightarrow \neg$M2.

When the holistic processes are said to be the weak pragmatic effects, LH does not entail the rejection of M1, however, it still entails the rejection of Minimalism since it is incompatible with M2. It is so, because LH so interpreted assumes that *all* expressions are affected by the weak pragmatics, while according to M2 there is *not-many* expressions vulnerable to the weak effects of that kind.

(1c) If $\forall x$SEM(x), then $\neg$(LH $\rightarrow \neg$M1) & $\neg$(LH $\rightarrow \neg$M2) & $\neg$((M1&M2) $\rightarrow \neg$LH).

It is not a surprise that if the processes allowed by LH were the purely semantic processes exclusively, the theory would be perfectly reconcilable with Minimalism – as Minimalism does not imposes any restrictions upon processes of that kind. As a matter of fact, in such a case LH could be seen as the version of Minimalism which narrows down the general formulation of the theory to the effect that *every* meaning depends (purely semantically) on every other meaning (non-holistic minimalists could maintain that e.g. only a limited number of expressions enter such correlations). However, although the compatibility is beyond a doubt here, it seems more like a purely technical and artificial compatibility than the real substantial possibility of merging the theories in question. The reason why it is so is that the assumption "$\forall x$SEM(x)" is itself highly unpleasant – not only in the framework of LH but in general[42] – since as a result of banning even the weak pragmatic effects from having an impact on propositions, it excludes the possibility of using indexicals. Thus, I think it is sensible to ignore this option in further discussions.

(2) Heterogenous interpretations. Holistic formatting can of course include more than one type of process and what is more – each of these types can have a different share in the whole. Since M2

[42] See footnote 33 above.

concerns "not-many" expressions, to speak of the shares I will also use this highly imprecise quantifier as well as its equally imprecise converse: "many." By combining these two quantifiers with the three possible types of processes we end up with as many as twenty possible interpretations of what LH could be with regard to what the holistic processes are like. However, to reduce the list to only four relevant options it is enough to notice that:

(2a) If $\exists x$STR(x), then LH $\rightarrow \neg$M1

This is actually the stronger variant of (1a). It is enough for LH to allow the strong pragmatic effects at all – no matter to what extent – to entail the rejection of M1 (which does not allow any strong pragmatics) and thereby to be incompatible with Minimalism. Thus, all the interpretations of LH involving the strong pragmatics can be judged as leading to its incompatibility with Minimalism.

The further reduction of the list is possible when we realise that it does not make any difference if we choose to take either MANY or NOT-MANY expressions to be sensitive to purely semantic processes, because exercising both options in the framework of LH keeps the theory compatible with Minimalism (since Minimalism is very friendly to the purely semantic effects – see 1c above). Thus, what we are actually left with are the two following options.

(2b) If MANYx WEAK(x), then LH $\rightarrow \neg$M2.

According to this interpretation, regardless of what else is possibly involved in the holistic formatting of meaning, if MANY expressions are affected by the weak pragmatic effects, it entails the rejection of M2 and thus LH and Minimalism appear incompatible.

(2c) If NOT-MANYx WEAK(x), then $\neg$(LH $\rightarrow \neg$M1) & $\neg$(LH $\rightarrow \neg$M2) & $\neg$((M1&M2) $\rightarrow \neg$LH).

If LH sees meaning formatting as a matter of not-many expressions being affected by the weak pragmatic effects, the theory seems perfectly compatible with Minimalism, since that is exactly what minimalists stand up for.

The upshot is that there are three[43] interpretations of LH on which it appears incompatible with Minimalism (1b, 2a, 2b), and 2c as the only one which makes the theories apparently compatible.

It has to be remembered though, that what these interpretations capture are just logical correlations and nothing has been said so far about the actual possibility of accepting by holists the premises that all those interpretations start from. Even though I think there are good reasons to argue that local holists are not obliged to accept assumptions made in the incompatibility scenarios, I will not discuss it in detail because that would show merely that LH is *not incompatible* with Minimalism, which obviously is not equal to justifying that the theories are indeed compatible. Instead, I will focus on arguing that the compatibility scenario (2c) can be actually endorsed within LH and thus I will give an argument for the claim that LH and Minimalism can be reconciled.

In 2c it is stipulated that not-many of the expressions potentially forming a sentence is sensitive to the weak pragmatic effects. Since LH assumes that all expressions are sensitive to changes governed by the holistic rules, the natural question to ask is what happens to the rest of expressions – i.e. to those that are not vulnerable to the weak pragmatics. The natural answer is that what fills up the domain of holistic processes are the purely semantic factors.[44] The resulting picture would be something along the lines (I use square brackets to speak about meanings):[45] consider a sentence-type of the form "$\alpha\ \beta\ \gamma$", built of three simple expressions, each of which has its literal meaning – respectively: $[\alpha]$, $[\beta]$ and $[\gamma]$. The proposition expressed by this sentence-type may be approximately represented as $[[\alpha][\beta][\gamma]]$.[46] Now imagine that someone utters the sentence (S) "$\alpha\ \beta\ \gamma$" in a context *C*. In accordance with the principle of LH the literal meanings go through the holistic machinery before reaching their destination in a proposition.

[43] I omit here 1a since it is encompassed by the wider case 2a.

[44] The strong pragmatics has already been eliminated in 2a.

[45] The picture I give should be treated only as an aid to understand the idea of LH under the given interpretation. It is a simplification though, as it draws on the assumption that propositions are structured, which is definitely not compulsory for holists.

[46] For the sake of simplicity I ignore here any cases of direct reference and objects rather than meanings being elements of propositions.

The actual proposition expressed by (S) in *C,* composed of meanings after their holistic *check-in,* may be symbolised as [[α*][β*][γ*]].[47] The interpretation of LH under investigation assumes that what is transforming the literal meanings of "α," "β," and "γ" into the meanings appearing in the actual proposition are the weak pragmatic effects [in not-many cases] and purely semantic effects [for what reminds]. So for example in the case discussed let us stipulate that what has turned [α] into [α*] was some weak pragmatic effect, while [β] and [γ] were affected in the purely semantic way.

As clear and straightforward as this account may appear, let us consider if there are any possible reasons that would make it unavailable for local holists and thus make LH incompatible with Minimalism. Apparent vagueness of what are the purely semantic processes may be considered as one such obstacle. I think this is a reasonable doubt because it seems at least somewhat obscure what are those mysterious processes affecting meanings in contexts which at the same time involve no contextual factor. It may seem a bit ridiculous at first but my proposal is to assume that these purely semantic processes do nothing particularly spectacular. To put it plainly, let's take these processes to be *blockers* of the pragmatic effects in the sense that they protect literal meanings from being affected by the (weak or strong) pragmatic effects in contexts. For instance, the purely semantic process which transforms [β] into [β*] boils down to preserving [β] from the pragmatic/contextual influences (analogously in the case of [γ]). The result is that these literal meanings keep their original form when entering propositions, i.e. [β] = [β*] and [γ] = [γ*]. Roughly speaking, in the case of (S) used in *C,* the holistic machinery has determined that the literal meaning of "α" in these circumstances has to be contextually adjusted (and became [α*] which is not identical to[α]), while literal meanings of "β" and "γ" do not require any pragmatic modifications (so that [β] and [γ] as such became components of the proposition expressed by (S) in *C*).[48]

[47] "[*e**]" stands for the meaning expressed by the expression "*e*" as it appears in a given sentence *S* used in a given context *C*.

[48] In a sense it can be said that purely semantic processes are the bridge between types and tokens and are the deep constitutional properties of expressions that make them possible to be used in real communication.

Eventually it appears that LH interpreted as the theory according to which holistic processes formatting meanings consists of the weak pragmatic effects (in not-many cases) and purely semantic processes can be successfully reconciled with Minimalism.

4. CONCLUSION

The three main conclusions from the above investigations are as follows:

I. Contrary to common opinion, Minimalism also faces the problem of so called instability (although for different reasons and in a slightly different manner than e.g. Contextualism or Meaning Holism).

II. It is neither the case that Global Holism entails the rejection of any of the two main minimalist theses, nor that Minimalism leads to the rejection of GH. If such rejections were to actually occur, either global holists or minimalists would have to endorse the principle of essential changeability/unchangeability which are neither compulsory nor the first-choice options for both sides. Instead, the theories are relatively easy to reconcile.

III. Neither Local Holism leads to the rejection of Minimalism, nor the other way round. Furthermore, among numerous possibilities of interpreting what are the holistic processes formatting meaning there is the option that enables us to reconcile the theories in question. Even if some might say that this interpretation is not the most common or favourable variant of holism, it seems quite sensible and is a good way of getting holism closer to the accounts postulating less context-sensitivity than various forms of Contextualism.

All in all, it looks like Meaning Holism is not doomed to play the role of an older sibling to Contextualism as it is possible to reconcile it with its more steady and noble cousin that is Minimalism. I believe that such a position works *for* holism as it makes it a possibly attractive complement to currently fashionable theories.

REFERENCES

Bach, K. (2005). Context ex Machina. In Zoltán Gendler Szabó (ed.), *Semantics Versus Pragmatics* (pp. 15–44). Oxford: Oxford University Press.

Bilgrami, A. (1998). Why Holism is Harmless and Necessary. *Philosophical Perspectives*, *12*, 105–126.

Block, N. (1986). Advertisement for a Semantics for Psychology. *Midwest Studies in Philosophy*, *10*, 615–678.

Block, N. (1994). An Argument for Holism. *Proceedings of the Aristotelian Society*, *95*, 151–170.

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

Borg, E. (2012). *Pursuing Meaning*. Oxford: Oxford University Press.

Brandom, R. B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.

Brandom, R. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge: Harvard University Press.

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

Cappelen, H., Lepore, E. (2006). Précis of Insensitive Semantics. *Philosophy and Phenomenological Research*, *73*, 425–434.

Churchland, P. M. (1993). State-Space Semantics and Meaning Holism. *Philosophy and Phenomenological Research*, *53*, 667–672.

Field, H. (1977). Logic, Meaning, and Conceptual Role. *Journal of Philosophy*, *74*, 379–409.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.

Fodor, J. A., Lepore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Fodor, J. A., Lepore, E. (2002). *The Compositionality Papers*. Oxford: Oxford University Press.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Harman, G. (1993). Meaning Holism Defended. *Grazer Philosophische Studien*, *46*, 163–171.

Jackman, H. (1999). Moderate Holism and the Instability Thesis. *American Philosophical Quarterly*, *36*, 361–369.

Jackman, H. (2017). Meaning Holism. In Edward Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University.

Kawczyński, F. (2018, forthcoming). Meaning Holism and Contextualism[s] – Friends, Foes, or Else?

King, J., Stanley, J. (2005/2007). Semantics, Pragmatics, and the Role of Semantic Content. In *Stanley 2007* (pp. 133–181). Oxford: Oxford University Press.

Kolbel, M. (2004). Faultless Disagreement. *Proceedings of the Aristotelian Society*, *104*, 53–73.

Lormand, E. (1996). How to Be a Meaning Holist. *Journal of Philosophy*, *93*, 51–73.

Odrowąż-Sypniewska, J. (2013). *Kontekstualizm i wyrażenia nieostre*. Warszawa: Semper.

Pagin, P. (2006). Meaning Holism. In Ernest Lepore and Barry C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 213–232). Oxford: Oxford University Press.

Peacocke, C. (1997). Holism. In Bob Hale and Crispin Wright (Eds.), *A Companion to the Philosophy of Language* (pp. 227–247). Oxford: Wiley-Blackwell.

Penco, C. (2001). Local Holism. In P. Bouquet, (Ed.), *Lecture Notes in Artificial Intelligence* (pp. 290–303). Dordrecht: Kluwer.

Récanati, F. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.

Récanati, F. (2010). Truth-Conditional Pragmatics. Oxford: Oxford University Press.

Rovane, C. (2013). The Larger Philosophical Significance of Holism. In Ernest LePore and Kirk Ludwig (Eds.), *A Companion to Donald Davidson* (pp. 395–409). Oxford: Wiley Blackwell.

Stanley, J. (2000). Context and Logical Form. *Linguistics and Philosophy*, *23*, 391–434.

Stanley, J. (2002). Making It Articulated. *Mind and Language*, *17*, 149–168.

Stanley, J. (2005). Semantics in Context. In Gerhard Preyer and Georg Peter (Eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth* (pp. 221–254). Oxford: Oxford University Press.

Stanley, J. (2007). *Language in Context: Selected Essays*. Oxford: Oxford University Press.

Stanley, J., Szabó, Z. G. (2000/2007). On Quantifier Domain Restriction. In *Stanley 2007* (pp. 69–110). Oxford: Oxford University Press.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 77–112
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.05
Rozprawa

Adriano Marques da Silva*

I-SEMANTICS: FOUNDATIONAL QUESTIONS[1]

Summary: What is the scope of a semantic theory consistent with the theoretical assumptions adopted by the generative program? In this paper I will show that the linguistic theory generically known as generative grammar is an extremely coherent Scientific Research Program and within this descriptive framework it's possible to characterize the main features of an I-semantics. First, will be presented the hardcore of the generative program, its heuristics and Chomsky's criticism towards formal semantics. Second, I will compare two approaches: the denotational approach by Larson and Segal and the intensional approach by Paul Pietroski. I argue in favor of Pietroski's approach, because it is more coherent with the core assumptions of the generative program. The main argument is that syntax, in the context of the generative program is explanatory and, in this very context, semantics is not. Therefore, in order to account for the explanatory role of syntax in the generative program it is necessary to review certain foundational assumptions commonly accepted in formal semantics.

Keywords: syntax – semantics interface, generativism, philosophy of linguistics.

1. INTRODUCTION

This paper attempts to answer the following problem: What is the relationship between the notion of an internalized linguistic competence, as conceived by the generative program, and a semantic theory?

* Instituto Federal da Paraíba (Cajazeiras-Brazil). E-mail: adrymarques@gmail.com

[1] I am grateful to the two anonymous referees for the helpful remarks.

More specifically, what is the scope of a semantic theory consistent with the theoretical assumptions adopted by the generative program? Section 2 introduces descriptive concepts to characterize the generative program, Lakato's Methodology of Research Programs. In 2.1, I characterize the hardcore of the generative program. In 2.2, I introduce the methodological foundations of the generative program. Section 3 presents Chomsky's criticisms against extensional semantics. In section 4, I present the extensional approach to I-semantics, proposed by Larson and Seagal. In Section 5, I present the proposal offered by Paul Pietroski. After comparing and contrasting these proposals, I argue in favor of Pietroski's approach, because it is more coherent with the core assumptions of the generative program and it expands the positive heuristics of this research program. The main argument is that syntax, in the context of the generative program is explanatory and, in this context, semantics is not. Therefore, in order to account for the explanatory role of syntax in the generative program it is necessary to review certain foundational assumptions commonly accepted in formal semantics.

2. LAKATOS'S METHODOLOGY

The generative linguistic theory has been presented as a scientific research program (SRP), which requires us to define such term. It was coined by the Hungarian science philosopher Imre Lakatos in 1978. It is related to his notions of the evolution and history of science. To Lakatos, scientific knowledge distinguishes itself from other forms of knowledge because it is structured around a number of untestable propositions (testable as defined by Karl Popper, 1959) that express the basic assumptions on which the theoretical approach is founded. Lakatos named such a set of propositions a hard core. Thus, the hard core is supposedly formed by a set of metaphysical propositions regarded as untestable by methodological decision. Additionally to the hard core, there is the heuristic, a set of methodological procedures that delimit the scientific research according to the program in question. The heuristic selects and organizes the problems and questions to be answered along the research as if it were a work plan, selecting topics for investigation and describing how they will be approached. The heuristic sets the methodological rules that guide empirical research.

The heuristic may be negative if it sets rules that indicate which directions should be avoided in the research and that ban any changes to the hard core of the program (i.e. that control what can be absorbed), or positive if it determines the rules that point towards the directions to be followed, thus controlling the limits beyond which the research can expand. In summary, according to Lakatos, a scientific research program contains a hard core of basic theoretical assumptions, based on which, the approach and even the object of study are defined.

According to Lakatos, a scientific research program progresses as it proposes various theoretical models that are different from each other, because auxiliary hypotheses are formulated during the process of scientific inquiry to handle data as they are collected and adjusted to the model. This is what makes them different from each other. According to Lakatos, having different models should not be a problem, provided that they share a common hard core and heuristic. The advantage in adopting such an approach, and what ensures the success of science in comparison with other types of knowledge, is that new hypotheses replace older ones, and new theoretical propositions originate therefrom.

This does not require one to forgo core objectives and issues. They remain preserved in the hard core and heuristic, which adds flexibility and efficiency to the process of producing scientific knowledge.

In regard to the application of Lakatos's scientific theory to gauge the progress of a theoretical model, there are important points to be made. Firstly, according to Lakatos, if two theories (T1 and T2) are part of a scientific research program, T1 is superior to T2 if T1 has more empirical content, i.e. if it explains more facts that T2, or if T1 has more heuristic power, i.e. if both the facts previously explain by T2 and the new facts explained by T1 receive a more appropriate treatment in terms of descriptive, explicative-predictive, depth and usability potential, as conceived by Ludlow (2011), according to whom the simplicity of a theory is directly linked to the simplicity of its use. In summary, testing a theory is an 'internal' process and is based on its ability to explain more facts in a program, and to do so more efficiently.

Secondly, if T1 is superior to T2, it is evident that there will be a trend to promote T1 to the detriment of T2, which will lead to a number of changes to the heuristic of the program. Such changes

may be of two types: creative changes, which cause the positive heuristic of the program to expand; and degenerative changes, which force changes upon the hard core of the program, or ad hoc changes to the heuristic itself. It is the balance between the quantity and the intensity of each type of change that determines the size and the speed of such alterations, and that provides support for the evaluations. On the other hand, as one might assume, the said evaluations are far less dependent on the subject judgment of the one who conducts them. They are more closely linked to the overall behavior of the program in relation to its developments. Thus, the evaluation of an SRP is performed according to such developments: a program is progressive if its theoretical development predicts its empirical development; or it regressive if its theoretical development is delayed in relation to its empirical development, requiring ad hoc explanations.

Considering Lakatos's propositions as they have been described so far, it is possible to define some parameters to guide a preliminary evaluation. The objects of these evaluations are obviously the components of a scientific research program, namely its hard core and heuristic. From such a perspective, one may ask: what is the hard core and the heuristic of a generative program? What are its characteristics? These questions support the evaluation described in the following sections.

2.1. THE HARD CORE

The hard core of the Generative research program can be summarized in the two following propositions, and a third one can also be added. Without further discussion, I would like to propose that the nucleus of the GG consists of the following statements:

The sentences in a given language are determined, at least in part, by states of the mind/brain, and states can be defined in terms of internalized knowledge of that language, which is rooted in the mental/brain structures of human beings, and it is called I-language:

The nature of these states of mind/brain can be described by theoretical models that represent the computation involved in the generation of the sentences, generating a theory of I-language, which is called grammar (of I-language);

The acquisition and development of internalized knowledge is mostly determined by an innate, biological predisposition, as per

a Universal Grammar.[2] As a branch of cognitive psychology, the generative program has as its object of study the internal cognitive states of the mind/brain of the speaker-listener.[3]

The hard core of a research program is comprised of a set of propositions regarded as untestable, in the Popperian sense, and only assumed. In other words, propositions that are 'metaphysical' and that reveal the point of view that will give the very definition of the object of study, etc. Therefore, one could expect that it cannot be criticized, as the initial assumptions derive from the specific choices made in each research work according to the phenomena it intends to analyze. What generative grammar intends to do is to construct a computing device, capable of forming and transforming representations, that can 'simulate' the linguistic knowledge a speaker of a natural language has in his mind/brain. It is this 'nucleus', constantly present in the history of generative grammar history, that allows us to say we have one and the same research program, in spite of the various deep changes the theoretical device has undergone. The hard core of a generative research program has proved extremely fruitful, as evidenced by the countless relevant research projects that have been conducted on the matter. It is, therefore, at least for the moment, safeguarded from direct attacks against its hard core.

On the other hand, hard-core assumptions must be the only ones that are untestable. Therefore, the other assumptions, the heuristic ones, must be tested and are subject to falsification. The restrictions regarding the number of conjectures and assumptions must be extremely severe, according to what could be considered a 'stricter' interpretation of formal rigor – phenomena must be explained within the boundaries of the hard core and the heuristic. Explanations must strictly fit within the hard core and the heuristic rather than shape them.

---

[2] See Chomsky (2005) about the importance of general laws of nature or computation, outiside the cognitive endowment, to the development of the internalised knowledge.

[3] It is worth remembering that the internalism defended by Chomsky is a methodological perspective in linguistics that does not deny the legitimacy of the study of certain aspects of the E-language. He argues that all linguistic theory (i.e. the study of each level of linguistic articulation) should be identified by the study of these aspects. In addition, he 'suspects' the feasibility of a scientific study of E-language without an in-depth study of I-Language.

According to Boeckx (2006) the Minimalist Program (MP) should be considered as part of the model that resulted from the last major elaboration of the program, the Principle and Parameters Theory. MP would consist of methodological guidelines to help linguists to apply 'Ockham's razor' to the Principles and Parameters Theory (P&P), eliminating what was unnecessary. In order to understand the MP role, we must understand better the working hypothesis of human languages proposed by the GG program. According to Chomsky, languages are biological systems that men use to speak about the world: describe, refer to, ask, communicate with one another, articulate thoughts etc. Those 'things' we do with language constitute what Chomsky calls the conceptual intentional system. On the other hand, as an 'expressive' medium, language must be associated with a production and reception system, of motor-sensorial nature, capable of allowing for the production and reception of sounds that constitute the linguistic expressions. Chomsky labels this second system the articulatory-perceptual system. Thus, the human language must be able to contact (be an interface of) not only the conceptual-intentional system (C-I), but also the articulatory-perceptual system (AP).

The spirit of formal rigor seems to have been incorporated more explicitly by Minimalism, since it assumes that all theory constructs that are not required by the theory should be eliminated and that new propositions should be limited to those that are fully explainable within the context of the theory, i.e. those that are empirically motivated according to the theory. Such an explicit statement suggests that the rigor has not always been construed as described or even maintained in previous stages of Generativism, notably in the Principles and Parameters model (Chomsky 1981). However, the commitment of Minimalists to such rigor still represents progress to a certain extent.

2.2 THE HEURISTIC

Analyzing the heuristic of Generativism involves reflecting upon two central questions: what are the methodological rules employed in the generative program and how capable are they of meeting the requirements of the program's hard core? What is the relationship between such rules and the heuristic assumed by the generative program and

how much have they expanded this heuristic? These questions emerge as it becomes evident that, despite the unequivocal achievements of the generative theory as regards our understanding of human language, there is some tension between heuristic matters and the methodological rules they involve. The methodological rules hitherto adopted by classic Generativism force the expansion of the negative heuristic to the detriment of the expansion of the positive heuristic.

2.2.1 POSITIVE HEURISTIC

The Generative Program aims to formulate a model that is capable of explaining the linguistic phenomena. From its very foundation it opposes the taxonomic concept of linguistics, according to which the purpose of language sciences is solely to observe linguistic data (i.e. statements) and characterize them according to certain taxonomic categories (e.g. words, morphemes, vowels, consonants etc.). From the perspective of scientific research programs, one can affirm that the longevity and originality of the Generative Program derives from its capacity to pose intriguing questions, suggest relationships between apparently trivial phenomena, formulate empirical and complex generalizations and principles that can explain phenomena seemingly unassociated with each other and found in languages that would be considered radically different at first glance. Minimalism chooses negative data as the core *explanandum*.[4] The goal, as Chomsky emphasizes in a number of excerpts, is not creating formal devices that can generate sentences in a particular natural language.

The level of appropriateness of the descriptions and theoretical vocabulary employed is a critical empirical problem, and it requires that the linguistics hypotheses be constantly refined. As the patters observed in a language are compared against other data and languages, it is possible (in principle) to achieve more reliable generalizations and, therefore, formulate linguistic principles that will integrate the theory of the Universal Grammar, the theory about universal linguistic principles. Ludlow (2011) explains that such principles are indispensable for the Generative Program, as they allow a number of problems and phenomena to be unified into one common vocabulary.

[4] To a discussion about the Lakatosian nature of the Minimalist program see Boeckx (2006).

2.2.2. NEGATIVE HEURISTIC

The hypothetico-deductive model provides the formal principles that form the 'protective belt' of the Generative Grammar.

Such principles ultimately indicate the boundaries that limit the expansion of the generative program. This means that the explanations must be formulated by using the formal resources available based on the 'analysis technology' provided by linguistic theory. However, such explanations sometimes depend on 'holistic inferences', i.e. on the linguist's grammaticality judgments. These inferences are holistic because data are usually analyzed by means of informal methods, and not by quantitative, statistical or other mathematical methods beyond the formal model provided by syntactic theory. Empirical generalizations, therefore, play a dual role in linguistic theory: on the one hand, they are used to construct theories; on the other hand, they are used as evidence to confirm theories. The apparently circular nature of this procedure makes it difficult for one to obtain an independent criterion to assess the status of each theory (i.e. the status of the hypotheses assumed by the theories). In other words, it is difficult to distinguish between the phenomenon under analysis and the hypotheses provided to explain it.[5]

If the analysis principles remain uncriticized, the program's negative heuristic is forced to expand. The 'protective belt' of the theory must be expanded so that the principles assumed are maintained in face of evidence to the contrary.

Thus, the program does not expand its positive heuristic or propose principles that can explain linguistic phenomena in an actual, systematic way. As we know, this situation became clear in the 1980s. There was a wave of highly specific, idiosyncratic parameters used as resources to safeguard the principles adopted.

3. CHOMSKY'S CRITICISM OF EXTENSIONAL SEMANTICS

Some examples that are problematic for formal semantics, as traditionally conceived (Chomsky 2000), are provided below:

(1) France is hexagonal and it is a republic.

[5] For a discussion about the reability of data in syntatic research, see Ott (2017).

(2) This administration does too little for the average Brazilian, whose children will inherit the social security deficit.
(3) Hamlet lived with his parents in Denmark.
(4) Unicycles have wheels.

If there is an extensional semantics theory underlying natural languages, then the domain of the entities indicated by lexical items should be, at least in principle, specifiable. It is necessary to assume that the sentence values are determined compositionally and, more importantly, that each sentence value corresponds to a truth value. Vagueness is a huge problem: how can a function be determined if the extension of the predicate is vague, if it cannot be accurately determined? In (1), it seems difficult to pin down the type of entity that could be considered the bearer of the predicates 'to be hexagonal' and 'to be a republic'. It is possible to find contexts in which sentence (1) has a meaning and is true. However, as Chomsky questions, would that suggest that there is something that is hexagonal and a republic? In other words, what is the bearer that is capable of satisfying such distinct properties? Therefore, despite being intelligible, sentence (1) has a meaning that cannot be determined compositionally on a prima facie basis. The problem found in sentence (2) is about specifying what the noun phrase 'the average Brazilian' refers to. What is its extension? How to determine such an extension? Sentence (3) provides an admittedly complex example involving fictional entities. It is possible to find circumstances in which (3) has a meaning and is true. However, what is the truth bearer of the sentence?

Sentence (4) seems to involve counterintuitive consequences, since (4) can only be true if each unicycle has wheels (i.e. more than one wheel). This means to say that the truth conditions seems to authorize instances such as:

(5) John has a unicycle.
(6) Therefore, John's unicycle has wheels.

We intuitively know, however, that sentence (4) will be true if each unicycle has one (and only one) wheel. It is easy to notice that the truth in sentence (4) does not guarantee that sentence (5) is true. On the other hand, the truth in sentence (4) causes sentence (6) to be true:

(7) Cars have wheels.
(8) John has a car.

(9) Thus, John's car has wheels.

Let us analyze the following sentences:

(10) Beavers are mammals.

(11) Beavers build dams.

In (10), the predicate 'to be a mammal' applies to the 'beaver' species. In (11), the same predicate applies to the prototypical group of beavers (i.e. the ones that live in the woods, not in laboratories etc.). This small sample shows the difficulties found when attributing truth conditions to natural language sentences. These examples show one main feature of natural languages: lexical items have a flexible structure, they introduce the vagueness and flexibility that are typical of natural languages. In all examples, determining the truth conditions seems to depend on a heterogeneous set of factors, which causes the attribution of truth conditions in each sentence to vary slightly.

According to Chomsky, the isomorphism between the language and the world commits the semanticist to the existence of exotic entities. The following examples show that there is no correspondence between linguistic categories and ontological categories:

(12) The flaw in the argument is obvious, but it escaped John's attention.

(13) The average family has 2.3 children.

Chomsky's argument is that if there is a bi-univocal relationship between the structure of linguistic items and the entities denoted by them, then the noun phrases 'the flaw in the argument' and 'the average family' presumably denote entities whose ontological status is obscure at best.

Chomsky argues that there is a mismatch between the type of individuation that we intuitively attribute to objects and substances and the type of individuation provided by formal semantics, as shown in the example below.

(14) Val(x, water) = 1 iff $x = H_2O$

Does the lexical item 'water' in the phrase 'The water of the Tietê River' denote the chemical substance H_2O? If not, assuming that the meaning of the phrase is formed by what its lexical items denote seems questionable. Let us suppose that 'water' denotes the chemical substance H_2O in this phrase. We know that a cup of tea

proportionally contains more H_2O than the Tietê River.[6] However, since the lexical item 'tea' does not designate the chemical substance H_2O (but rather a mixture of water and herbs) and, additionally, 'tea' and 'water' are different lexical items (therefore, they have different denotations), one can conclude that the Tietê River waters have proportionally more H_2O than a cup of tea.

Chomsky uses this example to show that the use of the term 'water' depends on a complex set of social conventions, and that the criteria of individuation of this item is not as simple as the formal semanticist would assume. They involve complex application conditions that do not depend on the chemical composition of water (or on the grammatical structure of the lexicon). Chomsky emphasizes that the use of a word can be determined for certain purposes, but that would be a normatization that is not associated with the laws of nature. They are theoretical constructs (e.g. 'matter', 'weight', 'c-command') and, therefore, they do not need to satisfy the myriad of intuitive applications that these terms have in everyday language. 'Water' and 'tea' are not terms from scientific theory. They are items used in our ordinary speech. Their application and identification criteria are vague and variable according to the context. Speakers are the ones who use words and sentences based on their perspectives and intentions. Therefore, there is no nomological relationship between the levels of articulation that form the natural language (phonetics, syntax and semantics) and the world.

As Putnam (1973) emphasized, linguistic items on their own do not suffice to determine whether an object falls under the extension of a concept or not. The link between language and the world is governed by convention. According to Chomsky, the reference relationship is not between language and extra-linguistic objects, as it is not established by the 'linguistic community'. It is mediated by a plethora of intentions, conventions and perspectives. From Chomsky's point of view, this relationship is beyond scientific inquiry. It is even beyond the possibility to provide a coherent description.

[6] As we know, the Tietê River currently contains a wide variety of chemical compounds from the waste discharged in its waters on a daily basis. We also know that the amount of solute in a cup of tea (i.e. the chemical compounds of the tea) is proportionally smaller than the amount of solvent (i.e. water or H_2O, if you will).

Chomsky (2000) discusses a number of points related to contemporary semantics and ponders on the general format of semantics within the generative, Minimalist Program.[7] In this model, the faculty of language (FL) is a computational system (CS) that acts on a lexicon, generating a phonetic form (PF) and a logical form (LF[8]) for each lexical item. PF and LC, in their turn, interface with other cognitive systems. Additionally, there is a Spell-Out operation that feeds the interfaces.

The concept of SEMs[9] is very important in the Minimalist Program. It is a theoretical construct that Chomsky uses to represent the potential semantic perspectives provided by lexical items as they are computed by the faculty of language. SEMs would be the inputs for the conceptual-intentional module, i.e. the reference and categorization perspectives that are available to speakers and that allow them to engage in communicative activities so they are able to deal with the world that surrounds them. These perspectives do not come directly from the world. Rather, they are conceptualization capacities that enable interaction between speakers and the world. They are hypothetical mental entities that translate the linguistic inputs for the conceptual-intensional module. Chomsky is not concerned with the definition of 'meaning'. It is about the contribution of the I-Language (algorithmic procedure internalized by speakers) to the generation of specific human skills. In this sense, SEMs are syntactic entities. They are theoretical descriptions provided by the GG about the linguistic meaning. They include semantic and categorical features, but not the language-world relationship as a theoretical, explanatory term. The explanation of the language-world relationship can be seen as the ultimate ('bold') objective of the semantic theory. It is the inquiry's ideal finish line, but not a starting point. This idea is still in an embryonic stage in the Minimalist Program.

---

[7] See Chomsky (1965, 1965, 1977, 1986) and his criticism towards formal semantics and his view according to wich meanings are generated by and internal to the human mind.

[8] A logical form, according to generative syntax, does not mean a formal sentence with a first-order predicate logic, but rather a conversion of lexical items into structural descriptions that explain the categorical and semantic requirements of each item.

[9] See Ludlow 2003.

Chomsky does not oppose formal semantics, but he criticizes some of its assumptions. His position may be summarized in the following points (Chomsky 2012):

(i) Rejection of the assumption that there is a transparent relationship between the structure of the language and the structure of the world;
(ii) Separation of colloquial terms from scientific ones: science invents its own concepts; it does not discover essences, not of water, mountains or meanings;
(iii) Rejection of ontological questions outside of a well-defined theoretical framework;
(iv) Separation of language from the use of language: for research purposes, certain mathematical properties that are part of the human language are studied (i.e. the I-Language);
(v) Rejection of the autonomy of meaning. Sentences and words do not carry full meanings that can be separated from a complex inter-relation between beliefs, context etc.

D'Agostino (1986) points out that Chomsky's position on the denotational interpretation of semantics is closely related to the empirical and theoretical work developed by him and other collaborators. Such work consists of theses that emerge from reflection upon empirical research. As we can see, the propositions above are indeed, to a greater or lesser extent, associated with the generative program and the methodological guidelines that govern empirical research. Propositions (ii)-(iv) are the direct reflex of a methodological choice: internalism. Propositions (i) and (v), on the other hand, summarize Chomsky's philosophical position on the nature of meaning.

Chomsky's position on semantics absorbs the criticism triggered in the second half of the twentieth century by the pragmatic turn (Taylor 1985), according to which many semantic problems accumulated by the analytic philosophy based on a formal and logical tradition, a tradition that prevailed at the beginning of the twentieth century, were solved or even dissolved by the analysis of the use of language. Chomsky affirms that his conception of 'meaning' is closer to the one defended by Austin and Wittgenstein. In a number of papers and books (Chomsky 1977, 1986, 1995, 2000, 2012) he insists that linguistic items (i.e. words and sentences) are used in a myriad of functions:

referring, describing, affirming, insisting, joking. The use of language is a form of action, and actions are free. They cannot be subjected to scrutiny, although it is possible to describe certain interesting correlations (e.g. the speech act theory). Linguistic items are used in certain circumstances to refer to or denote aspects and events of the world. They are used to represent things, the state of things, circumstances etc. There is no doubt about this point. However, he argues that this does not result in a reference relationship between one linguistic item and the other and between linguistic items and what they refer to. Linguistic items do not refer to anything. People are the ones who refer to something when using linguistic items. It is not necessary or uncontroversial to assume, for instance, that a verb such as 'to sing' corresponds to a unary predicate, which is satisfied by individuals who sing. Furthermore, since it is not possible to determine the extension of vague predicates (i.e. it is not possible to determine the members that 'fall under' the extension of the predicate), interpreting the extension of a predicate as its semantic value seems questionable. The metalinguistic formulas employed in formal semantics would be relegated to the status of 'hybrid expressions', semi-formal paraphrases which do not characterize functions.

Chomsky incorporates such criticism, but he does not accept the assumption that linguistic meaning is determined by use. According to Chomsky, such a perspective means that language acquisition is a process of introjection of a practical skill, and that children simply learn to mimic and reproduce adults' patterns of speech based on the observation of the linguistic behavior of the 'speaking community'. Thus, linguistic competence would be equal to a manifestation of a certain type of social behavior.

Linguistic theory is not expected to provide a 'linguistic creativity science'. Chomsky believes that the science of language has something to say about that. He emphasizes, however, that there is a difference between explaining the creative (intelligent) behavior and explaining what makes that creative behavior possible.

4. I-SEMANTICS: DENOTATIONAL APPROACH

Larson and Segal (1995) support the theory that semantics should be seen as a field of cognitive psychology. They try to include their proposal into the generative program, borrowing not only the formal resources developed by generative grammarians, but also the conceptual assumptions that guide this program.

There are two basic foundational problems related to the formulation of an 'I-semantics'. They consist of the definition of its object of study (i.e. the domain of the research) and the nature of the phenomena encompassed by this approach. Such problems can be formulated as follows:

(P1) What is the object of study?
(P2) What are the theoretical goals?

A possible starting point would be to analyze how grammar influences the organization of semantics. According to Larson:

> As speakers of English, we know facts about syntax: for example, that expressions divide into categories like verb, noun, preposition, and adjective, that verbs and prepositions typically precede their objects in English, that words in a sentence cluster into constituents. In addition, we know facts about the semantics, or meaning structure, of English: that sentences are related as synonymous or contradictory, that they are true under certain circumstances, that certain notions do not correspond to possible worlds (Larson 1995, p. 361)[10]

In order to explain this ability, Larson and Segal claim that it is necessary to assume that speakers have semantic knowledge. The I-semantics, according to them, would have the theoretical goal of explaining the speakers' tacit, internalized semantic knowledge:

> To view the subject matter of semantics as linguistic knowledge is to locate the place of semantic theory within the general enterprise initiated by Noam Chomsky (...) for whom linguistic theory is a theory of real knowledge of speakers. This project contrasts with a variety of other commonly held views of the subject matter. (Larson, Segal 1995, p. 16)

Therefore, the I-Semantics has to explain the knowledge underlying the speakers' semantic competence, the knowledge that makes

[10] Some aspects of the original proposition by Larson and Segal (1995) have been developed in previous papers. Furthermore, it is adopted by Borg (2004) as technical guidance in her defense of semantic minimalism.

speakers able to evaluate the truth condition of a sentence; to investigate whatever gives the speakers the capacity to judge truth or falsehood in the sentences; and to clarify the knowledge underlying these judgments. Semantic theory, in particular, has to demonstrate how strictly semantic aspects, independently from context, determine the truth and satisfiability conditions of every lexical item (i.e. demonstrate that the lexical item 'cat' refers to a cat, and that 'is on the rug' corresponds to an event etc.).

Larson and Segal affirm that the knowledge of the truth conditions of a sentence can be analyzed as an instance of a disquotational scheme:

(6) *The cat is on the rug* is true iff the cat is on the rug.

A semantic theory should infer the technical counterparts of the intuitive semantic judgments, particularly of disquotational truths such as (6). We would then come closer to the idea that the knowledge of the meaning of an S-sentence corresponds to the knowledge of its truth conditions. The sentence to the left, in italic type, is a sentence of the metalanguage denoting a sentence of the object language. To the right we have the truth conditions that must be satisfied in order to make the sentence in the object language true. Therefore, the sentence on the right is a metalinguistic one, which expresses the corresponding state of things designated by the object language. By intuition, if speakers know of this equivalence, they know the meaning of the sentence 'The cat is on the rug'.

'Semantics' and 'Syntax' are usually regarded as fundamentally different fields of inquiry. Such a distinction seems reasonably clear in logical-mathematical terms. 'Syntax', on the one hand, is merely a set of rules of good formation based on primitive symbols. 'Semantics', on the other hand, provides the satisfiability conditions for well-formed sentences. However, this distinction requires the assumption that there is a bi-univocal relationship between structure and meaning. Lewis (1972) introduces the idea that there are <syntax, meaning> pairs and interpretation restrictions: certain meanings are not attributed to certain structures. However, Larson and Segal point out that this picture is incomplete. According to Lewis, 'badly formed sentences' (i.e. 'faulty', 'incomplete' structural descriptions) do not carry any meaning. Lewis assumes that sentences in the natural language are

the counterpart of well-formed formulas. Therefore, if a formula is not well structured, it is not semantically interpreted. By analogy, if a sentence is faulty, it cannot be semantically interpreted. What Larson and Segal introduce is the generative perspective, in which 'faulty' sentences have a meaning, as do their grammatical counterparts. That is to say that a sentence, although 'faulty', is interpretable; therefore, it cannot be excluded from the research.

According to the perspective adopted by Larson and Segal, the I-semantics must infer the sentence-meaning pairs. This idea can be summarized as follows, considering the relationship between sentence '*S*' and proposition '*p*':

(7) *S* means *p*.

There would be a structural representation of *S*, represented by *X*, and it would be possible to conclude that '*X* means *p*'. Similarly, if there are negative data, an interpretation restriction, we would have:

(8) *S* cannot mean *p*.

This would be the same as to say that there is a description of *S*, i.e. *X*, based on which we cannot conclude that '*X* means *p*'. Although they assume the 'work division' between syntacticians and semanticists is analogous, Larson and Segal understand that the faculty of the language contains a module that is specifically semantic, and, therefore, there is a basic distinction between syntax and semantics. Ultimately, the authors aim to characterize the 'semantic module'. It would be responsible, according to Larson and Segal, for the tacit knowledge of the semantic properties and relationships present in the natural language.

Larson and Segal assume that the speakers' capacity to understand sentences comes from a tacit, unconscious knowledge of model-theoretical axioms. According to Larson and Segal, when there is a structural description, *X*, the semantic theory has to demonstrate that the interpretation of such structure is stable, unique and determined by the lexical items that compose it and its structural arrangement.

One could affirm that the objective is similar to that of syntactic theory, since said theory is to demonstrate that lexical items generate a univocal structural description when they are grouped according

to certain compositional principles. Considering grammar has a syntactic and a semantic component (even if in different modules), Larson and Segal modify Lewis's perspective regarding the object of study of semantics. They do not deny that ontological commitments are presumed in everyday speech: the speakers assume the existence of certain objects and relationships between such objects. However, according to Larson and Segal, the semantic theory should not expect from ontology an answer about 'what exists', about the kinds of entities that are referred to, so that it can be regarded as a discipline. On the contrary, they claim semantics is expected to provide, within the limitations arising from the field of research itself, support to the clarification of this kind of inquiry. Larson and Segal suggest that the counterexamples presented by Chomsky neither refute nor are they counterexamples against attempts to build an extensional semantic theory for natural language. They believe these cases should be considered and answered individually. They would not undermine the semantic research, but rather encourage it.

Larson and Segal assume that, when children acquire a language, they acquire the capacity to map linguistic signals onto concepts. They assume there is a bi-univocal connection between linguistic structure and conceptual structure. Therefore, a lexical item such as 'dog', for example, corresponds to the concept of DOG; the syntax combines the lexical items and, as a consequence, it is responsible for combining them with the corresponding concepts. By accepting these assumptions, it is possible to provide a 'psychologicalized version' of the T-schema. The seemingly unsolvable examples are then solved:

(15) France is hexagonal and it is a republic iff FRANCE IS HEXAGONAL AND IT IS A REPUBLIC.

(16) This administration does too little for the average Brazilian, whose children will inherit the social security deficit iff THIS ADMINISTRATION DOES TOO LITTLE FOR THE AVERAGE BRAZILIAN, WHOSE CHILDREN WILL INHERIT THE SOCIAL SECURITY DEFICIT

(17) Hamlet lived with his parents in Denmark iff HAMLET LIVED WITH HIS PARENTS IN DENMARK.

This typological solution is not satisfactory at all. Our encyclopedic knowledge tells us that France is an institution, that it has space-time

coordinates etc. We also know the cartographic representation of the territory of France corresponds to a hexagon, or at least resembles one. Therefore, as Chomsky claims, the lexical item 'France' provides us with certain notions. Some of this item's peculiarities are related to the history of France, others to the speaker-listener's geographic knowledge, or their personal experiences with French cuisine etc. The speakers have beliefs about these peculiarities and can express them in different circumstances. However, it would not be reasonable to expect that a paraphrase such as (1) would be able to refer to all of these uses. From the linguistic standpoint, 'France' is a term of a certain type (N) and has some semantic features (-agentive, +thematic, -animate etc.). Such features allow for certain semantic perspectives. Linguistic expressions trigger representations whose elements do not necessarily coincide with the linguistic structure in which these expressions are applied. Therefore, the psychologicalized version of the T-schema shows little explanatory advantage.

Larson and Segal assume that there is a bifurcation between rules of sentence formation for well-formed sentences in a formal language (syntax) and the formal interpretative resource that provides satisfiability conditions (in a model) of well-formed sentences (semantics). It means that, while building a semantic model, the semanticist has certain pre-theoretical expectations. The indisputably accepted expectation is that there is a dichotomy between syntax and semantics.

This technical distinction (as well as technical notions such as 'reference', 'truth', and 'satisfiability', critically analyzed by Chomsky) cannot be taken as a basic explanatory principle to explain linguistic phenomena. Consider the categorical and semantic constraints imposed by the nature of lexical items such as the following verbs, which are well-known examples given by Chomsky:

(1) John is eager to please.
(2) John is easy to please.
(3) John is eager that he please relevant parties.
(4) John is easy that relevant parties please him.
(5) # John is eager that relevant parties please him.
(6) # John is easy that he please relevant parties.

It is clear that there is an interaction between the meaning of the lexical items and the argumentative and thematic structure that each

one of them imposes. None of these restrictions is explained by the psychologicalized version of the T-schema:

(T) 'to please' is a transitive verb that corresponds to the PLEASE concept;
(T) 'eager' is a predicate that indicates the EAGER concept;
(T) 'easy' is the direct object that indicates the EASY concept.

It is important to emphasize that generative grammars were not conceived for the purpose of generating grammatical or well-formed sentences. They are heuristic tools used by syntacticists to unveil the computational principles underlying the grammatical operations. The phrase 'syntax-semantics interface' can be deceiving, since it suggests that there is a line between the two levels of linguistic articulation: on the one side, there is 'syntax'; on the other side, there is 'semantics'. A robust proposal must explain relevant data and not graft the explanation onto semantic formalism. It seems that there are more complex, interesting phenomena than the so-called 'syntax-semantics interface' appears to suggest:

> In general, one should not expect to be able to delimit a large and complex domain before it has been thoroughly explored. A decision as to the boundary separating syntax and semantics (if there is one) is not a prerequisite for theoretical and descriptive study of syntactic and semantic rules. On the contrary, the problem of delimitation will clearly remain open until these fields are much better understood than they are today. Exactly the same can be said about the boundary separating semantic systems from systems of knowledge and belief. (Chomsky 1965, p. 159)

Chomsky affirms that semantics is a form of syntax. What does that mean? Does that mean to say that semantics can be reduced to syntax? I believe that Chomsky's assertion can be construed as follows: What a formal semanticist truly does is a form of syntax. In other words, the difference between syntax and semantics is purely nominal. As we know, in formal semantics, it is assumed that there is a bifurcation between the rules of formation of well-formed sentences in a formal language (syntax) and the formal interpretative resource that provides satisfiability conditions (in a model) of well-formed sentences (semantics). That means to say that syntax offers the set of interpretable sentences and semantics provides a set of interpreted sentences.

Larson and Segal assume that there is a relatively transparent relationship between the grammatical form and the logical form[11] and a relatively transparent relationship between the logical form and the structure of the world.

The authors defend a certain form of semantic internalism, a psychologicalized version of formal semantics, according to which a semantic theory must be constructed as a system of mental representation of the world. Once this kind of semantic internalism is admitted, it is then necessary to provide answers to a series of problems: how to bring the notion of an I-Language, which is internal, individual and intensional, closer to the notion of reference, which is external to the individual? Is it possible to find referential semantics in the I-Language?

There is an unresolved problem that must be considered: what exactly is the research object of an I-semantics? Although they were nominally committed to the goals of the generative program, Larson and Segal do not consider this an unresolved problem to be examined. Therefore, they assume that semantic theories are extensional theories, i.e. theories about the truth conditions of sentences. Consequently, an I-semantics should be conceived in these terms.

This approach becomes intelligible (or at least coherent) in an E-language, taking on an extensional definition, i.e. assuming, from the very beginning, that the language flows as a finite set, S, composed by a finite number of sound-meaning pairs or well-formed formulas – interpreted formulas (in a model). Thus, the nominal objectives of Larson and Segal (i.e., integrating formal semantics into the generative program) are undermined by these assumptions. This version of I-semantics lies outside the scope of the generative program.

5. I-SEMANTICS: INTENSIONAL APPROACH

In the preface of Events and Semantic Architecture, Pietroski presents, in a clear and concise way, the essential objective of his proposal:

[11] Larson and Segal interpret the 'autonomy of syntax' as a strict division between syntactic competence and semantic competence. They would be different cognitive domains that should be 'connected' and 'mapped' by the semantic theory.

> One of my goals in writing this book (...) has been to get free of some assumptions (...) that a semantic theory for natural language will associate predicates with sets and sentences with true values. (Pietroski 2005a, p.1)

The author explicitly rejects assumptions that are the groundwork of the extensional semantics. He proposes a revision of such assumptions and, consequently, of the objectives that define the semantics of natural languages. However, he does not suggest that the technical tools inherited from the logical-formal tradition should be completely abandoned.

The argumentation used by Pietroski to sustain his proposal can be schematically summarized in the following items:

(A1) Meanings are internal properties of linguistic expressions;
(A2) Meanings are instructions for the construction of concepts;
(A3) Lexicalization consists (at least partially) of a creative process of abstraction.

Although logically independent, (A1)–(A3) are contrary to the assumption that there is a clear relationship between syntactic structure and semantic content. (A1) is a methodological assumption that serves as the modus operandi used by generativism in the study of linguistic phenomena. It should be noted that the semantic internalism defended by Pietroski must be seen as a methodological perspective regarding the study of meaning rather than a thesis about the nature of the semantic content. In the following, I will examine the importance of (A1), how syntax, in the context of the generative program is explanatory and, in this context, semantics is not. Therefore, in order to formulate an I-semantics, it is necessary to review certain foundational assumptions commonly accepted in formal semantics

6. HOMOPHONY RESTRICTIONS AND NEGATIVE DATA

In Pietroski (2005b), the author tries to explain what generativism can teach us about the nature of meaning and the nature of semantic theories. The answer offered by generativism is essentially negative: it teaches us what meaning cannot be, certain interpretative restrictions that deserve attention.[12] Let us see a summary of his argument. The

[12] One should notice that this does not mean, as Larson and Segal assume, that the structure descriptions provided by the generative grammar are incor-

generative perspective, Pietroski claims, does not accept the hypothesis that a natural language associates linguistic items to model-theoretical conditions of truth, because the I-Language imposes only a few intelligibility restrictions. It imposes a format for possible, but not achievable, constructions in natural languages. In summary, negative data are related to the relationship between word sequences and their intelligibility, between sound and meaning. According to Pietroski, it is necessary to explain these patterns. There are two kinds of restriction :

(i) The faculty of language imposes restrictions that are independent from any limitations imposed by other cognitive systems;
(ii) The source of these restrictions is in the interface between the faculty of language and other cognitive systems (perception, conceptualization etc.).

As an example of a restriction of type (ii), there are restrictions in the processing of connected sentences (there is a limit, imposed by the working memory, to the number of adjuncts that can be processed), cacophony, states of language (interface between the computational system and the articulatory system) etc. As an example of a type (i) restriction, Pietroski provides the following:

(1) The senator called the millionaire from Brasília.
(2) The senator called the millionaire, and the millionaire was from Brasília.
(3) The senator called the millionaire, and the call was (made) from Brasília.
(4) # The senator called the millionaire, and the senator was from Brasília.

Here we have negative data, Pietroski explains: sentence (1) has the meaning indicated in (2) or (3), but it does not have the meaning suggested in sentence (4). That means sentence (1) admits a finite

porated *only* as empirical evidence for the semantic theory or only as an analysis technique for sentences in the object language. On the contrary, they are about taking into consideration the generative *principles* responsible for the combinatory restrictions. In short: it is not about the application of the analysis technique when examining *particular* cases. If that were the case, such restrictions would be of little interest. Grammar, as Larson and Segal tacitly assume, would serve as a boundary for the set of *sentences* of the object language that can be subjected to semantic interpretation.

number of paraphrases, i.e. there is a limited number of ways in which the structure of constituents can be arranged. Sentence (1) has the meaning indicated in paraphrases (2) and (3), but not in paraphrase (4).[13] At first glance, it seems possible to build an algorithm capable of associating sentence (1) to paraphrase (4). However, as Pietroski emphasizes, it is an empirical phenomenon, and speakers-listeners do not make such an association.

The structure of constituents that originates (2) and (3) respectively corresponds to:

(1') {[The senator] [called [the millionaire [from Brasília]]].}
(1'') {[The senator] [[called [the millionaire]] [from Brasília]].}

Therefore: in (1'), 'from Brasília' is an adjunct to the phrase 'the millionaire'; in (1''), 'from Brasília' is an adjunct to the verb phrase 'called the millionaire'. The ambiguity in (1) is the result of two different scope relationships, but not of three or four or four hundred.[14] This example indicates that the ambiguity phenomenon has several degrees. It is clear that the ambiguity in (1) is resolved when there is a context, but can we say that the structural ambiguity phenomenon does not exist, that it is not a phenomenon that requires explanation? The brackets in (1') and (1'') indicate that the homophony phenomenon (expressions that contain the same phonological representation, but different meanings, such as 'bank' [financial institution; slope of land adjoining a river]) is subject to structure restrictions that are very specific and elaborate. Pietroski defends this as a good starting point for an I-Semantics, since it is a ubiquitous phenomenon, found in every natural language and in a large number of sentence constructions. Pietroski also emphasizes that there are no 'maximally homophonous' languages, i.e. although it is logically possible that there is a potentially infinite number of homophones in every language (e.g. alternative pairs of brackets), homophony is subject to restrictions (as in sentence (4), for example). Homophony (sounds that have different structural descriptions) is restricted. Not

[13] It is possible to infer that this semantic *relationship* is also subject to constraints, because sentence (1) does not lead to sentence (4).

[14] It is curious to notice that it is possible to find out what an ambiguous sentence *cannot* mean. However, it is virtually impossible to exhaustively specify what a sentence *can* mean, in every possible context.

all combinations are homophonous in the same way. That is especially the case for 'negative data', the meanings that are not allowed by the sentences. This example shows that ambiguity happens in different degrees: sentence (1) is ambiguous because it is related to two interpretations, but not three, thirty or three hundred. In other words, certain structures express a number n of meanings, but not $n + 1$.

(5) # The called the millionaire senator from Brasília.

It is important to notice that sentence (4), which has a different meaning, also expresses a coherent thought. It does not have/show the deviancy as in sentence (5). There is not a semantic or pragmatic restriction that bans paraphrase (4). It is a structural constraint. If by saying (4) the speaker actually meant (2), we would have understood 'what he meant' and also noticed that the speaker 'expressed himself/herself badly'. One could claim that, in the example above, the interpretative restriction is produced by the predicate 'to call', which would be semantically satisfied by $<\alpha, \beta, \upsilon>$, an ordered triple, being two individuals and one space-time location: <millionaire, senator, from Brasília>. However, this is an empirical hypothesis about the example analyzed. Thus, it should not be regarded as a phenomenon that semantic theories must explain. Moreover, from a strictly extensional standpoint, it does not matter how functions are specified. Once the interpretations are extensionally individuated, it does not matter what the pairing procedure is, only which pairs are generated. Therefore, the difference between (1) and (4) would be simply extensional.

These phenomena are a strong indication that syntax mediates the connection between sound and meaning, but it does not determine such connections. From a technical perspective (from the formal repertoire of first-order predicate logic, for example), on the other hand, syntax determines the interpretations, because it provides the model with a set of potentially interpretable formulas. It is an idealization that is inadequate to the study of the semantics of natural languages.

Another well-known phenomenon in the generative literature is anaphora. Consider the following sentences:

(5) John said he is going out.
(6) He said John is going out.

If we were to ask: who is going out? In (6), the answer is clear: John. However, in (5), if we were to ask the same question, would the answer be John? It is impossible to determine. These restrictions produce interpretative semantic effects.

The generative program has made explicit that 'negative data' include the phenomena that involve the complex interrelation between linguistic items. The 'island' phenomenon, for instance, suggests that there is a complex interaction between displacement operations and syntactic configuration. We can, for example, use the phrase from the sentence 'John wrote a book' to generate the sentence 'This is the book John wrote'. However, if we try to take the same phrase from the sentence:

(7) John married the woman who wrote a book.

We will generate an ungrammatical sentence:

(8) * This is the book which John married the woman who wrote___

As Stainton (2006) claims, the grammatical relationships generate very important interpretation restrictions. The semantics of verb phrases, for example, is subject to the thematic relationships imposed by the syntactic structure:

(9) Caesar killed Brutus.
(10) The toast pressed charges at the police station.

The distribution of thematic roles is clear: the external argument of the verb 'to kill' plays an active role, while the internal argument, 'Brutus', plays a passive role. I believe these examples are very suggestive, since they show that interpretation restrictions can violate our encyclopedic knowledge (we know Brutus killed Caesar, not the other way around) and our beliefs (we know toasts cannot press charges and dogs do not know how to do their taxes etc.). Despite all we know about toasts, animate and inanimate beings, the most reasonable meaning is not the one expressed in the sentences, but rather the 'bizarre' meaning. This example shows even more clearly that the interpretation restrictions imposed by language might not respect our beliefs, i.e. they do not have the meaning we (our common sense) would expect.

Chomsky's work always analyzed interpretation restrictions, even though he did not name that study 'semantics'. In fact, the examples provided are a small sample of a solid pattern, a converging result originating from many lines of empirical research conducted in the past 50 years: slight differences in the pattern of the connection of constituents produce considerable interpretative effects. The examples Pietroski offers are not used as rhetorical pieces to confirm a specific semantic hypothesis; they are a small sample of a phenomenon typical of natural language.

The examples also show that the notions of grammaticality and intelligibility do not coincide, but that they interact in a complex, intricate way. It is customary to assume that 'semantics' and 'syntax' are fundamentally distinct fields of research. This distinction seems reasonably clear in terms of logics and mathematics, in which syntax is merely a set rules for good formation, based on primitive symbols of the system, and 'semantics' provides satisfiability conditions, in a model, for well-formed sentences. According to this technical definition, syntax provides a set of structures that can be interpreted. However, in order for such a distinction to be valid, one must admit that there is a function that relates the 'semantics' and 'syntax' domains, that there is a bi-univocal relationship between structure and meaning. In a model, it is neither possible to correlate different interpretations to the same well-formed formula, nor to assign the same interpretation to different formulas. Logical systems do not tolerate ambiguity. In logical systems, syntax and semantics are different, unlike what happens in natural languages. Chomsky's criticism is directed at how this technical division influences the usual conceptions about natural languages and, more importantly, their semantics. As we know, even sentences considered semantically 'strange' are accepted and considered grammatical. This is a phenomenon the grammar theory cannot ignore. It suggests that the notion of grammaticality is granular; it does not coincide with the technical distinction between well-formed sentences versus badly-formed ones. The formal system prevents the formation of ill-formed (regardless of the 'level' of bad formation), ambiguous sentences etc. The goal of GG has been, from its very beginning, to explain computational restrictions to the pairing of sound and meaning. Certain sequences are not recognized by the speaker as sentences because they violate the structural

principles that govern the connection of constituents, and not because the word sequences are random.

These structural constrains underpin the interpretation of the sentences. It constrains, but does not determine, the meaning of the sentences. I believe this must be emphasized, because restricted homophony provides negative data; it is a *phenomenon* that, justifiably, must be taken into consideration in a semantic theory of natural languages. I believe that, in this field of research, semantic phenomena are the explanatory context of an I-semantics. The restricted homophony phenomenon clearly shows that the notions of well-formed formulas and grammaticality do not coincide; they are not even analogous: grammatical sentences may have unexpected, 'atypical' meanings. The distinction between syntax and semantics, in natural languages, does not follow the technical distinction adopted in logical-mathematical language.

Considering that negative data are the essential object of analysis of the generative approach, how should the objective of an I-semantics be described? What is the role of the logical-semantic metalanguage in the explanation of negative data? Should they be explained in terms of truth conditions, satisfiability and reference?

7. INTERNALISM AS METHODOLOGICAL STRATEGY

The core assumption made by Pietroski to sustain his proposal can be summarized in the following item:

(A1) Meanings are internal properties of linguistic expressions.

In this section, I am going to maintain that (A1) is a methodological assumption associated with the modus operandi of generativism in the investigation of linguistic phenomena. It should be noted that the semantic internalism defended by Pietroski must be seen as a methodological perspective, and not as a thesis about the nature of meaning. From this standpoint, semantic internalism is the perspective according to which linguistic theories do not assume or imply the existence of objects and properties external to the cognitive state of speakers. In short, Internalism is a conjecture about the object of study of the language science, rather than a doctrine about the nature of linguistic meaning. I believe this is the interpretative key to adequately

understand Chomsky's considerations about semantics and characterize the internalism defended by Pietroski.

There are, as discussed in the previous sections, certain interesting, not random, patterns, highly specific phenomena that are far from trivial, which are found in the interaction between the internalized combinatory system and lexical (categorical and functional) properties. The most important point is that these phenomena provide methodological suggestions to semantic inquiry. It is indisputable that certain semantic phenomena depend prominently and decisively on the speakers' system of beliefs and world knowledge. Therefore, we assume that the speakers' knowledge varies considerably. However, as Larson and Segal speculate, if speakers of different levels of knowledge and education experience a certain class of semantic phenomena in a systematic, regular and (relatively) uniform way, the source of such knowledge and its basic mechanisms must be explained. If in the course of language acquisition all children achieve the same kind of generalization, the same kind of semantic-structural knowledge, despite dialectal, cultural and idiosyncratic differences, we will see a version of the argument of poverty of stimulus (POS) applied to 'meaning', to semantics. In other words, if it is possible to identify the features that characterize semantic knowledge, the problem of poverty of stimulus is once more inserted into the semantics domain. It would be possible to formulate the problem in a coherent and precise way. Cook (2007), for instance, affirms that in no natural language does the determinant lexicalize equinumerosity. There is no quantifier of the 'Equi' type, so that in the sentence [Equi] [child] [had] [slice of pizza] = the number of children and number of slices of pizza is the same. That means to say that there is an equivalence relationship between the members of the set denoted by 'child' and by 'slice of pizza.' In order to express equinumerosity, one must use circumlocution, a complex sentence, such as: 'for every child and every slice of pizza, each child had one, and only one, slice of pizza.'

Chomsky (2000) reminds us that linguistics has a large volume of accumulated knowledge about morphology and syntax, i.e. about how words are formed and what joins them together in syntactic structures. He also reminds us, on the other hand, that very little is known about concepts. The most natural strategy in this case, he claims, would be to start with what we know and, then, expand the explanatory power of

the theory.[15] In other words, it is necessary to know if the phenomena to be analyzed form a unit, if they contain a more basic phenomenon, if they form a cohesive set of phenomena.

As in syntax studies, we would have a general methodological principle, the order into which the research is to be broken down, namely:

(i) Analyzing what is implemented (i.e. the format of the representations);
(ii) Investigating how the representations are implemented.

The study of meaning within the generative program would adopt an internalist strategy. At first glance, we can define it as a methodological precept that can be condensed into the following 'advice': observe how the internalized computational component produces sentences, i.e. sound-meaning pairs. According to this perspective, we must admit, from the beginning, that we know little about what, on the one hand, linguistic expressions are and what, on the other hand, the meanings are. Similarly, the internalism defended by Pietroski can be defined as a methodological precept: as far as semantic theory cannot say anything interesting about the relata in question (i.e. linguistic expressions and meanings), it seems hasty to tackle the relationship between statements (i.e. acts in which the signifier and signified are condensed together) and truth conditions. In short, it seems appropriate to start with stage (i), with the analysis of the properties of the phenomenon analyzed. In semantics, as in the case of syntax, the difference between the types of structure found in the several natural languages is expected to be relatively small, since it is accepted, by hypothesis, that the possibilities, the available alternatives in the course of acquisition are biologically determined. Ultimately, the structures should not violate any of the UG principles.

Pietroski breaks the factors that affect the truth conditions of a sentence down into two broad categories:

[15] A brief look at a manual of formal semantics will reveal that this caution is not shared by most semanticists. Judging by the topics contained in manuals, the formal study of natural language semantics is in an extraordinarily advanced stage and dismisses any methodological or metatheoretical considerations of this kind.

(i) Structural properties of linguistic expressions;
(ii) Extra-linguistic factors concerning the communicative situation/ context under analysis (e.g. general assertability conditions).

In the first one, the properties are formally analyzable and have been investigated throughout the Generative Program. As for the second category, they concern a heterogeneous set of factors, presumably less tractable ones. Chomsky has never defined, in his considerations about semantics, the term 'meaning'. He assumes that linguistic expressions have a meaning; or rather, that meaning is a property that linguistic expressions possess (just as they possess phonetic, morphological, and syntactic properties, among others). It seems to me that this is not a problem, but rather a virtue, because it does not mean that one should deny that words have an open texture, moldable to the communicative intentions, or to the perspective of the speaker etc.

Once it is determined that semantics deals with linguistic properties, the generic, most general question about meaning becomes expendable ('what is meaning'?). Whatever meaning is, according to Chomsky's perspective, it is necessary to explain certain non-trivial phenomena, certain properties and relations.

Chomsky and Pietroski remind us that a massive number of heterogeneous elements determine the truth conditions of an assertion. Nonetheless, one can undoubtedly assume that technical terms used in metalanguage (e.g. 'valuation', 'satisfiability' etc.) have an empirically discernible counterpart, instantiated in the use of language. Indubitably, there is a relationship between meaning and use, but it is much more tenuous and intricate than is usually assumed. However, the problems associated with the relationship between content and intension are complex, and extremely difficult to solve. Chomsky would say they are mysteries. It does not seem clear why the hypotheses about these problems should be taken as criteria to decide upon and evaluate theories about the semantic phenomena found in natural languages.

The attempt to establish a systematic semantic theory that is capable of explaining the complex language-action-world triangulation seems to face insurmountable difficulties. However, it is important to clarify that Pietroski does not oppose extensional semantic theories: "I fully endorse the strategy of supposing that the core semantic notions are extensional" (Pietroski 2005b, p. 287). More specifically: "...I am not

objecting that theories of meaning can be formulated in a metalanguage by an extensional logic" (*idem*, p. 285).

Pietroski argues that resorting to truth conditions is a useful heuristic resource, an idealization adopted for research purposes. The author treats the problem as a metatheoretical question. It is useful to treat semantic theories as theories of truth. From a methodological point of view, it is a somewhat advantageous idealization. However, his argument is that adopting this assumption is not independently motivated. Additionally, one might say that his project intends to prove that, in practice, the denotational interpretation of logical-semantic metalanguage can be disregarded without explanatory losses. The idea is not to abandon formal semantic theories, but to reassess their value and explanatory role. It is legitimate to use the formal apparatus of model theory to describe certain semantic phenomena, but this does not mean that such phenomena should be understood from the standpoint of the metalanguage used to describe them. With this perspective in mind, we progressively (i.e., as we become able to deal with complex semantic phenomena) re-evaluate the role of semantic theories.

It is about the methodological decision to combine the heuristic of the Generative Program with the resources offered by semantic models, rather than the incorporation of heterodox conceptual assumptions into the core of basic assumptions that comprise generativism (as proposed by Larson and Segal). The extensional apparatus may help semanticists elucidate certain lexical properties of linguistic items, i.e. properties derived from the interaction between the way syntax connects constituents and their features. Pietroski believes that a modest Davidsonian typology is necessary (indispensable, in some cases) in the semantic analysis, but semanticists do not need to believe that the semantic model postulated by him associates sentences with states of things. That is, the typology employed in semantic theories should not confuse semanticists, leading them to believe that linguistic expressions denote semantic values. Formal metalanguage offers hypotheses, which can be reviewed, on the semantic properties of predicates. It is a necessary idealization for a rigorous investigation of complex semantic phenomena.

This framework generates semantic types and allows us to accurately codify hypotheses on semantic composition in natural languages.

That is, the formal apparatus allows for the refinement of the distinctions between what is said, what is asserted, and the meaning of the words. However, by definition, these aspects are not within the scope of the internalist study promoted by the Generative Program.

Methodological internalism is a strategy consisting of:

(i) Turning the focus of the research away from the language-world relationship;
(ii) Investigating the internal properties of sentences (semantic typology).

From this methodological perspective, the research should not start from general assumptions about the meaning in order to draw conclusions about natural language. On the contrary, the starting point is observing the phenomena ubiquitously present in human language (e.g. restricted homophony, recursion etc.). From this point, an attempt is made to devise a research plan and, gradually, the outlines of an I-semantics will be delineated.

Pietroski's proposal constitutes a research program of a generative nature because it is consistent with the methodological internalism adopted by the generative program and it offers explanations for phenomena that are dear to the generative program, such as negative data. Pietroski's proposal offers an expansion of the positive heuristic of the generative program. We can say that Pietroski's hypotheses do not underestimate the interpretative effects triggered by syntax, but they also do not overestimate the relation between truth and meaning. The expansion of the positive heuristic of the generative program happens at the expense of the revision of the assumptions of formal semantics. Pietroski offers an alternative hypothesis about the semantic composition of natural languages.

8. CONCLUSION

In this article my aim was to characterize an I-semantics, to define the scope of a semantic theory consistent with the theoretical assumptions adopted by the generative program. In 2.1 was presented the hardcore of the generative program. In 2.2, I introduced the methodological foundations of the generative program. Section 3 showed Chomsky's criticisms against extensional semantics. In section 4,

I explained the extensional approach to I-semantics, proposed by Larson and Segal. In Section 5, I explained the proposal offered by Paul Pietroski. After comparing and contrasting these proposals, I argued in favor of Pietroski's proposal, because it is more consistent with the main assumptions of the generative program, and it expands the positive heuristics of the generative program, while the extensional proposal does not expand the positive heuristics of this research program. The main argument was that syntax, in the context of the generative program is explanatory and, in this context, semantics is not. In order to account for the explanatory role of syntax in the generative program it is necessary to review certain foundational assumptions commonly accepted in formal semantics.

REFERENCES

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

Boeckx, C. (2006). *Linguistic Minimalism*. Oxford: Oxford University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1977). *Essays on Form and Interpretation*. New York: North Holland.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht, Holland: Foris Publications.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (1995). Language and Nature. *Mind*, 104, 1–61.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2005). Three Factors in Language Design. *Linguistic Inquiry*, *36*(1), 1–22.

Chomsky, N. (2012). *The Science of Language: Interviews with James McGilvray*. Cambridge: Cambridge University Press.

Cook, V. (2007). *Chomsky's Universal Grammar*. *An Introduction*. Malden, MA: Blackwell Publishers.

D'Agostino, F. (1986). *Chomsky's System of Ideas*. Oxford: Clarendon Press.

Heim, I., Kratzer, A. (1998). *Semantics in Generative Grammar*. Cambridge: Blackwell Publishers.

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.

Larson, R. (1995). *Semantics*. In L. Gleitman (Ed.), *Invitation to Cognitive Science, Vol. 1. Language* (pp. 361–380). Cambridge: MIT Press.

Larson, R., Segal, G. (1995). *Knowledge of Meaning: an Introduction to Semantic Theory*. Cambridge: MIT Press.

Lewis, D. (1972). General Semantics. In D. Davidson and G. Harman (Eds.), *Semantics of Natural Language* (pp. 169–218). Dordrecht–Boston: D. Reidel.

Ludlow, P. (2003). Referential Semantics for I-Languages? In L. Antony and N. Hornstein (Eds.), *Chomsky and his Critics* (pp. 140–161). Malden, MA: Blackwell Pub.

Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. Oxford: Oxford University Press.

Ott, D. (2017). Strong Generative Capacity and the Empirical Base of Linguistic Theory. *Frontiers of Psychology*, *8*, 1617.

Pietroski, P. M. (2002). Function and Concatenation. In G. Preyer and G. Peter (Eds.), *Logical Form and Language* (pp. 91–131). Oxford: Clarendon Press.

Pietroski, P. M. (2003a). Semantics and Metaphysics of Events. In K. Ludwig (Ed.), *Contemporary Philosophy in Focus: on Davidson* (pp. 137–162). Cambridge University Press.

Pietroski, P. M. (2003b). The Character of Natural Language Semantics. In A. Barber (Ed.), *Epistemology of Language* (pp. 217–256). Oxford: Oxford University Press.

Pietroski, P. M. (2005a). *Events and Semantic Architecture*. Oxford: Oxford University Press.

Pietroski, P. M. (2005b). Meaning Before Truth. In G. Preyer and G. Peters (Eds.), *Contextualism in Philosophy* (pp. 253–300). Oxford: Oxford University Press.

Pietroski, P. M. (2005c). Meaning Before Truth Conditions. Handout; available in: http://www.terpconnect.umd.edu/~pietro/research/papers/Arizona1.pdf

Pietroski, P. M. (2006a). Character Before Content. In J. Thomson and A. Byrne (Eds.), *Content and Modality: Themes from the Philosophy of Robert Stalnaker* (pp. 34–60). Oxford: Oxford University Press.

Pietroski, P. M. (2006b). Interpreting Concatenation and Concatenates. *Philosophical Issues*, *16*, 221–245.

Pietroski, P. M. (2007). Systematicity via Monadicity. *Croatian Journal of Philosophy*, *7*, 343–374.

Pietroski, P. M. (2008). Minimalist Meaning, Internalist Interpretation. *Biolinguistics*, *4*, 317–341.

Pietroski, P. M. (2010). Concepts, Meanings, and Truth: First Nature, Second Nature, and Hard Work. *Mind and Language*, *25*, 247–278.

Pietroski, P. M. (2010b). *Describing I-junctions*. Not published; available on: http://terpconnect.umd.edu/~pietro/research/papers/ijunctions.pdf

Pietroski, P. M. (2011). Minimal Semantic Instructions. In: C. Boeckx (ed.). *The Oxford Handbook on Linguistic Minimalism* (pp. 472–498). Oxford: Oxford University Press.

Pietroski, P. M. (2012). Semantic Monadicity with Conceptual Polyadicity. In M. Werning, W. Hinzen, and E. Machery (Eds.), *The Oxford Handbook of Compositionality* (pp. 129-148). Oxford: Oxford University Press.

Pietroski, P. M. (2015). Framing Event Variables. *Erkenntnis*, *80*, 31–60.

Pietroski, P. M. (2017). Semantic Internalism. In J. McGilvray (Ed.), *The Cambridge Companion to Chomsky* (pp. 196–216), Cambridge: Cambridge University Press.

Pietroski, P. M., Hornstein, N. (2002). Does Every Sentence Like this Contain a Scope Ambiguity? In W. Hinzen and H. Rott (Eds.), *Belief in Meaning: Essays at the Interface* (pp. 43–72). Frankfurt: Hansel-Hohenhausen.

Pietroski, P. M., Hornstein, N. (2009). Basic Operations: Minimal Syntax-Semantics. *Catalan Journal of Linguistics*, *8*, 113–139.

Putnam, H. (1973). Meaning and Reference. *Journal of Philosophy*, *70*, 699–711.

Stainton, R. J. (2006). Meaning and Reference: Some Chomskian Themes. In E. Lepore and B. C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 913–940). Oxford: Oxford University Press.

Taylor, C. (1985). *Philosophical Papers. Vol. 1: Human Agency and Language*. Cambridge: Cambridge University Press.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 113–138
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.06
Rozprawa

Joanna Klimczyk*

'OUGHT', AGENTS AND AMBIGUITY THAT MATTERS[1]

Summary: According to a well-homed view in linguistic semantics, deontic logic and logic of agency, some 'ought' sentences, like 'Kate ought to write the report', are ambiguous between the so-called *agentive* sense as when Kate is the agent of writing the report, and the non-agentive, or evaluative sense as when, in the light of some norm or things being ideal, the proposition that Kate writes the report would come out true. Within this approach to the semantics of 'ought', the ambiguity in question is not due to any semantic ambiguity of the word 'ought', but the ambiguity traced to *Kate writes the report*. We may call the view in question, after Schroeder, the *agency-in-the-prejacent theory*, or APT for short. APT's explanation of ambiguity has been put under heavy criticism by Mark Schroeder's 2011 influential paper. Schroeder tried to undermine APT by exposing its central theoretical drawbacks, their being: (i) that APT badly overgeneralizes because if ambiguity is in *Kate writes the report*, then it should equally well be preserved under the non-agentive interpretation of 'Kate ought to write the report', but it is not, and (ii) that APT also undergeneralizes, since it 'inscribes' the same ambiguity as observed in 'Kate ought to write the report' to a sentence that lacks it, e.g. 'Bill ought to

* Polish Academy of Sciences, Institute of Philosophy and Sociology. E-mail: jklimczyk@ifispan.waw.pl

[1] Thanks to Teemu Toppinen for helpful remarks on an earlier version of this paper and to the anonymous referee for their extensive and helpful comments. The paper is written under the support of an OPUS 12 grant funded by the National Science Centre on the project 'Substantive semantics for normative language (on the basis of the analysis of 'ought' sentences). Decision no. DEC-2016/23/B/HS1/02921.

kiss Lucy'. I argue that both the 'overgeneralization problem' and the 'undergeneralization problem' are harmless for the criticized view, since Schroeder's two central arguments against the respective problems are seriously defective. Also, the third problem identified by Schroeder, that APT cannot accommodate the deliberative sense of 'ought', is mistargeted. I argue that identifying the *salient* property of the deliberative ought is crucial for assessing whether APT is able to accommodate it or not, and that Schroeder failed to recognize this properly.

KEYWORDS: 'ought', deliberative 'ought', Schroeder, agency, authorship, sentential ambiguity

1. INTRODUCTION

It is a common view in linguistics that some noun-phrase-verb-phrase (NP-VP) sentences are *agential*, in the sense that they inform us that some agent *does* something, whereas others are merely *circumstantial*, in the sense that they simply say that something *happens* to someone. According to this view, a sentence such as 'Kate gets a sun hat' is ambiguous between a reading on which Kate is the agent of getting a sun hat and the reading on which Kate is merely a *patient*, or better an *experiencer* in the relation between Kate, the subject, and the getting of a sun hat. Many have thought that allowing for sentential ambiguity is *all* we need to successfully account for the ambiguity of troublesome '*ought*'-sentences, that is *some* sentences of the form '*S* ought to φ' such as 'Larry ought to win the lottery'.

Among adherents of this view are prominent philosophically-oriented logicians such as Nuel Belnap and John Horty. These authors hold that because some sentences (expressed in agential grammar[2] by default) such as 'Kate gets a sun hat' or 'Larry wins the lottery' are ambiguous as prejacent[3] sentences, applying a deontic operator 'ought'

[2] 'Agential grammar' is my phrase, which I use as shorthand for the general syntactic structure that makes room for a subject-argument place. Briefly, I use 'agential' to mean agentive in grammar and 'agentive' when refering to the content of agential 'ought' sentences.

[3] 'Prejacent' is a technical term in modal logic and modal semantics, and means the proposition embedded under a modal operator. In broader linguistic contexts, 'prejacent' is construed as the meaning of the sentence, or proposition. See von Fintel 2006.

to them makes the respective *ought*-sentences no less ambiguous. So far so good. If the sentence 'Larry wins the lottery' strikes us as being *grammatically* ambiguous with respect to whether Larry is the agent of winning, or the patient of a happy arrangement of things that render him the lottery winner, it is natural to predict further that the agential 'ought' sentence 'Larry ought to win the lottery' will preserve the very same ambiguity, because ambiguity is traceable to the *prejacent* and not to the meaning of the word 'ought' (Schroeder 2011, p. 9–10).

However, this quite natural explanation of the ambiguity found in 'ought' sentences of the considered sort has been forcefully put into doubt by some philosophers, recently most rigorously by Mark Schroeder (2011).[4] Schroeder admits that there is good reason to admire the *agency-in-the-prejacent theory* (henceforth *APT* for short), as he calls it, for its simple and seemingly convincing explanation of the ambiguity observed in the respective 'ought' sentences. Nevertheless, he provides examples that are supposed to undermine the initial explanatory attractiveness of *APT*.

As I understand Schroeder's criticism, the main problem he sees with *APT* is that it is inadequate as a *theory*. We expect of an explanation that enjoys theoretic significance (i) that it should properly *generalize*, which is another way of saying that it works well for all sorts of cases it should account for, and, on the other hand, (ii) that it should not *undergenerate*, meaning that it does not leave unexplained any case that it should explain. According to Schroeder, *APT* fails on both counts. On the one hand, *APT* is guilty of overgenerating, since it ascribes the ambiguity observed in one type of 'ought'

[4] Schroeder's paper enjoys the reputation of being an influential recap of the key problems with the orthodox explanation of ambiguity observable in some agential 'ought' sentences. The paper was originally published in *The Philosophical Review* in 2011 and next nominated and selected for inclusion in the 2011 *Philosopher's Annual:* a journal that publishes the ten best articles in philosophy of each year. Interestingly, some participants in the debate (e.g. Wedgwood, Broome) in the propositionalist camp found the criticism well-taken, and admitted that Schroder accurately pinpointed the central weakness of the approach, on which 'ought' is a raising verb expressing a propositional operator. The weakness in question is that if 'ought' relates agents to propositions, then why not think that it relates agents to *arbitrary* propositions? Schroeder dubbed the problem the Basic Problem. I critically evaluate it in (MSb).

sentences – agential 'ought' sentences – to some other type of 'ought' sentences, i.e. 'ought' sentences expressed in non-agential grammar. If the sentence 'Kate ought to get a sun hat' is ambiguous between an agentive and non-agentive reading, then the sentence 'It ought to be (the case) that Kate will get a sun hat' should also be ambiguous in the very same way, since these two sentences have the same prejacent *that Kate will get a sun hat*. However, the sentence 'It ought to be (the case) that Kate will get a sun hat' does not give rise to an agentive reading, at least not in normal circumstances. Moreover and more importantly, as Schroeder's objection goes, the stipulated extension of ambiguity, of the sort discussed above, generates unreliable predictions about the meaning of *some* agential 'ought' sentences, like 'Larry ought to win the lottery'. According to *APT*, the sentence in question admits both agentive and non-agentive readings, whereas it looks like the agentive interpretation of the sentence in question is blocked, since it is not in Larry's control to bring about that he wins the lottery. Therefore, *APT* does not deliver on its theoretical promise: it accounts for ambiguity observable in some agential 'ought' sentences, but not in all. And if *APT* is meant to explain ambiguity observable in *some* agential 'ought' sentences and not *all*, then we should be told what makes some agential 'ought' sentences such that *APT* applies to them and not to others. Otherwise, the explanatory value of *APT* is illusionary. We would like i to work for an arbitrary agential 'ought' sentence, or at least to know a principled way of figuring out the cases to which *APT* applies and those to which it does not. On the other hand, *APT* badly undergenerates since it 'inscribes' ambiguity into prejacents that obviously lack any. Consider the sentence 'Bill ought to kiss Lucy'. Now, if the sentence in question is ambiguous between agentive and circumstantial interpretations, as *APT* predicts, it has to be explained in terms of an ambiguity to be found in *Bill kisses Lucy*. But is it not too far-fetched to assign to the prejacent 'Bill kisses Lucy' the reading on which Bill is simply a patient in the relation between him and the kissing of Lucy?

Besides the two central problems that, on Schroder's diagnosis, *APT* shows vulnerability to, it falls short of satisfying a natural expectation coming from its promise to accommodate an agentive sense of 'ought'. If *APT* is to make room for an agentive reading of sentences like 'Kate ought to write the report', then it should be able to capture

the deliberative sense of 'ought', especially as agentive 'ought' is often about what the subject of the 'ought' *ought to do*. We may call the corollary problem of *APT*, when interpreted as an account of the agentive sense of some agential 'ought' sentences, the *deliberative ought exposure problem* (or *DOEP* for short). *DOEP* arises because if the central hallmark of the deliberative 'ought' is settling the agent's practical issue about what *to* do, *APT* lacks resources to bring out that sense straightforwardly. The source of the problem is taken to reside in the limits of the propositional interpretation of 'ought' which poorly, if at all, captures the relation between agents and *actions*. Action interpreted as a property of an agent makes it explicit who is to be the doer of the required action and thence nicely represents the sort of relation that the deliberative ought is supposed to represent, whereas action couched in terms of a proposition does not. Consider again the sentence 'Kate ought to write the report'. If the sentence in question is to be read as relating Kate to the action-proposition that she will write the report, how is this interpretation to do justice to the possible interpretation of the 'ought' as being primary about what *Kate* ought *to do*? Crudely, in what way does stipulating that the meaning of the sentence is *it ought to be the case that Kate will bring about that she will write the report* capture the alleged sense of the sentence as settling *Kate's* deliberative problem regarding what to do? These three alleged defects lead Schroeder to conclude that *APT* is unrescuable. Consequently, he concludes that we should reject it and replace it with one that does better – viz., the naïve view that Schroeder himself advocates. The naïve view is simply that some 'ought' sentences indeed are ambiguous because the *word* 'ought' is ambiguous.

The naïve view has an obvious advantage over *APT* in that rejecting sentential ambiguity implies dismissing the trouble-making claim that there is *one* unifying syntactic pattern that serves well in expressing the content of *any* ought-sentence.[5] However, I believe that Schroder's

[5] Schroder's view stands in opposition to the unifying account, which is still the dominant position. According to the unifying account of the meaning of 'ought', 'ought' always expresses a propositional operator *O*(p). Among adherents of the unifying account are Finlay and Snedegar 2014, Chrisman 2012a, 2012b, Finlay 2014, Cariani 2013 as well as Broome 2013 and Wedgwood 2006 (however these two philosophers propose certain improvements to the paradigmatic unifying account).

rebuttal of *APT* as offering no good explanation of ambiguity is too hasty. At the end of the day, it is *APT* that wins, although seeing this requires entertaining a *novel* interpretation of sentential ambiguity, different from the one considered in the literature. Or so I will argue.

2. THE MAP OF THE TERRAIN

This paper has two main objectives. First, to show that, Schroeder's bold declarations notwithstanding, the two objections he raises – that *APT* overgenerates and undergenerates – are in fact not damaging to the criticized view. In a nutshell, I will argue that the appeal to bad *over*generalization is unsuccessful for two reasons. First, because the examples analysed by Schroeder do not support his conviction that the *deliberative* 'ought' *cannot* be expressed in the syntax in which 'ought' relates agents to propositions and not agents to actions. Second, and relatedly, Schroder's unfortunate choice of the illustration of the apparently *bad* overgeneralization that *APT* gives rise to is a mark of a more systematic error he commits. The error in question, I will argue, lies in failing to recognize (or else misconstruing) the genuine character of the ambiguity detectable in the salient agential 'ought' sentences. The ambiguity in question does not concern availability of the circumstantial and the agentive readings of the relevant sentence, but rather the availability of a reading on which the agent of 'ought' is the sole initiator and executor of that 'ought' and a reading on which she is not. In other words, I think that Schroeder quite rightly observes that the relevant ambiguity has to do with the *deliberative* sense of 'ought';[6] however, in my view, he misinterprets what

[6] More precisely, with what philosophers typically tend to refer to in terms of the "deliberative" 'ought' but which is a misnomer for the sort of 'ought' that is what I like to call a 'full-blown' *first-personal 'ought'*, that is, 'ought' saying what the *deliberating* agent ought to do (herself). The label "deliberative ought" is often misleading since the notion "deliberative" is more capacious than the notion "first-personal ought". There can be deliberative 'oughts', i.e. oughts providing (good) answers to the question regarding what the deliberating agent ought to do, yet such that they are not delivered by the deliberating agent herself because, say, they are proposed by a trustworthy adviser. What is crucial in the case of such 'delivered' 'oughts if they are to fit well in the extension of the notion "deliberative ought" being at stake is that they have inscribed a requirement that the owner of the ought in question is, if possible, the sole producer of the required action.

this deliberative 'ought' is *really* about.[7] It is not so much about settling the question of what is advisable for the deliberator to do so that she does it, but rather about settling the question of what is advisable for the deliberator to do *herself* without delegating the essential part of the action to another party or parties.

As for the second challenge raised by Schroeder, the objection that *APT* undergenerates because the very same ambiguities that we observe in the prejacent of certain 'ought' sentences of the form 'S ought to *phi*' do not arise for some other 'ought' sentences of the very same form, I will dismiss it as irrelevant. Schroeder's case against *APT* is built on a somewhat sophisticated observation that in standard cases it is enough to passivize the original agential 'ought' sentence to see that the explanation of ambiguity offered by *APT* does not work. Two reasons explain why the test for passivization may *seem* a reliable test for evaluating the explanatory virtue of *APT*. One points to the phenomenon I am inclined to call the *fleeting character of ambiguity* that *APT* seems vulnerable to, and the other is inspired by the lesson derived from the passivization of 'ought' sentences that express the so-called deliberative 'ought', and which I call the *vanishing ambiguity problem*. Both problems concern the *dynamic* character of the ambiguity that adherents of *APT* are forced to predict about problematic 'ought' sentences. Both problems pose real challenges to *APT if* Schroeder's criticism is on target. I shall consider them in the next section in order to demonstrate how big a challenge Schroeder's two objections have pressed upon adherents of *APT*.

[7] Terminological confusion regarding the sense we are really after when we ask what the meaning of the deliberative 'ought' is has a long tradition. To my knowledge, Williams (Williams 1981) was the first author who used the term 'deliberative' as a cognate for my term 'truly agential' introduced in my book manuscript *Normativity that matters. On the meaning of practical 'ought' sentences* (MSa), and hence, because of the quite unfortunate choice of the label, he can be found guilty of misdirecting our attention to the properties that do not reflect the *agentive* essence of 'ought', i.e. the very property that makes it clear that the agent herself is required to *phi*. In my terminology this feature of 'ought' is referred to as 'authored'. More on this in section 5.

3. STRENGTHENING SCHROEDER'S CHALLENGE: AMBIGUITY AND DYNAMISM

The crux of the dynamic ambiguity problem, as I think of it, resides in its *dynamic,* and hence unstable character, which is very undesirable, if one's theory has an ambition to offer reliable predictions as to the character of the expected ambiguity. A good theoretical explanation of semantic ambiguity observed in sentences of the *same* type would be one that predicts *the same* ambiguity in 'Larry ought to win the lottery' as in 'Bill ought to kiss Lucy', or in 'Peter ought to feel the smell of sweet perfume in the room'. However, as Schroeder sensibly observes, whatever ambiguity we find in 'Larry ought to win the lottery' (if any at all),[8] it cannot be found in 'Bill ought to kiss Lucy'. It strikes one as far-fetched, to say the least, to propose that 'Bill kisses Lucy', apart from the obvious agentive reading, admits the circumstantial reading, on which it is said that kissing Lucy simply happened to Bill as winning the lottery happened to Larry. Winning the lottery is a sheer fluke, something you cannot *plan* to achieve, but kissing someone is not the sort of thing over which you have no control (at least not in ordinary cases).

The dynamic ambiguity problem in fact amounts to the *fleeting* and *unprincipled* character of the posited ambiguity, since it turns out that *what* ambiguity a particular agential 'ought' sentence exhibits cannot be easily explained by sentential ambiguity alone.[9] We can grant that the *sentences* 'Larry wins the lottery' and 'Bill kisses Lucy' are ambiguous, but this alone is not informative as to *the sort* of ambiguity that these two sentences exhibit. However, had *APT* provided a good explanation of

[8] Many philosophers, including Schroeder, tend to think that the agential 'ought' sentence 'Larry ought to win the lottery' is not ambiguous at all because the prejacent 'Larry wins the lottery' is not ambiguous: it is only sensible to think of Larry as of the patient of the relation between Larry and the winning since it is not in Larry's capacity to make his lottery ticket win. If the central argument of this paper is not flawed, then even seemingly not agentive 'ought' sentences like the one about Larry suffer from some sort of ambiguity of normative significance regarding the proper bearer of responsibility for the ought in question.

[9] In section 5 I will show that this objection is misguided, since ambiguity between the circumstantial and the agential interpretations is *not* the only ambiguity that *is* to be observed in a sentence. See also my "'Ought', ownership and agentive ought. Remarks on the semantic meaning of 'indexed ought'", forthcoming.

ambiguity, our knowing that ambiguity is *in* the prejacent would at the same time be knowing the *character* of the respective *ambiguity,* namely that it admits the agential reading and the circumstantial reading. But this is not so, since whatever ambiguity we may detect in the sentence 'Larry ought to win the lottery' is not observable in the sentence 'Bill ought to kiss Lucy'. So whatever theoretical virtue we owe to *APT*, it is not the one we are really after: we expect of *APT* that it will *explain the* disturbing ambiguity in the interesting agential 'ought' sentences, and not that it will give rise to new ones! If the *dynamic ambiguity problem* is genuine, then *APT* offers no solution to it. Worse, it actually increases our epistemic frustration, since what we learn from it is what we know from the outset, namely that the ambiguity is *in* the prejacent. But much more important is that what we *cannot* learn from *APT*, as I showed, is what sort of ambiguity is (or what sorts of ambiguity are) traceable to prejacents.

This objection is not a petty one: the problem that it brings out is not that *APT* does not indicate and enumerate the *potential* ambiguities that can be discovered in the sentences of the form '*S* ought to *phi*', but rather that it is hardly possible for *APT* to identify and enumerate *all* of the ambiguities that might be discovered in agential 'ought' sentences. If two NP-VP sentences randomly selected are ambiguous, and *APT* only manages to identify the ambiguity in one of these two sentences but not in the other, there is good reason to doubt that it offers a good explanation of the phenomenon in question.

In light of the *dynamic ambiguity problem*, Schroeder's objection is perfectly in order. Later on I will propose what I deem to be a promising rebuttal of this complaint, which shows that *APT* gives a *good* explanation of the observed ambiguity, but only once we revise our views on the character of the ambiguity that truly matters. If some NP-VP sentences are ambiguous in *more* than one way, as I think they are, and there is *a* way in which *any* of the considered 'ought' sentences are ambiguous, then Schroeder's objection loses much of its initial attractiveness.

Let us now consider the second worry that Schroeder's objections towards *APT*, were they successful, would give rise to. This is what I call *the vanishing ambiguity problem* (*VAP* for short). *VAP* is the reverse of the *dynamic ambiguity problem*, since if we assume that the relevant ambiguity is a dynamic phenomenon, then it might be

that an ambiguity originally detected in some sentence in a certain context will simply disappear into thin air in a different context. Now, it is necessary to emphasize that not *every* 'disappearance' of previously observed ambiguity is going to raise a worry for the friend of *APT*. Specifically, no harm to the explanatory value of *APT* is done if the ambiguity in question is removed by the context of interpretation. Let us grant that the prejacent 'Bill kisses Lucy' admits both the agential interpretation and the circumstantial interpretation (which we have assumed it *ordinarily* does not). Suppose next that the context dismisses the agentive interpretation as inapplicable to the case at hand. Imagine that you are a friend of Bill who knows of Bill that at the moment in which he kissed Lucy he was under the influence of drugs. There is no doubt that whatever one is doing when one is 'stoned', one is not exercising one's agency in the proper way. So it is not the *disambiguation* that *APT* cannot account for but the literal *disappearance* of the previously observed ambiguity. This serious defect is what the test from passivization purports to uncover. In the rest of the paper I shall be arguing that, despite appearances, the *dynamic ambiguity problem* is not a problem for a proponent of *APT*.

4. WHAT IS WRONG WITH THE OBJECTION FROM OVERGENERATION?

As I said, many participants in the debate over what explains the ambiguity of (some) sentences of the form '*S* ought to φ', where *S* is the subject and φ stands for action, subscribe to *APT*. Many, but not all. Schroeder remains unpersuaded by it. He has two general problems with *APT*. The first, and more minor, is that *APT* includes some, but not all, of the hallmarks of deliberative uses of agential 'ought' sentences. Second, and more major, is that among those hallmarks that are left out is the crucial one, precisely the one which stands behind the name of the considered use of agential 'ought' sentences, and which is directly important for advice in the sense of settling the deliberator's question of what to do. If 'ought' sentences that are claimed to express the deliberative 'ought' do not fulfil the promise, it is bad. So let us now see *how* badly *APT* fares when it comes to accommodating the deliberative sense of 'ought'.

Consider the following pair of 'ought' sentences (Schroeder 2011):

(A) 'Larry ought to win the lottery'
(B) 'It ought to be that Larry wins the lottery'.

The sentences presented in (A) and (B) have the same prejacent, which is 'Larry wins the lottery', however they seem to differ essentially in their meanings.[10] The linguistic explanation of our intuitions regarding the semantic difference between the two sentences is that, normally, the subject position is semantically significant. Oversimplifying, if a sentence has a subject argument place occupied, our first reaction is to interpret it as saying something about the agent. If the syntax is semantically illuminating, then consequently, whatever ambiguity, if any, is detectable in the sentence expressed in (A) should not be present in the sentence expressed in (B). But that conflicts with the prediction made by *APT* that the same prejacents ensure the same meanings. Given that the sentence in (A) has a clear agentive, or deliberative meaning (in Schroeder's terminology), (B) should also have it. However, in light of the constraints Schroeder places on the deliberative 'ought', the deliberative reading of the sentence presented in (B) is unavailable.

The argument that Schroeder supplies for his claim is simple: if the essence of the deliberative 'ought' resides in its direct relevance to the decision-making of the deliberator about *what to do*, then the agent's pondering over what *ought to be* is completely useless. Many states of affairs ought to be and ought not to be. The planet ought to be inhabited by happy animals, and it ought not to be destroyed by careless human global policy. It ought to be the case that Luckless Larry wins the lottery, and certainly it ought not to be that Larry's lucky lottery ticket disappears. But how can asking oneself the question about what ought to be the case bring the deliberator closer to settling what *she ought to do*? In Schroeder's view it cannot, so he concludes that 'ought' expressed in the sentence presented under (B) is not a deliberative sense of 'ought'.

[10] As it should be clear these days "meaning" is the philosophical term of art. It can mean various things to various philosophers in various contexts. Here I work with two assumptions that seem uncontroversial across the board, namely that sentential 'meaning' can be construed either in terms of the sentence's truth-conditions or as a proposition, or propositions expressed (if you are pluralist about the content as I am).

Schroeder's argument is simple, and appears to be intuitive; nevertheless, I am unconvinced. I think that in providing a negative answer to the question about whether the deliberative 'ought', as he characterizes it, can be expressed by a sentence evoking the raising syntax with the expletive 'it', he confuses two things that should not be conflated: one is formal and quite unimportant and the other is substantive and significant. It is also my view that once we identify the source of the confusion, we will see that Schroder's objection is toothless.

So let us separate the important from the unimportant. The *unimportant* issue concerns the grammatical form in which the deliberative question is typically couched, and the *important* issue concerns whether a particular syntax allows transmission of the intended *deliberative* content. Let me begin by explaining why the question regarding the grammatical translation of the deliberative 'ought', when properly construed, is irrelevant.

Consider the case of a twin brother of Luckless Larry,[11] Fearful Leo. Leo knows of himself that he is a lazy and fearful sort of person who does not like to exercise his capacity for agency. Being an agent implies bearing responsibility for the consequences of one's actions and this is exactly what Leo detests. Specifically, he hates making any decisions, no matter what particular area of human concerns the decision is supposed to have an impact on. Leo systematically avoids taking responsibility for his life as most people try to avoid snakes. He prefers to go with the flow of things: whatever will be, will be – that is his credo. However, there is one thing in his attitude that Leo is really tired of; it is his repugnance to decision-making *inasmuch* as it makes living with Leo very difficult for the ones he loves. For that reason, Leo finally decides to consult a specialist and go through psychotherapy. The therapist advises Leo, as a part of the therapy programme, that he should use a certain linguistic gimmick when he is about to ponder what to do. The recommended trick serves to ward off the demons of autonomy. Here it is: whenever Leo is about to deliberate over what he ought to do in order not to be overwhelmed by the frustrating feeling of responsibility for how his life will go, he should consider his potential courses of actions, not under the guise of performances done by him, but rather impersonally, under the guise of the states

[11] Luckless Larry is the protagonist of Schroeder's example targeted at demonstrating that there are uses of agential 'ought' sentences, such as 'Larry ought to win the lottery', whose surface logical form is deeply misleading: it suggests that the sentence is about *Larry* the agent who stands under the deliberative 'ought to do' relation with respect to winning, though such an interpretation is unavailable – it is not in Larry's capacity to bring it about that he wins. See Schroeder 2011, p. 8.

of affairs that should be effectuated. So, for instance, instead of asking himself whether he ought to exercise regularly to maintain a slim figure and good health, he is supposed to rephrase the question in more friendly (to his ears) terms by asking himself whether it ought to be that he exercises on a regular basis. This slight change in grammar would cause a great change in Leo's feelings, the psychotherapist told Leo, and after several attempts at deliberating in accordance with the recommended method, Leo now must admit that his doctor was absolutely right. He can come to decisions without having this unbearable feeling of having made one!

Does the fact that Leo hides the truly deliberative character of the question posed to himself, expressing it in non-agential grammar, make any difference when evaluating whether the 'ought' that Leo actually considers is directly relevant to Leo's settling the question about what *Leo* ought *to* do? Obviously, it does not. Whether a question does or does not have a particular character cannot be reliably read from the grammar alone. By itself, syntax lacks *intrinsic*, substantive meaning.

5. WHAT IS THE DELIBERATIVE 'OUGHT' *ALL* ABOUT?

If, as the example with Fearful Leo suggests, the deliberative 'ought' can be expressed in non-agential grammar, the thing that requires investigation is what makes a particular syntax well-suited to transmit the intended *deliberative* content. Now, to answer that question, we first need to establish what this *deliberative 'ought'* is. Once we have done this, we can move on to a more formal and sophisticated problem regarding the most appropriate logical form and grammatical interpretation of that 'ought'.

Schroeder enlists five hallmarks of the deliberative ought, of which three seem to be crucial, to my mind, when it comes to revealing the very nature of the *deliberative* 'ought' as he understands it. These are the following:

(i) direct relevance to advice, where what is meant by 'direct relevance' is that the 'ought' in question genuinely answers one's query regarding what to do;
(ii) close connection to the notion of accountability; and
(iii) tight connection to the notion of obligation.

Having in mind these three substantive features of the deliberative 'ought', we can appreciate the worry underlying Schroeder's

scepticism that non-agential grammar constructions are unable to convey the truly agential, or deliberative content.

If the evaluative 'ought' is used in contexts in which we are interested in what *state of affairs* it is desirable to obtain, it seems warranted to worry that a syntax that serves to express one kind of normative content may prove ineffective in expressing some other, especially if that other content is of a completely different normative character. The question I want to address now is whether the raising syntax typical of the evaluative and non-agential use of 'ought' is suitable as the bearer of the deliberative content in the above-presented sense. More precisely, I will explore whether the raising syntax allows us to exhibit the three listed *substantive* hallmarks of the deliberative 'ought'. Schroeder considers this task to be unfeasible. I disagree. I think that we can express what is called the deliberative 'ought' in the raising syntax once we properly interpret the sense of the term 'deliberative' that is *really* at stake, which in my view is to be achieved once we are *explicit* about what makes the deliberative 'ought' *different* from the evaluative 'ought', and at the same time closely related to the 'ought' of obligation. What makes the so-called deliberative 'ought' special does not have much to do with the fact that 'ought' is truly useful for advice, because it is not very clear what is actually meant by 'usefulness' or 'relevance to advice',[12] but rather it is tightly connected with the *first-person perspective*, and particularly with the requirement that the owner of ought (to use a nice phrase borrowed from Broome) be also the producer of the demanded action. Since misidentifying the very essence of the deliberative 'ought' is quite naturally to be followed by a *formal* misrepresentation of the true character of the 'ought' in question, Schroeder's mistaken idea as to what this deliberative 'ought' consists in quite naturally led him to raise mistaken objections about the formal *mis*representation of this sort of 'ought'.

In what follows I am going to argue that *APT* sustains Schroeder's criticism once we decipher the relevant meaning of the deliberative 'ought' which, surprisingly, does justice to Schroeder's own view

[12] In my book manuscript (MSa) I argued that the notion of advice is ambiguous with respect to its practical character. This view on the ambiguity of advisability places me in opposition to the dominant view, which is the unambiguous account of advisability put forward by Schroeder 2011, Chrisman 2012a; 2012b, and Finlay 2014.

on the nature of this sort of 'ought'. Moreover, if my explanation of the ambiguity proves defensible, then it nicely, yet, perhaps, surprisingly, generalizes to *any* ought-sentence, irrespective of whether it expresses normative, epistemic or any other content. Most importantly, however, it undermines Schroeder's challenge against *APT*. If I am correct, then it is not true that the ambiguity observed in agential 'ought' sentences cannot be satisfactorily explained in terms of sentential ambiguity.

5.1 THE AMBIGUITY THAT MATTERS

To give a preview of my idea, think of a paradigmatic non-agential, non-normative 'ought' sentence like 'The sun ought to rise early tomorrow'. Evidently, the 'ought' in the considered sentence has been used in the epistemic sense, meaning something to the effect that the state of affairs including the sun's early rising tomorrow is likely to be the case. The considered 'ought' claim evokes the raising syntax, which is the proper syntax to express the epistemic sense of 'ought'. Now, even the epistemic 'ought' claim warrants the stipulated authored interpretation that I take to be a mark of the deliberative use of 'ought' when this sort of 'ought' is properly construed. That is, the prejacent sentence 'sun rises' 'hides'[13] ambiguity with respect to whether the sun rises 'out of itself' (that is, in the *authored* fashion in my terminology), so to speak, or rather is *made* to rise by some external forces. I further explain the proposed ambiguity by alluding to the intrinsic feature of the *prejacent proposition*,[14] which is coarse-graininess that does not allow us to recognize immediately the *relevant* meaning of p:[15] whether p is

[13] I assume, perhaps controversially, that semantic ambiguity is not to be restricted only to the sort of ambiguity that is something that strikes us from the very first encounter with the sentence, since very few ambiguities are detectable in this way. *Seeing* and *not seeing* ambiguity is a matter of training in recovering the admittable (in the light of a linguistic theory and everyday pragmatics) senses of the sentence. If the idea of ambiguity (explicit ambiguity by default) makes sense, then its reverse must also make sense. *Implicit* ambiguity is the ambiguity we have got used to disregarding as bringing out the sense of the sentence that is completely irrelevant for our successfully grasping the relevant message conveyed by the sentence.

[14] 'Prejacent proposition' is the proposition denoted by the sentence.

[15] What I have in mind is best illustrated by an example. To avoid complexity by beginning with non-agential propositions, take a sentence expressed in

about the state of affairs that obtains, or about the state of affairs that obtains due to agential contribution. But if *p* truly means '*A* made *p* be the case', the next question that immediately arises is whether *p* truly means '*A* made *p* be the case *herself*' or *p* rather means '*A* made *p* be the case *due* to the contribution of *B*'. But if *p* can be read as meaning '*A* made *p* be the case *due* to the contribution of *B*', it remains further unclear whether things are such that the relevant meaning of *p* is '*A* made *p* be the case due to the contribution of *B* herself', or rather the relevant meaning is '*A* made *p* be the case due to *B*'s making it the case that *C* finally brought it about that *p* is the case'. If this view of mine is tenable, then it turns out that the grammatical translation of propositional 'ought' through an 'ought that' construction introduces endless sentential ambiguity regarding who the *truly* normatively relevant owner of 'ought' is – the person responsible for performing the 'ought to do' in question. I contend that *this* is a serious problem that any interpretation of the so-called 'deliberative' 'ought' in *terms of* propositional 'ought' faces; it seems unable to block the regress of

---

agential grammar such as 'Larry wins the lottery'. Now, my claim is that 'Larry wins the lottery' is essentially ambiguous in a twofold way: (i) it is unclear in the way *APT* predicts: we do not know whether what is said is that *Larry* is the agent of winning, or rather what is said is that winning falls upon Larry, as something that happens to him; and (2) it is unclear whether what the sentence says is that it is Larry *himself* that makes it the case that he wins, or rather what the sentence says is that Larry *wins*, which in turn admits the interpretation that Larry wins due to the contribution of Betty, or some other person the sentence is silent about. Note that the sort of ambiguity that I ascribe to the prejacent sentence is much more troubling than the sort of ambiguity that *APT* posits, because the ambiguity between the agential and the circumstantial interpretation itself is unambiguous, that is, it does not prompt further inquiry regarding the proper meaning of each of the available readings, whereas the sort of ambiguity that I associate with the grammatical interpretation of the proposition itself is seriously worrisome for two reasons. One is that the ambiguity in question generates endless ambiguity, and the second is that it has a metaphysical underpinning: it is in the nature of the grammatical translation of *any* proposition that we cannot tell from it whether the sentence is true because the subject of the verb *himself* made it true, or rather the sentence is true due to the contribution of other factors, agential as well as non-agential. I introduced the problem of the regress of ambiguity observable in the propositionalist interpretation of 'ought' in the context of a discussion over what is the very thing that makes interpretation of 'ought' in terms of propositional 'ought' a challenging enterprise. See my book manuscript (MSa) and Klimczyk 2018, forthcoming.

ambiguity regarding the proper owner of normative 'ought'. Offering a solution to that problem is, in my view, equivalent to a successful defence of the idea that the deliberative 'ought' can be expressed via the relation between an agent and a proposition.

But now, let me return to the suggested and controversial idea that even non-agential sentences like 'The sun ought to rise early tomorrow' or 'The ball ought to be in the pocket' are ambiguous in the same way in which 'Larry ought to win the lottery' or 'Peter ought to brush his teeth' are, though in ordinary circumstances we tend to ignore the sort of ambiguity in question. And we lose it out of sight for the obvious reason that what I dub the 'authored' aspect of 'ought' is *beside the point*. When I utter a prediction about the expected hour of tomorrow's sunrise, I am interested in the potential consequences of the early sunrise for my practical purposes: assume that I plan an all-day trip to the woods. Similarly, when I spell out the prediction that the ball ought to be in the pocket, I produce the judgment with a particular intention in mind, specifically with the intention of establishing the current state of the billiard game. If in the two sorts of circumstances that my examples illustrate, there is no place for the question of whether the sun is going to rise early tomorrow *out of itself*, and analogously for the question of whether the ball *itself* makes it the case that it has landed in the pocket; this finds explanation not in the unintelligibility of the questions posed but rather in their contextual inadmissibility. The fact that we do not normally notice the considered sort of ambiguity in non-agential sentences like the above-considered ones is not a mark that the proposal is mistaken, but rather that the context of interpretation renders the suggested readings unavailable or spurious.

If I am right and the question regarding the *authored* or *non-authored* character is in principle an intelligible question that applies to most uses of 'ought', this is very good news for a proponent of *APT* because it delivers the most desired evidence that *APT properly* generalizes. What I mean by stipulating that "*APT properly* generalizes" is that it provides a *global* explanation of the phenomenon of ambiguity observable in almost all, if not all, 'ought' sentences to the following effect: if an 'ought' sentence is ambiguous, it is ambiguous in the proposed sense. The "global" character of the advocated explanation is due to the intrinsic ambiguity of the grammatical interpretation of

the *proposition* 'that *S* φ-s'[16] from which *alone*, barred from the context in which it occurs, we cannot directly glean whether the *authored* or *non-authored* reading is the relevant one. I conclude with the bold conditional claim in favour of *APT*: *if* ambiguity possibly detectable in a prejacent sentence is reducible to ambiguity *in* the propositional interpretation of the prejacent sentence in question, in which case the ambiguity becomes in the important sense *fundamental,* then *APT* offers a truly powerful explanation of it. This is so because the controversial ambiguity turns out to be a more global and more primitive phenomenon than has been initially assumed. It is global because it is predictable of any arbitrary 'ought' sentence whatsoever, and it is primitive because it is essentially connected with how a proposition gets grammatically translated, which makes the ambiguity under consideration more a *syntactical* phenomenon than a semantic one. If I am right, then sentences as diverse as 'Larry ought to win the lottery', 'Bill ought to kiss Lucy', 'Ann ought to brush her teeth' are equally ambiguous because their prejacents *are* ambiguous with respect to *who* is the agent that bears responsibility for making these 'oughts' be the case. And prejacents of the considered sentences are ambiguous because it is in the nature of such propositions to be vague with respect to the authored character of the expressed action.

6. WHAT IS WRONG WITH THE UNDERGENERATION ARGUMENT?

Let me now examine the second challenge posed by Schroeder, which is that *APT* badly undergenerates. The undergeneration argument, as I call it for short, is based on the idea that the test for passivization reveals something important about what readings of a particular prejacent are available, and for that reason is conclusive with respect to whether the prejacent is or is not ambiguous.

I think that the argument that builds upon passivization, as it stands, is toothless. I consider it to be toothless because it is irrelevant, and it is irrelevant because it is difficult to imagine what important information *about* semantic ambiguity we can infer from the trivial

[16] *S* stands for subject in general: agential (me, your boss, teacher) and non-agential (tree, ball, sun). *S* is the default subject of the verb, which is grammatically absent.

information that some 'ought' sentences are subject to passivization and others are not.

Recall what *the* allegedly bad thing that, according to Schroeder, the passivization test reveals is: the 'ought' sentence that originally had *explicitly* agential *meaning* like 'Bill ought to kiss Lucy' in consequence of a petty transformation in grammar to the effect of 'Lucy ought to be kissed by Bill' acquires a vivid *non-agential* meaning. The original sentence was saying something about what *Bill* AGENT ought to do, but the sentence that went through passivization no longer says what Bill AGENT ought to do. This objection seems to be decisive to many as it is based on an unquestionable platitude in linguistics that active and passive constructions do not differ as to the truth conditions of the propositions expressed. On a natural reading, as Schroeder stresses, the proposition *that Bill kisses Lucy* is true if and only if the proposition *that Lucy is kissed by Bill* is true too.

I have no quarrel with such a description of the case. What I disagree with, however, are the consequences derived from this obvious observation, that propositions expressed in active and passive form have the same truth conditions; namely I oppose the idea that active and passive constructions express the *very same* thought. According to my view, the thought is not so much the *uttered sentence* but the *uttered sentence properly logically interpreted*[17]. Having said that, this implies that, on my interpretation, whether sentences like 'Bill kisses Lucy' and 'Lucy is kissed by Bill' express the same thought or not depends on the accepted logical interpretation of the prejacent proposition. Differently put, on the stipulated view it is not the prejacent proposition denoted by the sentence that is the bearer of truth conditions *but* the relevant logical interpretation of it. And if, in uncovering the thought expressed by a sentence, the focus is put on the most discourse-apt logical[18] interpretation of the prejacent proposition, because on my view the discourse-apt logical form of the sentence is the bearer of the sentence's meaning, then it seems obvious that the proposition *that Billy kisses Lucy* and the proposition *that Lucy is kissed by Bill* have different truth conditions. On the most natural reading, truth

[17] Thanks to the referee for pressing me to clarify my view.

[18] In *Normativity that matters* I explain the meaning of normative 'ought' sentences in terms of their discourse-apt logical form.

conditions for the former are given by the state of affairs including Bill's sole involvement in the kissing, whereas truth conditions for the latter are given by a general state of affairs in which Lucy is the patient of Bill's action. Since in the passivized sentence the emphasis is put on the obtaining of the relevant states of affairs, the scope of Bill's agency involved in making that state of affairs obtain is not clear. If we tend to interpret the sentence 'Lucy is kissed by Bill' as having a default meaning equivalent to the meaning given by the sentence 'Bill kisses Lucy', this seems to be so because we assume that the most natural context of utterance of the sentence 'Bill kisses Lucy' is one in which in kissing her, Bill exercises his agency *to the full*. Be that as it may, it is worth bearing in mind that the putative equivalence of propositions is not what comes out directly from the respective sentences, but rather is something that follows from a theory of meaning that one subscribes to. If you believe that sentences express some minimal proposition that remains unaffected by the context in which the sentence is uttered, then nothing in the proposition *that Lucy is kissed by Bill* itself gives rise to the stipulation that it necessarily has the same truth conditions as the proposition *that Bill kisses Lucy*. The uttered sentence 'Bill kisses Lucy' *may or may not* express the same proposition as 'Lucy is kissed by Bill'. The point I am making here is that the proposition *that Bill kisses Lucy* is true when the proposition *that Lucy is kissed by Bill* is true, but this is so *not* because these two propositions have the same truth conditions, but because their truth conditions happily overlap when the truth conditions of the former are given by such distinct propositions as *that Lucy is kissed by Bill AGENT* and by *that Lucy is kissed by Bill as a result of Bill's being manipulated to do the kissing without knowing what he is doing*. To sum up, my suggestion is that in order to establish whether sentences using active and passive constructions express or do not express the same thought, we need to get to the *proper* logical interpretation of the proposition expressed. The significance of my doubts should not be underestimated, since if 'Bill kisses Lucy' and 'Lucy is kissed by Bill' do not express the same thought, the key motivation behind helping oneself to the passivization test in the role of a powerful argument for the unaccommodated intuition that there is no *real* difference in meaning between the passivized sentence and its non-passivized original is undermined. Moreover, and in a sense more importantly – at least more importantly from the point of view of the

dialectics of Schroeder's criticism[19] – if he errs in his claim that 'Bill ought to kiss Lucy' and 'Lucy ought to be kissed by Bill' express the same content, passivization is completely ineffective in further undermining *APT*'s capacity to capture the interesting sense of ambiguity. It would make sense to resort to passivization if it was crystal clear in the first place that the prejacents 'Bill kisses Lucy' and 'Lucy is kissed by Bill' do have the same *content*, which is something I doubt.

Now, in order to establish whether a passivized sentence and its non-passivized counterpart express or do not express the same thought, we must settle what is understood by 'thought'. I assume that 'thought' in general is a term interchangeable with the term 'content', and 'content' is typically denoted by a multitude of propositions from which the particular proposition that is the most accurate expression of a thought hangs on the details of the context in which the thought occurs.[20] Differently put, the thought is the proposition expressed by a sentence *and given* by the logical interpretation that best matches the speaker's understanding of the situation that gave rise to the thought in question. This means that if we consider a sentence like 'Bill kisses Lucy', then according to the proposed approach, the thought expressed is given by the logical interpretation of the proposition 'that Bill kisses Lucy' that best corresponds to the speaker's grasp of the situation. Crudely, if the logical interpretation that best expresses what the speaker thinks is the proposition *that Bill is such that Bill kisses Lucy*, then the thought expressed by the sentence 'Bill kisses Lucy' is that BILL[21] kisses Lucy. I also take it that the proper logical interpretation of the uttered sentence is *what is said*

[19] Schroeder's charge is tricky because if the passivization test is apt, then it might be that *APT* fails *not* necessarily because the predicted ambiguity in 'Bill ought to kiss Lucy' is not to be found in 'Lucy ought to be kissed by Bill' but perhaps because 'Lucy ought to be kissed by Bill' *remains* ambiguous in the way 'Bill ought to kiss Lucy' is but *APT* misidentifies its nature in the first place.

[20] I assume that in general each possible propositional representation of a particular thought is complete, that is, is truth-evaluable, yet neither of the individual propositions on their own that expresses the thought, expresses it completely. In other words, I am of the opinion that only a conjunction of propositions making up the contextually salient content of thought is almost successful in representing *the* content of one thought.

[21] Here I use capital letters to indicate a peculiar form of agency as when the agent of some action is to be the sole producer of it.

by the token of an utterance under consideration. Now, it strikes me as not question-begging that the thought expressed by the sentence 'Bill ought to kiss Lucy', irrespective of the *precise* character of what it is *all* about, is primarily about *Bill*. Similarly, it is my strong conviction that the thought expressed by the sentence 'Lucy ought to be kissed by Bill', again quite irrespective of the *precise* character of what it is *all* about, is primarily about *Lucy*. The non-passivized sentence says something about what Bill ought to *do* and its passivized counterpart says something about *to whom* Bill ought to do the thing in question. If the content of these two sentences differs with respect to the general subject matter, how can it be that they, in principle, express the same thought? And if they do not express the same thought, because at some general level they are concerned with different objects, how can it be that the claim about the ambiguity regarding their senses arises in the first place as *a* relevant issue?

We do not tend to suspect semantic ambiguity whenever we encounter a pair of sentences expressing different propositions with the same truth conditions. The sentence 'Hesperus is the personificated name for the planet Venus' and the sentence 'Phosphorus is the personificated name for the planet Venus' have the same truth conditions, but we do not predict any sentential ambiguity about them. A similar observation strikes me as true for an arbitrary sentence in English. Take the sentence 'Susan is drinking white liquid from the glass, the liquid which is one of the main ingredients for baking chocolate cakes' which is obviously true if and only if the sentence 'Susan is drinking milk from the glass, the liquid which is one of the main ingredients for baking chocolate cakes' is true. Does the fact that these two sentences *say* different things while referring to the same state of affairs warrant the worry about the ambiguity that is to be discovered in the proposition presented under one description but which is missing when we propose an alternative description of the same state of affairs? I think not.

However, I expect that to this suggestion of mine Schroeder would reply by saying that the point of his critical note is not so much that the two 'ought' sentences telling us about an event from the lives of some film protagonists Bill and Lucy *strictly speaking* express the same proposition because they actually have *different* prejacents, but rather the idea is that we, *the interpreters*, "naturally" *understand* these two sentences as expressing the same thought, because we have in

mind the *relevant* scenario, namely, in the example at hand, the idealistic scenario in which "two lovable but romantically ill-fated characters ought to get together" (Schroeder 2011, p. 12). If that were the line of his reply, then I suppose that the general and quite correct lesson he wants to teach us is as follows: when considered against the background of a specific narrative, or in the light of a determined context of interpretation, the passivized sentence and the non-passivized original *tend* to *be given the same* meaning, where what is meant by "meaning" is what is said, or what is asserted by the respective sentences, and not their truth conditions.

Again, that strikes me as perfectly right as a thesis about how we come to understand the meaning of a sentence *from within* a story. When we are gripped by the narrative, we follow the plot with the general intention of getting the depicted *facts* right, which typically implies no eye on the interpretational subtleties. This means that although from the point of view of the communicative intention of the producer of the statement 'Bill ought to kiss Lucy' the grammatical construction used to build the sentence he utters may be important, as it serves to put emphasis on that it is *Bill* and not, say, Ricky or Mark who ought to kiss Lucy, it is not so important from the perspective of the interpreter who wants to get an overview of what is going on. I assume that getting the general idea of what is going on is the basic intention governing one's watching a film.

However, and here is my objection to Schroeder's argument, the pragmatic reasons that speak for attributing the same content to sentences expressing literally *distinct* propositions in the context of *learning* a story (where learning a story consists in arranging all the available and relevant information in order that it provides us with an *intelligible* outlook on what this all is about) do not belong to *the sort* of reasons speaking for a general thesis in philosophical semantics saying that prejacents that differ with respect to their subject-matter have the same content.

To conclude: from the fact that a pair of sentences that have the same truth conditions *say* different things, nothing follows about these sentences being ambiguous in a *different* way, which is what Schroeder's argument from undergeneration assumes to be the case. In fact, nothing directly follows about the potential ambiguity in what is said. Even more, there is nothing in what the test for passivization

reveals that would justify our interest in the unambiguous character of these sentences in the first place. Schroeder's argument is ill-advised because he takes the test for passivization to settle what the sentence *means*, whereas in fact it can only be helpful in uncovering the plural content of what *is said*. For that reason I find it difficult to see in what way the passivization could be instructive in deciding whether the *explanation* of ambiguity given by *APT* is any good.[22]

7. CONCLUSION

In this paper I set myself two kinds of objectives: critical and constructive. In the critical part these objectives were three in particular: (1) to show that Schroeder's criticism targeted at the theory, according to which ambiguity observable in some 'ought' sentences resides in the sentence, and not in the meaning of the word 'ought', is not as powerful as it is thought to be; (2) to justify the claim that Schroeder's arguments against treating the deliberative 'ought' in terms of the propositional ought is, overall, unsuccessful, and finally (3) to show that the reason for which Schroeder's challenge towards *APT* is mistargeted is that he underestimates the possibility that there might be more than one ambiguity observable in the prejacent sentence. On the constructive side, I wanted to render attractive the two-tier conditional claim, saying the following: *if* there is an interesting ambiguity to be observed in agential 'ought' sentences, it concerns the issue of *who* is to be the essential executor of it, and thence the bearer of responsibility for making the relevant 'ought' proposition true, and *if* the *right* kind of ambiguity concerns authorship of the required action, then that very ambiguity should be recoverable from *any* 'ought' sentence whatsoever. Even if you, the reader, think that these tasks have not been successfully completed, I hope that you will grant as much that a new and interesting perspective in the philosophical study on the meaning of 'ought' has been brought to further examination.

[22] The confusion arises because the phrase 'sentential meaning' can mean different things: for some and in some contexts it is synonymous with 'truth conditions of the sentence', and for others and in some contexts it can mean 'content', or 'the proposition expressed'. Now, what passivization shows is that we should not identify the truth conditions of the propositions expressed by a passivized sentence and its non-passivized originals with the content expressed by these sentences.

REFERENCES

Belnap N., Horty J. F. (1995). The Deliberative Stit: A Study of Action, Omission, Ability and Obligation. *Journal of Philosophical Logic*, *24*, 583–644.

Broome, J. (2013). *Rationality Through Reasoning*. Chichester, UK: Wiley-Blackwell.

Brandom, R. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.

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

Cariani, F. (2013). 'Ought' and Resolution Semantics. *Noûs*, *47*(3), 534–558.

Chrisman, M. (2012a). 'Ought' and Control. *Australasian Journal of Philosophy*, *90*(3), 433–451.

Chrisman, M. (2012b). On the Meaning of 'Ought'. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (vol. 7, pp. 304–332). Oxford: Oxford University Press.

Finlay, S., Snedegar J. (2014). One Ought Too Many. *Philosophy and Phenomenological Research*, *89*(1), 102–124.

Finlay, S. (2014). *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press.

Fintel, K. von (2006). Modality and Language. In D. M. Borchert (ed.). *Encyclopaedia of Philosophy*. Second Edition. Detroit: MacMillan Reference USA. http://web.mit.edu/fintel/modality.pdf

Glüer, K., Wikforss A. (2009). The Normativity of Meaning and Content. In Edward N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)*. https://plato.stanford.edu/archives/spr2016/entries/meaning-normativity

Glüer, K., Wikforss A. (2010a). *Es brauch die Regel nicht. Wittgenstein on Rules and Meaning*. In D. Whiting (Ed.). The Later Wittgenstein on Meaning (pp. 148–166). Basingstoke: Palgrave Macmillan.

Glüer, K., Wikforss A. (2010b). The Truth Norm and Guidance: A Reply to Stegliech-Petersen. *Mind*, *119*, 757–761.

Hattiangadi, A. (2007). *Oughts and Thoughts. Rule-Following and the Normativity of Content*. Oxford: Oxford University Press.

Higginbothan, J. (2006). Truth and Reference as the Basis of Meaning. In M. Devitt, R. Hanley (Eds.). *The Blackwell Guide to the Philosophy of Language* (pp. 58–76). Oxford: Blackwell.

Horty, J. F. (2001). *Agency and Deontic Logic*. Oxford: Oxford University Press.

Horty, J. F. (2012). *Reasons as Defaults*. Oxford: Oxford University Press.

Klimczyk, J. (2018) 'Ought', Ownership and Agentive Ought. Remarks on the Semantic Meaning of 'Indexed Ought'. *Studia Philosophiae Christianae*, forthcoming.

Klimczyk, J. (MSa) *Normativity That Matters. On the Meaning of Practical 'Ought' Sentences*.

Klimczyk, J. (MSb). *Why the Basic Problem is not a Problem?*

Miller, A. (2010). The Argument from Queerness and the Normativity of Meaning. In M. Grajner (Ed.). *Truth, Existence and Realism*. Frankfurt, Heusenstamm, Paris, Lancaster, New Brunswick, NJ: Ontos Verlag.

Ridge, M. (2014). *Impassioned Belief*. Oxford: Oxford University Press.
Shackel, N. (2014). A Problem for the Unity of Normativity. *Analysis*, *74*(3), 404–411.
Schroeder, M. (2011). *Ought*, Agents and Actions. *Philosophical Review*, *120*(1), 1–41.
Wedgwood, R. (2006). The Meaning of 'Ought'. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (vol. 1, pp. 127–160). Oxford: Oxford University Press.
Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.
Williams, B. (1981). 'Ought' and Moral Obligation. In B. Williams, *Moral Luck* (pp. 114–123). New York: Cambridge University Press.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 139–159
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.07
Rozprawa

Erich Rast*

METALINGUISTIC VALUE DISAGREEMENT

Summary: In a series of publications Burgess, Plunkett and Sundell have developed a metalinguistic negotiation view that they call 'Conceptual Ethics.' I argue that their position adequately captures our intuition that some cases of value disputes are metalinguistic, but that they reverse the direction of justification when they state that speakers 'negotiate' the best use of a term or concept on the basis of its prior social role. Borrowing ideas from Putnam (1975b), I instead suggest distinguishing two meanings of general terms and value predicates. Core meaning represents the lowest common denominator between speakers and is primarily based on our needs to coordinate behavior. In contrast to this, the noumenal meaning of a general term or value predicate is intended to capture an aspect of reality and represents what a term *really* means. Like many other disputes about theoretical terms, terms for abstract objects, and predicates, metalinguistic value disputes are about noumenal meaning on the basis of a shared core meaning. This direction towards reality is what sets the account apart from mere metalinguistic negotiation.

Keywords: metalinguistic negotiation, value disagreement, relativism, meaning theory, externalism

* New University of Lisbon. E-mail: erich@snafu.de

Work on this article was conducted with funding under grants SRFH/BPD/84612/2012 and PTDC/MHC-FIL/0521/2014 of the Portuguese Foundation for Science and Technology. Many thanks to Marcin Lewinski and all other members of our project team as well as the participants of the PHILANG 2017 Conference for useful suggestions and commentaries.

1 INTRODUCTION

Consider the following conversation:

(1) **a.** Alice: Capitalism is good.
b. Bob: No, it isn't.

The dialogue illustrates an old conundrum of lexical semantics that has become important for a recent debate on relativism and contextualism in the philosophy of language. Suppose Alice has a certain logical combination of criteria $A_1, \ldots, A_n$ in mind that account for the truth-conditions of her use of 'good' in that particular case, but that Bob has a different logical combination of criteria $B_1, \ldots, B_n$ in mind. They implicitly disagree about the meaning of that particular use of 'good' in the given context. Then it seems that they are talking past each other, because Bob's reply does not contradict the content of Alice's assertion. He may agree with her about the question of whether capitalism satisfies criteria $A_1, \ldots, A_n$, but implicitly disagrees with her implicit assumption that these criteria provide an adequate lexical decomposition of her particular use of 'good' in the given situation. The problem is to explain how such disagreements are possible, in which sense they are metalinguistic and in which sense they are substantial. Plunkett and Sundell (2013) argue that speakers can have *substantial* verbal disputes and discuss the following similar examples:[1]

(2) **a.** That chilly is spicy!
b. No, it's not spicy at all!

(3) **a.** Secretariat is an athlete.
b. Secretariat is not an athlete.[2]

(4) **a.** Waterboarding is torture.
b. Waterboarding is not torture.

(5) **a.** Lying with the aim of promoting human happiness is sometimes morally right. In fact it often is!
b. No, you are wrong. It is never morally right to lie in order to promote human happiness.

(6) **a.** **A** tomato is a fruit.
b. No, a tomato is not a fruit.

---

[1] See Plunkett and Sundell 2013, p. 15, 16, 19, 20, 22.

[2] Two sports reporters are discussing a horse in a race. One is calling it an athlete, whereas the others point out that only humans can be athletes. This example is originally from Ludlow (2008). See Plunkett 2015, p. 840–845.

Before going on, it is worth pointing out that there is an innocuous variant of these examples. Consider (1) again. Bob might associate the same criteria A_1, …, A_n with Alice's use of 'good' but disagree about the claim that capitalism satisfies them. I call this case the *direct content-based disagreement*.[3] It is not problematic and will not be discussed further. Instead, I will discuss readings of the examples according to which Alice and Bob implicitly disagree about the right meaning or interpretation of the predicative complex in the given context, and maybe also disagree about the meaning of the logical subject of predication like 'capitalism' in (1).[4] These readings give rise to a metalinguistic analysis like it has been suggested by Sundell (2011), Burgess (2013), Burgess and Plunkett (2013), Plunkett and Sundell (2013, 2014), and Plunkett (2015). My central thesis is that although there are such metalinguistic disputes, Burgess, Plunkett and Sundell's (in short: BPS) metalinguistic negotiation view (aka 'Conceptual Ethics') paints a skewed, or at least incomplete, picture of them. According to BPS, it is often not the existing social role or function of the expressions under consideration that motivates the dispute, but the speakers' concern with a shared reality. In that respect, I shall argue, metalinguistic value disputes are not substantially different from many other disputes about general terms, singular terms for abstract objects and their corresponding predicate expressions.

2 APPROACHES TO VALUE DISAGREEMENT

There are some semantic positions in the philosophy of language that need to be mentioned to put the current debate into the proper historical perspective, even though I agree with BPS they ultimately

[3] By 'content' I mean content in the sense of Kaplan's character/content distinction, that is, the notion of semantic content that is the result of saturating indexicals and evaluated with respect to circumstances of evaluation. The above characterization does not necessarily apply to broader notions of semantic content which may have their place in other approaches. What is important for the current purpose is that Alice and Bob disagree on the basis of the same meaning and the same contextual resolution of indexical expressions.

[4] To keep things simple, I will assume in what follows that the meaning of either the logical subject expression or the predicative complex is fixed. In principle, however, a metalinguistic dispute can be about both expressions at the same time.

fail as an adequate explanation of the types of disputes in examples like the ones above.

2.1 CONTEXTUALISM

First, what Alice and Bob have 'in mind' in example (1) could be spelled out in contextualist terms in a framework based on Kaplan (1989). In a context with Alice as evaluator, the semantic content of (1–a) would yield a semantic content with her criteria, whereas in a context with Bob as evaluator the semantic content would contain his criteria.

However one spells out the details of this approach, it does not seem to be adequate in general, though, since it merely restates the problem in a particular contextualist framework: According to this semantics Alice and Bob associate different semantic contents with Alice's original utterance, and so they do not really disagree and are talking past each other. Whatever attitude Bob has about *his* content could be compatible with any attitude about *her* content, and vice versa.

2.2 RELATIVISM

Second, as a solution to this problem a relativist semantics could be given, as it has been defended for predicates of personal taste and even evaluative language in general by various authors such as Kölbel (2002), Lasersohn (2005), MacFarlane (2005, 2014), and Egan (2014). There are many different brands of relativism and trying to characterize all of them would go beyond the scope of this article. Generally speaking, a typical assessment-relativist approach to value disagreement would stipulate that the same semantic content of (1) may be true with respect to Alice as a judge and false with respect to Bob as a judge.

Could this type of relativism be a solution to the problem of how to explain value disputes like (1)? As far as I know, nobody has ever seriously considered such semantics for disputes that are clearly value-based such as (1), (4), and (5), and with good reason. According to assessment-relativist semantics, the use of 'good' would be interpreted as 'good relative to a judge' – though not as part of the semantic content, but as part of the semantic evaluation mechanism. There are many metaethical theories with which this view is compatible, for example some forms of (ideal) appraiser subjectivism, but there are

also many moral theories with which it is plain incompatible such as value-absolutist readings of nonrelational good. Assessment-relativism of 'good' excludes those metaethical stances.

By the same token, a contextualist semantics for 'good' that always stipulates a hidden argument place for a benefactor would exclude any metaethical stance that asserts that certain uses of 'good' can be nonrelational, stating that '*x* is good' is true if and only if *x* is good at the time of utterance, false otherwise.

This raises a question: Can the semanticist tell moral philosophers that they are wrong? There can certainly be scenarios in which a semanticist may inform a moral philosopher that there is a logical problem with a certain moral conception or that it does not conform with what speakers ordinarily associate with the meaning of given value terms, and there are metaethical stances like quasi-realism whose express purpose is to make prima facie implausible metaethical theories compatible with semantics and our common-sense intuitions. In general, however, the answer must be No. Semantics can put weight and pressure on certain philosophical constructions but cannot decide them. If semantics provided a knockdown argument to a certain moral stance, for instance, then it would be moral philosophy in disguise, and the same can be said about any other claims of priority over domain-specific knowledge. For example, semantics cannot tell us whether atoms can be split or not. I will come back to this topic later.

In addition to this general worry, relativism also seems to be generally less plausible than ordinary contextualism, because it needs to come up with a complicated story to explain uses of value terms that explicitly involve a benefactor PP of the form 'for X', like in the following examples:

(7) Alice: Capitalism is good [for us].

(8) Alice: This sandwich is tasty [for the customers].

In both cases, it is not hard to come up with a scenario in which the PP with the benefactor needs to be inferred from the context. For example, in (8) it could be part of the common ground that Alice is a cynical chain restaurant manager who despises their own sandwiches and talks about the new product line *for the customers*. Examples like this are hard to explain from a purely relativist stance because they seem to require, at least in the most straightforward setting, a way

to accommodate the content of the PP at the level of Logical Form, which should propagate into the semantic content – which prima facie conflicts with the plain relativist semantics. But even if this problem can be solved technically, the general philosophical worry about expressions like 'good' in (7) is that these have clear-cut absolute uses, as the following variant of (1) indicates:

(9) **a.** Alice: Capitalism is good.
b. Bob: You mean for yourself?
c. Alice: No, I meant good {for us / in general / for everyone / for our country / for you / simpliciter / …}.

In light of the many options, claiming that 'good' is 'good for the speaker' or 'good for (some) assessor' by default seems philosophically dubious.[5] For 'good' in particular an ambiguity thesis that stipulates both a relational and nonrelational meaning seems more appealing, but even if only relational uses are allowed, then an indexicalist position according to which the benefactor is present as an open argument place that can be filled by a PP, or bound from the context, would explain examples like (9–c) much better. The burden of proof is on the relativist here, and at the same time there is nothing philosophically compelling about a relativist semantics for evaluative terms in general, even though it may be adequate for certain predicates of personal taste in spite of examples like (8).

2.3 SOCIAL EXTERNALISM TO THE RESCUE?

Putting relativism aside as being both empirically and philosophically unsatisfying, let me turn to general error theories. The kind of error theory I have in mind is not the one by Mackie (1977), but rather one based on some exaggerated form of social externalism. In this view, both Alice and Bob in (1) have the wrong lexical decomposition in mind, they are both wrong about the 'correct' meaning of her particular use of 'good' in the given conversational situation. Instead, like with 'elm' or 'arthritis', experts on goodness fix the meaning of 'good' in examples like (1). Maybe Alice agrees with the experts, maybe Bob agrees with the experts, but they might also both be mistaken if they

[5] Note that benefactor and assessor need not be the same and that the for-PP is primarily used to indicate a benefactor.

are themselves not experts on 'good'. Whatever they have in mind, according to the proposed view Alice's utterance ultimately means whatever the experts on goodness and democracy explain it means.

I argue in the next section that this account is more plausible than it might seem at first glance, but in the crude form presented so far it remains unacceptable. As Plunkett and Sundell (2013, p. 26–28) lay out, it is implausible to presume that experts generally fix the meaning of value terms, as they are the ones who most persistently disagree about these terms. Moreover, there may be no experts on 'good' and we also disagree about who counts as an expert. Would that be me or the Pope? Who decides? The problem becomes even more apparent with example (4). Who counts as an expert on torture? The one who tortures a lot? Donald Rumsfeld? Moral philosophers in general, or one in particular? Legal experts? To cut a long story short, although social externalism probably plays an important role in settling certain factual matters about 'elms' and the biological contexts in which a 'tomato' definitely *is* a fruit, it is implausible as a *general* solution to metalinguistic value disputes.

2.4 SEMANTIC PRIMITIVISM

There are two more interesting and influential responses to the problems raised by examples of metalinguistic value disagreement. The first one is *semantic primitivism*. As I understand this position, it states that value terms are not lexically decomposable in the sense that the predicates into which the term is decomposed exhaustively define the meaning of the value term.[6] For 'good' this position can be attributed to Moore (1903). In Moore's opinion 'good' cannot be defined by separate criteria, as I have suggested in the initial example, but

[6] Moore does not claim that a decomposition of 'good' and similar value terms in the sense of providing dictionary definitions is not possible, but rather that such a definition never provides a fully satisfying analysis of the term. This is compatible with the claim that a dictionary definition provides 'the' meaning of the term in question. However, in a truth-conditional setting does it follow from Moore's thesis that the decomposition does not exhaustively represent the truth-conditional contribution of the term and that the term must therefore have some primitive meaning in addition to whatever decomposition one proposes. I would like to thank an anonymous reviewer for comments that helped in making this formulation and the subsequent passage more precise.

rather has some primitive meaning. Words like 'good' are in Moore's view very similar to color predicates like 'yellow'. Although 'good' may be given some explanation in the form of a dictionary entry, this decomposition in Moore's opinion does not *define* the meaning of 'good'. In the more general context of contemporary truth-conditional semantics (which was of no concern to Moore), semantic primitivism asserts that a decomposition only explicates its meaning but does not deliver its full truth-conditional contribution. Consequently, there cannot really be any metalinguistic disagreement about such terms, at least not a form of disagreement that could be settled by a dispute about which lexical decomposition represents the correct or adequate meaning for a particular use of the term. Maybe speakers could resolve the disagreement in other ways, by pointing to instances of good things, for example, just like you may point to yellow objects in order to teach someone who speaks a language that does not have a lexically realized predicate for yellow what English 'yellow' means.[7] But it seems more plausible in such an account to stipulate that one or both of the speakers in value disputes like (1) fail to recognize the primitive property of being good that corresponds to Alice's use of 'good' or disagree about the question whether capitalism falls under it. In the first case one or both speakers are not fully competent and in the second case the disagreement is directly content-based. In both cases, the disagreement is not metalinguistic and our intuition has been explained away that value disputes like (1), (4) and (5) are at least partially about the terms involved.

The 'Paradox of Analysis' and Moore's Open Question argument have been used to argue for that position, and there is also a general worry that semantic decomposition could lead to a definitory vicious circle. Many attempts of defusing the 'paradox of analysis' have been made such as, for instance, Neo-Fregean approaches first proposed by Church (1946) and laid out in detail by Jacquette (1990), and since I believe that all of them are more or less successful and that there is no paradox, I will not further discuss the Paradox of Analysis in general here. The more specific open question argument is based on

[7] There are obvious concerns based on Quine (1964) about such an attempt, but we can ignore these for the sake of the argument. There are better reasons to reject Moore's conception that will be laid out below shortly.

the idea that one would need to justify any given decomposition for a use of 'good' into a logical combination of other criteria why the application of these criteria counts as good. For example, if being good is analyzed as whatever is commendable, one would have to ask the question why whatever is commendable counts as being good, thereby reasoning in circles. This is another version of the alleged paradox and in my point of view no more convincing than other formulations of it. On the contrary, if we explicate the meaning of a particular use of 'good', then we do not have to ask the additional question why the logical decomposition counts as being good, since that decomposition already explicates the meaning of that use of 'good'. The individual parts of such an explication do not even have to count as good, just as in an analysis of 'capitalism' each of the individual characteristics that together constitute a capitalist economic system do not themselves have to be counted as capitalist. When a general term, predicate, or term for an abstract object is explicated by semantic decomposition, then that complex meaning is constituted by the network of constraints of its individual parts and inferences that can be drawn from them within a holistic network of other such specifications of lexical meanings, much in the way computational ontologies work. Even if individual entities of such an ontology form part of another, more fine-grained ontology or are related in some systematic ways with another ontology, this does not constitute a vicious circle. Instead, the purpose and theoretical goals of the semantic analysis or explication dictates, from a practical point of view, how many levels of decomposition are appropriate and whether relations to other ontologies need to be explored. For semantics in general the very first level of decomposition is adequate; for a more philosophical analysis it is possible that further fine-grained analyses and ontological reductions would be more suitable. Much more would have to be said about this reply to semantic primitivism, but for lack of space I would like to leave it as is. This critique does not imply that there cannot be any primitive concepts, although perhaps their primitiveness is always relative to a given ontology.[8]

Let me end this section on semantic primitivism by pointing out that the burden of proof is on the primitivists side, and at least for value predicates their story seems unconvincing. Apart from more

[8] I would like to remain agnostic about this issue in this article.

general philosophical arguments against Moore on 'good' such as Geach (1956), there is also pervasive linguistic evidence that many, if not most, evaluative predicates are multidimensional, which puts them further at odds with a semantic primitivism that does not take into account multiple dimensions as part of the primitive meaning.[9]

2.5 METALINGUISTIC NEGOTIATION

A more compelling position takes the metalinguistic aspect of the examples discussed seriously. Burgess, Plunkett and Sundell have proposed a position they call 'Conceptual Ethics' in a series of papers with exactly that aim – see Burgess (2013), Burgess and Plunkett (2013), Plunkett and Sundell (2013), Plunkett and Sundell (2014), and Plunkett (2015). In their opinion, the disagreement in examples (2)–(6) is metalinguistic and they call the activity of discussing and determining how we *should* use words or concepts 'conceptual ethics', because it concerns normative issues. Such disputes are often worth having, they argue at length, because they are based on a concept's "… sociological facts about its sociological role" (Plunkett, Sundell 2013, p. 25), because there is something "…substantive at stake in how the relevant terms are used in the context […] and the speakers recognize this fact" (ibid.). As they lay out, these disputes also survive paraphrasing, a test devised by Chalmers (2011) to distinguish substantive from merely verbal disputes, so metalinguistic disagreement need not be merely verbal. According to their view, "…certain words (largely independent of which specific concept they express) fill specific and important functional roles in our practices" (Plunkett, Sundell 2013, p. 20), and discourse participants *negotiate* the best use of a term or concept on the basis of these existing functional roles. As Plunkett puts it, "[…] a metalinguistic negotiation […] is a dispute in which speakers each use (rather than mention) a term to advocate for a normative view about how that term should be used" (Plunkett 2015, p. 832).

[9] See Kennedy (1999), Kennedy and McNally (2005), Sassoon (2013), Sassoon and Fadlon (2017), McNally and Stojanovic (2017). It is unclear in which sense a meaning based on multiple dimensions could still be regarded as primitive, since it needs to somehow take into account different qualitative or quantitative orderings and their aggregation. In contrast to this, Moore considers 'good' to stand for a unary, natural predicate like 'yellow' does, for example.

A new area on this position is the attempt to combine a metalinguistic nature of disputes with the idea that these disputes are nevertheless substantial and worth having. However, despite the many examples they discuss, the function of the sociological role of a term or concept under discussion remains problematic in their approach. Suppose Alice presumes the definition of 'torture' preferred by the UN under which waterboarding would clearly fall under this concept, and Bob prefers the US definition that focuses on physical harm. Consider first that the prior sociological role of a word like 'torture' somehow could settle the dispute, i.e., plays the role of a corrigens and helps decide which is the 'right' definition. Then in a society in which torture is accepted and not sanctioned at all, waterboarding would be equally acceptable. So if the existing functional role ultimately settles the question, then Bob should have no problem with the UN definition, as long as he agrees with the existing practices, i.e., the lack of sanctions in this case. But it seems clear to me that this is not at all what Alice and Bob's are discussing in (4), the question under discussion in Alice's utterance is whether waterboarding *is* torture, irrespective of the existing social role of sanctioning torture. Consider second that the existing role does not settle the dispute in any way. Then it is not clear how the existing social role of 'torture' – that it is illegal, sanctioned, reprehensible, etc., in any civilized society – can help any of the discourse participants with their positions and why they do not merely talk past each other if they base their views on different definitions of what constitutes torture. Alice wants waterboarding to classify as torture, because she wants it to be sanctioned, and Bob does not want waterboarding to classify as torture, because he does not want it to be sanctioned. Since the social role does not settle the dispute in this scenario, they continue to talk past each other by propagating their favorite definition on the basis of different wants and desires.

It seems that a mixed approach is the most promising: The existing social and more broadly conceived functional role of a term or concept serves as a guideline for discussion, it partly settles the matter but in borderline cases a metalinguistic negotiation turns into a genuine value dispute.[10] The idea is perhaps that waterboarding is intuitively

[10] See Plunkett (2015, p. 851–852, 867) about mixed cases in other kinds of metalinguistic disputes.

no less reprehensible than another, similar practices that have already been classified as torture, and so it should also be regarded as torture.

However, upon some reflection it becomes clear that this line of reasoning is fallacious and that neither Alice nor Bob should or would want to argue that way. Alice could attempt to argue that waterboarding is torture *because* waterboarding is morally reprehensible and should be sanctioned, but that would make for a lousy argument. She could also argue that waterboarding is morally reprehensible and should be sanctioned, but for that argument she does not necessarily need to refer to torture at all and so it cannot represent the metalinguistic dispute in the example. It is simply another dispute. Instead, she needs to argue, if she intends to convince Bob rationally, that there is something wrong with the US definition, that it does not capture all aspects of what we commonly conceive as torture, and in a second step, that waterboarding sufficiently elicits many of these aspects for it to qualify as torture in this sense. She may then go on to explain that this concept of torture is faithfully represented by the UN definition, but this is again a separate issue. So what is under dispute really is the question whether waterboarding *is* torture, and only in a second step, she may intend to convey pragmatically, by stating (4–a), that it should also be sanctioned like torture.

Generally speaking, under normal circumstances the social practices associated with a value term result from the evaluative component of the value term and not the other way around. By regarding the metalinguistic dispute as an instance of negotiation, Burgess, Plunkett and Sundell reverse this direction of justification. That is in my opinion the main problem with their 'Conceptual Ethics', which otherwise adequately reflects our intuition that the dialogues in question are cases of genuine metalinguistic disputes worth having.[11]

[11] Sundell (2016) defends an even stronger thesis, that all value disputes can be explained as metalinguistic negotiation. This is criticized by Marques (2017), and apart from some assumptions about what counts as a possible explanation of value disputes (ibid., 42–43) that I consider too strong, I agree with her main objection that metalinguistic negotiation is neither necessary nor sufficient to have a value dispute and therefore cannot serve as the *only* explanation of value disputes. However, this critique only concerns the stronger thesis and not the weaker claim of Plunkett and Sundell (2013), Burgess and Plunkett (2013) , and Plunkett (2015) that many cases of value disagreement are metalinguistic. In my opinion, there is (at least) direct value disagreement, metalinguistic value disagreement based

In the next section I would like to offer a précis of their position that builds on this insight but brings something important back that metalinguistic negotiation, on the basis of an existing social role, cannot deliver: reality, insofar as it goes beyond social reality. I will also argue that metalinguistic disputes of the kind discussed so far occur in many other areas as well, almost inevitably whenever certain predicates, general terms and terms for abstract objects are involved, and that these types of disputes are a completely normal aspect of natural language use. They are part of the ordinary uses and functions of natural languages.

3 CORE MEANING AND NOUMENAL MEANING

Putnam's seminal article *The Meaning of 'Meaning'* is famously known for his Twin Earth thought experiment, but Putnam (1975b) also put forward a positive solution to the puzzle based on what he called 'meaning vectors'. This proposal is relevant to the above discussion. To quickly recap what it was about: A meaning vector contains syntactic and semantic markers, a stereotype in the sense laid out by Putnam (1975a), and a description of the external content of a term. For example, 'water' is a mass noun for a dispersed liquid substance with a stereotype that could be paraphrased as 'a drinkable colorless liquid essential for all life on earth' (the details or adequacy of this description do not matter here). In addition, the meaning vector contains 'H_2O' as a description of the external content of 'water', and as Putnam argued, the twin earth scenario shows that the meaning of 'water' cannot be adequately represented by the markers and a stereotype alone. However, Putnam (1975b) made it clear that in order to count as a competent speaker you do not have to know the externalist

on semantic underdeterminacy, which I lay out further below, and implicit value disagreement that I have investigated in detail […] and that forms the basis of Marques's main counter-argument against Sundell (2016); see Marques (2016, p. 47). However, if the central thesis of this article is correct, then there may be many more types of value disagreement anyway, since the noumenal meaning of value predicates is not fixed and people can endorse many different competing metaethical stances. Hence, the project of finding the necessary and sufficient conditions for an adequate explanation of any type of value disagreement seems to be doomed from the start.

content of a term, or otherwise nobody would have used the term competently before the rise of modern chemistry, and let us also not forget that at least in theory, as a remote possibility, it could happen that our views about chemistry are fundamentally mistaken and that water will turn out to be XYZ in the future.

Based on Putnam's proposal my suggestion is as follows. There is a *core meaning* that corresponds to the stereotype in his view. Mastery of this core meaning, be it implicit knowledge or an ability, is required by virtue of linguistic competence, but speakers only need to loosely converge on this type of meaning. It represents the lowest common denominator between competent speakers and primarily serves the purpose of communicating in order to solve cooperation problems. This type of meaning can be described in terms of truth-conditions like any kind of meaning, since the question of whether one should develop a truth-conditional semantics, as opposed to another type of semantics, primarily concerns methodology.[12] However, this type of meaning need not be truth-conditionally complete in the sense that the contribution of the core meaning of an expression to the whole meaning of the utterance will automatically make the utterance fully truth-conditionally evaluable. In a truth-conditional setting, the whole utterance may turn out to be a propositional skeleton (Bach 2004). In a more general understanding of meaning theory, geometrical approaches like Prototype Theory (Rosch 1983) and Conceptual Spaces (Gärdenfors 2000) may be more adequate for this type of meaning, because even though it is mandated by linguistic competence in a realistic approach different speakers will only converge imprecisely about a common core meaning.[13]

There is another kind of meaning that can be regarded a generalization of Putnam's specification of externalist content. I call it *noumenal meaning*, because it is directed towards external reality

[12] From the point of view of (desirable) methodological pluralism this is not a very good question. See Dekker (2011) for a defense of methodological pluralism in semantics.

[13] There are open problems with the logical combination of geometrical meanings and quantification, so these approaches cannot really serve as full replacements, but some advances have been made, see for example Aerts et al. (2013) and Lawry and Lewis (2016). Note further that a truth-conditional approach based on 'loose bundles' or propositional skeletons with nearness measures defined between them will face similar problems.

while at the same time not necessarily corresponding to something that exists in reality. In the case of natural kind terms that Putnam discusses in response to Kripke (1972) the noumenal meaning coincides with a description of externalist content and only denotes the natural kind provided that our current scientific theory is correct – the actual extension is fixed indexically. In his theory, Putnam does not need an externalist description of the extension of *other* general terms. In contrast to this, I understand noumenal meaning as that kind of meaning that reflects what a general term *really* means, i.e., it is intended to single out some particular aspect of reality that I call *noumenon*, following a long tradition in philosophy The noumenal meaning of 'water' is arguably H_2O, and the noumenon is H_2O in this case, provided that our current knowledge of physics and chemistry is correct. But it is important to realize that the story of Putnam-style semantic externalism cannot be extended to general terms and value predicates in general. Indexicalist externalism may be adequate for proper names and natural kind terms on the basis of current assumptions about science, whereas the meaning of other terms such as 'phlogiston' cannot be explained by reference to the actual world – phlogiston does not exist and consequently there is nothing like H_2O that represents its physical microstructure.

Going beyond what Putnam stated when he laid out his version of externalism, I would like to suggest now that according to the way we talk, in our actual linguistic practice, we commonly assume that expressions have a noumenal meaning that describes, singles out or otherwise captures an aspect of reality, but that we also often disagree about this meaning, and that this explains the above kind of metalinguistic disputes about value terms. Simply put, judging from the way we talk and the way language is supposed to work, we are all externalists by default but at the same time often disagree about what lexical decomposition of a term adequately describes an aspect of reality – and we even disagree about what counts as reality itself, one may add though, I will not address this more philosophical concern any further in what follows.

Take for instance the general term 'atom'. Its contemporary core meaning is something like 'extremely small building block of matter which can bundle together with similar building blocks to form molecules and is often depicted like a tiny solar system but is in fact way

more complicated'. Something like this. Its noumenal meaning used to be 'particle-like smallest indivisible building block of matter', but at some point there was disagreement about this lexical decomposition. Nature has informed us that atoms are divisible, so something with the original definition must have been wrong and the noumenal meaning had to be revised. In turn, this has likely triggered a revision of the core meaning, but probably rather slowly, because everyday talk was not under strong pressure to conform with reality in this case. It is our strife to get reality right that primarily motivates such disagreements.

Value disagreements like (1), (4), and (5) do not substantially differ from other cases of implicit or explicit metalinguistic disagreement. They are implicit in the examples but could be made explicit by a discourse participant at any time by asking questions like "What do you mean by 'good'?" or "How do you define 'torture'?" Under normal circumstances we strive for reality in such disputes *on the basis of a presumed shared existing core meaning* when we realize that the core meaning does not adequately capture reality in the context of the discussion. If we do not realize this, on the other hand, then there will be no explicit metalinguistic disagreement. As mentioned earlier, core meaning often suffices for our communicative purposes. There is no need to know exactly what the nature of time is when you ask "What time is it?" and want to catch the 18:30h train. I once ordered two tickets for a public outdoor swimming pool by uttering "Two tickets for normal adults, please." and the attendant answered: "Well, what does 'normal' really mean?" The joke was successful, because no noumenal meaning was under discussion; the core meaning sufficed for the cooperative behavior of selling and buying a ticket without any perks.

Although sometimes noumenal meaning depends on the larger theoretical context, for example definitions of 'normal' in the statistical sense versus 'normal' in the sense of a prototype, and sometimes operational definitions based on primarily practical considerations may be more or less appropriate and partly negotiable, noumenal meaning is generally directed towards reality and we generally intend it to be directed towards reality. We talk about numbers as if there was a Platonic realm of numbers, about values as if there were absolute values that we somehow perceive or intuit, and some of us talk about a particular god as if he or she existed. As the case of 'phlogiston' or 'Vulcan' reveal, however, the fact that we commonly assume that

certain expressions have a noumenal meaning does not imply that this meaning picks out a corresponding aspect of reality.

How does this theory explain value disagreement? The main difference to the contextualist position is that in a dual aspect theory linguistically competent speakers do agree on the core meaning and therefore never merely talk past each other. At the same time they do not merely negotiate how to best modify concepts in a way that suits prior social roles, though. Their disagreement can be about the noumenal meaning of value terms, which is not necessarily part of a shared lexicon, although experts may agree on the noumenal meaning of many expressions within their area. For value terms there may not be any such widespread agreement even among experts, but the way in which different speakers argue for and justify specific lexical decompositions that are supposed to reflect what a value term *really* means, i.e. decompositions representing noumenal meaning, is generally directed towards reality.

Finally, I would like to turn to the question of whether noumenal meaning is meaning at all, for an obvious critique off such a dual aspect theory is that only a core meaning is genuine meaning, since noumenal meaning is not required by linguistic competence and need not be shared or fixed. I believe this question to be ill-conceived, though, since Plunkett and Sundell (2013) and Plunkett (2015) show convincingly that many examples of metalinguistic disputes are clearly about the meaning of the terms involved. Should it be regarded as *linguistic* meaning? The answer can only be Yes, because noumenal meaning is associated with and tied to the terms involved, and by their existing core meaning also connected with the existing social role of a term. If the candidate for a noumenal meaning of a term deviates too far from its social role or core meaning, then we get jargon or technical definitions, or the candidate will be rejected. However, noumenal meaning is not linguistic in the sense of being understood by competent speakers on the basis of a shared lexicon, and for certain predicates, general terms, and terms for abstract objects there is also an important difference to Putnam's specifications of externalist content: When speakers dispute what an expression really means, then they need to have their own candidate for the noumenal meaning of that expression 'in mind', however that is spelled out in the detail; only then can the disagreement be considered implicitly or explicitly metalinguistic.

4 SUMMARY

How much does my proposal differ from Conceptual Ethics? Plunkett readily concedes that on the surface many metalinguistic disputes look like object-level discourse[14] and that it is possible that "[…] issues in conceptual ethics are settled by fully objective, mind-independent normative facts […] or that they are largely settled by facts about what the objective joints of reality really are […]" (Plunkett 2015, p. 860/1) while at the same time disputing "[…] the interpretation that takes them to be canonical disputes centered at object-level issues." (ibid., p. 867) However, as I have laid out above, BPS also focus on the social role of expressions and regard metalinguistic dispute cases of negotiating the best use of a term or concept, where they tend to couple the best use to a certain social role or function propagated by a discourse participant. I have argued that this way of looking at these disputes is inadequate in general, because it ultimately cannot explain why they are rational and substantive. If discourse participants merely battle their conflicting wants and desires of how to connect a term with some concept on the basis of an existing or a desired social role of that term, then they continue to talk past each other. So although BPS agree that some metalinguistic disputes might be based on the factual question whether the concepts used are more or less adequate to capture a relevant aspect of reality, their metalinguistic negotiation view emphasizes a normative aspect of these disputes that tends to make them irrational and mostly rhetoric. Instead, my suggestion is to take the 'best' in 'best use of a term or concept' to primarily mean 'best from an epistemic point of view', from the point of view of the theory of sciences, since in the end a general term has to capture a relevant aspect of reality adequately in order to be useful for theorizing about the world and for our conceptual systems.

To describe this general aspect of metalinguistic disputes in more detail, I have suggested, based on Putnam (1975b) and augmenting his original suggestion, that lexical meaning principally has two different aspects. When using sentences with general terms or terms for abstract objects speakers often implicitly or explicitly disagree about their *noumenal meaning*, since not even experts can be convincingly

[14] See Plunkett (2015, p. 828–830).

said to be able to determine what the respective terms *really* mean. This disagreement, however, always takes place on the basis of a prior agreement about the *core meaning* of the expressions involved, which is the meaning that is required by virtue of linguistic competence. When we use value terms like 'good' and value-laden terms like 'torture' we do not always know what these really mean in the given context of use, but we have a prior loose understanding of their core meaning. Disagreement about the noumenal meaning of terms is based on our strife to capture important aspects of reality that often go beyond the existing social function of these expressions associated with the core meaning. These types of disagreements are metalinguistic, because they implicitly concern the meaning of linguistic expressions. They can be substantial not only because of the prior social role of those expressions but also because our conceptual systems may capture relevant aspects of reality more or less adequately, and this adequacy is contested in corresponding metalinguistic disputes.

REFERENCES

Aerts, D., L. Gabora, S. Sozzo (2013). Concepts and Their Dynamics: A Quantum-Theoretic Modeling of Human Thought. *Topics in Cognitive Science*, *5*(4), 737–772.

Bach, K. (2004). Minding the Gap. In C. Bianchi (Ed.), *The Semantics/Pragmatics Distinction* (pp. 27–43). CSLI Publications.

Burgess, A. (2013). Conceptual Ethics I. *Philosophy Compass*, *8*(12), 1091–1101.

Burgess, A., D. Plunkett (2013). Conceptual Ethics II. *Philosophy Compass*, *8*(12), 1102–1110.

Chalmers, D. (2011). Verbal Disputes. *Philosophical Review*, *120*(4), 515–566.

Church, A. (1946). Review of "A Note on the "Paradox of Analysis"." Morton G. White, "The "Paradox of Analysis" Again: A reply." Max Black "Analysis and Identity: A Rejoinder." Morton G. White, "How Can Analysis Be Informative?" Max Black. *11*(4), 132–133.

Dekker, P. (2011). Jigsaw Semantics. In *The Baltic International Yearbook of Cognition, Logic and Communication 6 Formal Semantics and Pragmatics: Discourse, Context, and Models*, 1–26.

Egan, A. (2014). There's Something Funny About Comedy: A Case Study in Faultless Disagreement. *Erkenntnis*, *79*, 73–100.

Gärdenfors, P. (2000). *Conceptual Spaces*. Cambridge, MA: MIT Press.

Geach, P. (1956). Good and Evil. *Analysis*, *17*(2), 33–42.

Jacquette, D. (1990). A Fregean Solution to the Paradox of Analysis. *Grazer Philosophische Studien*, *37*, 59–73.

Kaplan, D. (1989). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In J. Almog, J. Perry and H. Wettstein (Eds.), *Themes From Kaplan* (pp. 481–564). Oxford, New York: Oxford University Press.

Kennedy, C. (1999). *Projecting the Adjective*. New York: Garland.

Kennedy, C., L. McNally (2005). Scale Structure, Degree Modification, and the Semantics of Scalable Predicates. *Language*, *81*(2), 345–381.

Kölbel, M. (2002). *Truth Without Objectivity*. London: Routledge.

Kripke, S. (1972). *Naming and Necessity.* In G. Harman and D. Davidson (Eds.), *Semantics of Natural Language*. Dordrecht: Reidel.

Lasersohn, P. (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy*, *28*(6), 643–686.

Lawry, J., M. Lewis (2016). Hierarchical Conceptual Spaces for Concept Combination. *Artifial Intelligence*, *237*, 204–227.

Ludlow, P. (2008). Cheap Contextualism. *Philosophical Issues*, *18*, 104–129.

MacFarlane, J. (2005). Making Sense of Relative Truth. In *Proceedings of the Aristotelian Society*, Volume 105, pp. 321–339.

MacFarlane, J. (2014, March 20). *Assessment Relativity: Relative Truth and its Applications*. Oxford: Clarendon Press.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin.

Marques, T. (2017). What Metalinguistic Negotiations Can't Do. *Phenomenology and Mind*, *12*, 40–48.

McNally, L., I. Stojanovic (2017). Aesthetic Adjectives. In J. Young (Ed.), *Semantics of Aesthetic Judgment* (pp. 17–37). Oxford: Oxford University Press.

Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Plunkett, D. (2015). Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy. *Inquiry*, *58*(7–8), 828–874.

Plunkett, D., T. Sundell (2013). Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint*, *13*(23), 1–37.

Plunkett, D., T. Sundell (2014). Antipositivist Arguments From Legal Thought and Talk: The Metalinguistic Response. In G. Hubb and D. Lind (Eds.), *Pragmatism, Law, and Language* (pp. 56–75). London: Routledge.

Putnam, H. (1975a). Is Semantics Possible? In *Mind,Language and Reality* (pp. 187–201). Cambridge: Cambridge University Press.

Putnam, H. (1975b). The Meaning of 'Meaning'. In H. Putnam (Ed.), *Mind, Language and Reality*, Number 2 in Philosophical Papers (pp. 215–271). Cambridge, UK: Cambridge University Press.

Quine, W. V. O. (1964). *From a Logical Point of View*. Oxford: Oxford University Press. second revised edition; first publ. 1953, 1961.

Rosch, E. (1983). Prototype Classification and Logical Classification: The Two Systems. In E. Scholnick (Ed.), *New Trends in Conceptual Representation: Challenges to Piagets Theory?* (pp. 73–86). Hillsdale: Lawrence Erlbaum Associates.

Sundell, T. (2011). Disagreements About Taste. *Philosophical Studies*, *155*(2), 267–288.

Sundell, T. (2016). The Tasty, the Bold, and the Beautiful. *Inquiry*, 59(6), 793–818.
Sassoon, G. (2013). A Typology of Multidimensional Adjectives. *Journal of Semantics*, *30*, 335–380.
Sassoon, G., J. Fadlon (2017). The Role of Dimensions in Classification Under Predicates Predicts Their Status in Degree Constructions. *Glossa*, *2* (1)(42), 1–4.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 161–178
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.08
Rozprawa

Federico L.G. Faroldi[*], Andrés Soria Ruiz[**]

THE SCALE STRUCTURE OF MORAL ADJECTIVES

Summary: In this paper we discuss how and whether moral adjectives fit a well-known semantics for gradable adjectives. We first test whether moral adjectives are relative or absolute adjectives. The preliminary results suggest that moral adjectives don't fall neatly under either category. In the second part we tackle the question of the scale of moral adjectives in a more theoretical fashion, i.e. by investigating their possible scales with mathematically precise tools.

Keywords: moral adjectives; scalar semantics, multidimensionality, meta-ethics

1. INTRODUCTION

"It is morally better to keep a promise than to save a life." Sentences like these show that moral adjectives such as 'good', 'bad' '(un)ethical', 'cruel', '(im)moral', 'virtuous' or 'despicable' are gradable, that is, they place their objects on a scale.[1] Can we say more about their semantics, and the types of scales they use? In this paper we apply available tests

[*] Centre of Logic and Philosophy of Science, Ghent University and Research Foundation – Flanders (FWO). E-mail: federico.faroldi@ugent.be

[**] Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL Research University, CNRS, Paris France. E-mail: andressoriaruiz@gmail.com

[1] There are thick and thin terms in the previous list. We will not, however, discuss that distinction here (see Väyrynen 2016 for a recent overview of this debate).

to see how and whether moral adjectives fit a well-known semantics for gradable adjectives.

But before that, why focus on moral adjectives? Moral adjectives form an easily recognizable class of words, although this class most likely does not carve a natural joint in language, so to speak. By way of introduction and justification, we note that the present project is inspired by three main trends that are very much alive in the current literature in philosophy and linguistics.

First, recent years have witnessed a surge of literature on the semantics and pragmatics of words that share some (if not many) semantic features with moral vocabulary, such as subjective terms – most eminently, predicates of personal taste, i.e. *fun* and *tasty* (Kölbel 2003; Lasersohn 2005; Stephenson 2007; Stojanovic 2007; Meier and van Wijnbergen-Huitink 2016 *a.m.o*), and, more recently, aesthetic adjectives too (see Liao and Meskin 2015, Liao *et al* 2016; Stojanovic 2016; Umbach 2016) – modals of various flavors (stemming from Kratzer's seminal contributions, see her 2012) and other purportedly normative expressions, such as *know* or *rational*. To our knowledge, moral expressions have received comparatively little attention.[2]

Secondly, and by contrast, moral vocabulary has long been discussed in meta-ethics. The tradition loosely inaugurated by Moore took very seriously the idea that the inquiry into the linguistic properties of moral words was the right way to gain philosophical insight into the nature of the moral concepts that we deploy. Ordinary language philosophers followed suit, and language-oriented analytic meta-ethics lived something of a heyday that lasted approximately until the second half of the XXth century (Darwall et al 1992).

Its decay coincided in time with the emergence of formal semantics, and more specifically, with the appearance of degree semantics, which is the third tradition upon which this project rests. Since the 1970s, much literature in formal semantics has focused on the semantic properties of gradable expressions, often tackling difficult and well-established problems in philosophy, such as vagueness (Cresswell 1976; Klein 1980; Barker 2002; Fara 2003; Kennedy 2007; Lassiter 2016).

[2] A curious example is MacFarlane's: his 2014 covers pretty much all the range of expressions that have been discussed in the contextualism-relativism-expressivism debate, but never does he tackle moral adjectives (even though he does devote a whole chapter to the modal *ought*, which is moral under one interpretation).

Thus, our aim is to employ the resources of degree semantics to explore the semantic properties of moral adjectives, starting from the basic observation that these expressions admit comparisons. A worry at the outset, however, is that from the fact that there is a comparative it doesn't follow that there is a scale. If we just consider *good*, for instance, and a general betterness relation, it is somewhat common in the literature to assume that it cannot be a total relation over all alternatives, because of incommensurability or various forms of non-comparability (for an introduction see Chang 1997), while other authors like Temkin (2012) additionally argue that the 'betterness' relation is not transitive. We acknowledge these worries and we hold their discussion, as well as a precise definition of what we mean by a 'scale', until section 3.

We assume, as it's standard, that the semantic value of a gradable adjective G is a function μ from an arbitrary type α (individuals, for instance) to degrees on a scale:

$$[\![G]\!]^{M,w}_{\langle\alpha,d\rangle} = \lambda k_{\alpha}.\mu_G(k)$$

The positive form of a gradable adjective G_{POS} predicates a property of the degree possessed by α:

$$[\![\text{is } G_{POS}]\!]_{\langle\alpha,t\rangle} = \lambda k_{\alpha}.\mu_G(k) \text{ is } P$$

We start (sect. 2) by considering the properties that the positive form of moral adjectives ascribes to their objects. In particular, we test whether moral adjectives are relative or absolute adjectives. To do this, we look at the entailment patterns of moral adjectives, their compatibility with modifiers such as *perfectly* and *slightly* and their sensitivity to comparison classes. The preliminary results point towards the fact that moral adjectives don't fall neatly under either category. Nonetheless, we take the available data to suggest that moral adjectives are relative, even though they display two features that have often been associated with absolute adjectives: they are insensitive to comparison classes and they admit modification by modifiers like *perfectly* and *slightly*. We present a hypothesis to account for these results.

In section 3 we turn to the properties of the measure function μ. Thus, we tackle the question of the scale of moral adjectives in a more theoretical fashion, i.e. by investigating their possible scales with mathematically precise tools. We show results of two classes: first,

boundedness properties are compatible both with ratio and interval scale and are, therefore, uninformative; second, we prove that moral adjectives cannot always have a ratio scale because of some non-additive cases. These results pave the way to the multidimensionality hypothesis.

2. RELATIVE OR ABSOLUTE?

Gradable adjectives are usually classified as relative (*rel*) and absolute (*abs*), depending on the property that the positive form ascribes to their object's degree (property P in the above formalization). The positive form of relative adjectives predicates that the object's degree surpasses a certain threshold, which is an intermediate point on a scale whose precise value is determined contextually; absolute adjectives are those whose positive form predicates that their object's degree is a scale endpoint. *Tall* is a typical example of a relative adjective: to be tall amounts to possessing a degree of height that exceeds a certain threshold, which is an intermediate point on the height scale.

Absolute adjectives are called *maximum* (abs_{max}) or *minimum* (abs_{min}) standard if the endpoint is the upper or lower scale endpoint, respectively. *Full* is an abs_{max} adjective: to be full is to have a maximum degree of the property of fullness (namely, to have as much content as capacity); *dirty* is an abs_{min} adjective, since to be dirty is to possess a non-zero degree of dirtiness.

In addition to this, gradable adjectives are positive or negative, depending on whether modification by *-er* (or *more*) denotes a higher degree on the relevant scale. Thus, *tall* is positive, while *short* is negative; *full* is positive while *empty* is negative; and *dirty* is positive while *clean* is negative.

Interestingly, common tests for this distinction used in the literature do not give stable results when applied to moral adjectives. In particular, while certain entailment patterns suggest that these adjectives are relative, their behavior with respect to PPs denoting comparison classes, as well as their admissibility of certain modifiers suggests that they are absolute (see Liao and Meskin 2015; Liao et al. 2016). Finally, some of these tests give different results for positive and negative adjectives, so we'll consider adjectives of both types, namely

virtuous/ethical/generous (positive) and *cruel/despicable* (negative). Let's review these tests briefly.

We'll start by looking at three tests according to which moral adjectives come out as relative (Kennedy 2007, sec 3). First, given that absolute adjectives in the positive form denote endpoints on a scale, the following schemata hold:

if *x* is abs_{max}, then *x* could not be abs_{max} -er.
if *x* is not abs_{min}, then *x* does not possess any degree of abs_{min}-ness.

To test these entailment patterns, we consider whether sentences that do not respect them are acceptable. That is, we consider whether the conjunction of the antecedent clause and the denial of the consequent are coherent. If not, then the pattern holds for that adjective. Consider the following two cases with *full* (abs_{max}) and *open* (abs_{min}), and compare them with *tall*:

1) # The glass is full, but it could be fuller.
2) # The door is not open, although it's ajar.
3) Natalia is tall, but she could be taller.
4) Matheus is not tall, but he possesses some height.

As the examples show, the relevant constructions are not coherent when we use absolute adjectives, thereby showing that these adjectives do not respect those entailment patterns. When we use a relative adjective like *tall* however, those constructions are coherent. Positive and negative moral adjectives pattern like relative adjectives in this test:

5) What she did was ethical, although it could have been more ethical.[3]
6) Being vegetarian is not virtuous, although it has some degree of virtuousness.
7) What he did was despicable, although it could have been more despicable.
8) Eating animals is not cruel, although it has some degree of cruelty.

[3] Some may balk at this: after all, it seems that, for what she did to be more ethical, she should've done something else altogether. What she in fact did cannot have more or less moral value than it actually has. We share the intuition, but we don't take that intuition to say anything about the scalar properties of *ethical* (or any moral adjective, for that matter). It does however, say something about the objects of moral evaluation.

Secondly, absolute adjectives also satisfy the following entailment pattern:

if *x* is more abs_{min} than *y*, then *x* is abs_{min}.
if *x* is more abs_{max} than *y*, then *y* is not abs_{max}.

Again, compare for illustration *full* and *dirty* to *tall*:

9) # The shirt is dirtier than the jacket, although the shirt isn't dirty.
10) # The glass is fuller than the vase, although the vase is full.[4]
11) Natalia is taller than Matheus, although she isn't tall / Matheus is tall.

With respect to these two patterns, positive and negative moral adjectives give slightly different results: positive adjectives do not show these patterns, hence they behave like relative adjectives. Negative adjectives show the second pattern, but our judgments about the former are less clear:

12) Although it isn't ethical, animal testing for scientific purposes is more ethical than for cosmetic purposes.
13) Donating to a charity is virtuous, and volunteering for a charity is more virtuous.
14) ?? Although it isn't despicable, hiding your office mate's keys is more despicable than eating their snacks.

Paying bribes is despicable, although accepting them is more despicable.

Using a negative moral adjective like *despicable* in a comparison does seem to suggest that both terms of comparison deserve the positive form. However, we are skeptical that this particular pattern of inference is an entailment. The fact that a sentence like (14), even if marked, is not completely ruled out, suggests that the inference is to some extent cancelable, so it might well arise due to implicature or some other mechanism.[5]

[4] Sometimes constructions like these are acceptable, cf. *the first theater was full, and the second was fuller.* However, when we say such things we are arguably speaking imprecisely. That is, we don't really mean that the first theater is full in the literal sense of not having any free seats, but rather that it is very crowded, or something like that (see Kennedy 2007, p. 23–24).

[5] Bierwisch (1989, p. 89) makes a similar observation about the pair of comparatives *besser* (*better*) / *schlechter* (*worse*): comparisons with *schlechter*, but not with

The previous observations strongly suggest that moral adjectives are relative. On the other hand, the behavior of moral adjectives with respect to their thresholds' sensitivity to *comparison classes* seems to suggest that they are *not*.

As we mentioned, the thresholds for relative adjectives are determined in context. More specifically, the value of a threshold can be shifted by reference to a comparison class: *tall for a basketball player* and *tall for a 5 year old* establish different thresholds for the positive form of the adjective *tall*. By contrast, absolute adjectives do not show such sensitivity to comparison classes: modification by a comparison class forces the interpretation that the positive form doesn't apply: the PP *for a TV antenna* in *straight for a TV antenna* does not shift the threshold of straightness, but rather is most naturally taken to mean *not straight*.

Liao et al. (2016) observe that *aesthetic* adjectives pattern in this respect like absolute adjectives: rather than shifting the relevant threshold, mentioning a comparison class suggests that the bare positive form doesn't apply (they also note that such constructions – namely 'aesthetic adjective + *for* comparison class' – appear very rarely in *corpora*):

16) Anyone who calls someone 'beautiful for an older woman' does not get my love.
17) Elegant for a Best Western.[6]

As it's clear, mentioning the comparison class in both cases suggests that the bare positive form does not apply. We think that the same applies to moral adjectives: explicit reference to comparison classes does not shift the threshold but rather suggests that the bare positive form wouldn't apply. Moreover, such constructions sound slightly marked to our ear:

18) Giving alms is a generous act (?? for a miser)
19) What they did was not despicable (?? for a vile person)

---

besser, invite the inference that the positive form *schlecht* (*bad*) applies to both *relata*. However, later (p. 206–207) he says that both comparatives in the pair *schöner* (*prettier*) / *häßlicher* (*uglier*) invite such inference. We acknowledge that the inferences are there, but the fact that they seem in general defeasible suggests to us that they are not entailments, and hence that the hypothesis that these adjectives are relative-standard is not decisively challenged by these observations.

[6] These are Liao et al (2016)'s examples (4a–b), one of the few instances of that construction that they found in *corpora*.

Lastly, the scales of gradable adjectives can be open on either, neither or both ends; and the acceptability of modifiers like *slightly* and *perfectly* (which pick out minimal/maximal endpoints) with positive and negative adjectives has been taken to reveal information about the type of scale lexicalized by a given pair of adjectives (Kennedy 2007). For example, the pair *dirty, clean* lexicalizes an upper-closed scale, as shown by the following pattern:

20) perfectly / ?? slightly clean
21) ?? perfectly / slightly dirty

Interestingly, moral adjectives (in particular, the pair *ethical, unethical*) show the same pattern:

22) perfectly / ?? slightly ethical
23) ?? perfectly / slightly unethical

Kennedy proposes the generalization that adjectives that lexicalize totally open scales are relative, while adjectives that lexicalize scales closed on one or two ends have absolute interpretations. If this generalization is correct, moral adjectives are absolute.

To summarize: Entailment patterns suggest that moral adjectives are relative; while lack of sensitivity to comparison classes and the fact that pairs of moral adjectives admit scale endpoint modifiers suggests that they are absolute.

More tests can and ought to be carried out, but we venture that the hypothesis that moral adjectives are relative is better supported by the data (i.e. the observations about entailment patterns appear particularly telling), whereas tests that suggest that moral adjectives are absolute are based on observations that can be explained alternatively: first, insensitivity to comparison classes *per se* does not show that moral adjectives are absolute, but simply that their thresholds' are rigid in a way that the thresholds of other relative adjectives are not. Why is this so? We may rehearse the following preliminary answer: whether an object falls under a certain moral concept, say *whether an action A is despicable*, depends – in a way to be carefully spelled out – on the moral values of the person who is considering that question. But, crucially, it does not depend on the actions that we may compare *A* with. Thus, it is to be expected that mentioning a comparison class does not change our evaluation.

Secondly, the association of open scales to relative adjectives and of partially closed scales to absolute adjectives is a generalization that has been challenged. In particular, Lassiter and Goodman (2013; 2015) challenge the claim that relative adjectives are associated with an open scale. Moral adjectives may be an exception to the second generalization.

Moreover, the acceptability pattern of modifiers like *perfectly/slightly* can receive an alternative explanation in terms of the *multidimensionality* of moral adjectives. Multidimensional adjectives (Sassoon 2013, 2016; called *evaluative* in Bierwisch 1989) are adjectives that denote properties that can be possessed relative to different respects or dimensions. The pair *healthy*, *sick* is a paradigmatic example: one can be healthy or sick with respect to various dimensions, such as blood pressure, cholesterol or blood sugar level. By contrast, there is but one dimension associated with an adjective like *tall* (i.e. height).

An available test for multidimensionality is the admissibility of "dimension-accessing" operators and modifiers, such as the PPs *with respect to ...* and *in some/most/every respect(s)*. Compare *healthy* (multidimensional) to *tall* (dimensional):

24) Natalia is healthy in some/most/every respect(s) / … with respect to blood pressure, but not cholesterol.
25) # Matheus is tall in some/most/every respect(s) / ...with respect to ?

Moral adjectives are multidimensional, as shown by the fact that they also admit such "dimension-accessing" operators and modifiers:

26) What she did was despicable with respect to its level of cold-bloodedness.
27) Natalia is virtuous in some/most/every respect(s).

Importantly, multidimensional adjectives admit maximum/minimum standard modifiers like *perfectly/slightly*. But when they combine with multidimensional adjectives, these modifiers do not reference endpoints on a scale, but rather, they quantify over the dimensions associated with the relevant adjectives: *perfectly healthy* means, roughly, *healthy in all respects/dimensions*; and *slightly sick* means *sick in some respect/dimension* (see Sassoon 2016, p. 10). Furthermore, those modifiers are not interchangeable: *perfectly* only admits multidimensional adjectives that have a *universal* interpretation, that is,

adjectives for which the bare positive form requires that the individual is above the relevant threshold for *every* dimension. By contrast, a modifier like *slightly* is only acceptable with multidimensional adjectives that are interpreted *existentially*, that is, where the bare positive form only requires for the individual to be above the threshold for *some* dimension. *Healthy* and *sick* are universal and existential, respectively (Sassoon 2016, p. 3, proposes the generalization that positive multidimensional adjectives tend towards universal interpretations, while negative adjectives tend towards existential interpretations). Thus, the following pattern emerges:

28) perfectly / ?? slightly healthy
29) ?? perfectly / slightly sick

Thus, we submit that that is how the acceptable APs in (22)–(23) are to be interpreted: *perfectly ethical* means *ethical in every respect*; *slightly unethical* means *unethical in some respect*. Importantly, these interpretations are consistent with the claim that such adjectives are relative.

This suggests the following modification of the lexical entries that we started out with. The semantic value of a multidimensional and *universal* gradable expression $G+$, in its positive form, and according to the multidimensionality hypothesis, would now be

$$[\![\alpha \text{ is } G+]\!]^{M,w,d} = 1 \text{ iff } \forall d \in \Delta_G, \mu_G^d([\![\alpha]\!]^{M,w}) \geq \delta_G^d;$$

while the semantic value of a multidimensional and *existential* gradable expression $G-$ (in its positive form) would be

$$[\![\alpha \text{ is } G-]\!]^{M,w,d} = 1 \text{ iff } \exists d \in \Delta_G, \mu_G^d([\![\alpha]\!]^{M,w}) \geq \delta_G^d,$$

where d is a parameter ranging on the set of relevant dimensions Δ_G (we refer the reader to Sassoon 2016 for more details).

Thus, we hypothesize that moral adjectives are multidimensional, relative gradable adjectives whose threshold is not sensitive to modification by comparison class.

Before we move on however, one may wonder: why venture such generalization? Why not entertain the possibility that, for instance, some moral adjectives are absolute-standard? There is an empirical and a theoretical reason for this: on empirical grounds, we simply have yet to come across a moral adjective that shows a markedly

different semantic behavior from what we observe here. Our choice of examples in this paper is not selective; we have picked moral adjectives at random and observe that they reveal a – relatively stable – semantic pattern. On theoretical grounds, we consider it a plausible hypothesis that all moral adjectives (and possibly other evaluative adjectives as well) have similar semantic properties due to their sharing certain mathematical or structural properties inherent to values in general. For instance, it's plausible that orderings of objects according to their moral value are only partial. Thus, one might reasonably expect that all moral adjectives give rise to partial orderings as well (thereby allowing for incomparabilities). However, we think that if the foregoing hypothesis were true, that would constrain the mathematical properties of the scales of these adjectives (that is, the properties of the measure function μ, see next section), but not necessarily the properties of their thresholds. In other words, that moral adjectives track mathematical properties of values is compatible with moral adjectives being relative or absolute standard. So we take there to be theoretical reasons to expect some uniformity at the level of the internal scale of these adjectives, and empirical – but of course defeasible – reasons to expect uniformity at the level of thresholds for the positive form.

3. A FORMAL PERSPECTIVE ON SCALES

In this section we offer an additional argument for our thesis, partially independent from the preceding section, as we slightly change our point of view. Instead of taking a "bottom-up" approach, as it were, starting from linguistic tests, we now adopt a "top-down" perspective. Even if the fact that a moral adjective is multidimensional allows for each dimension to have different kinds of scale, plausibly the dimensions have to be aggregated or combined into a single scale, and we can still explore what are the properties of *that* scale based on the observable linguistic properties of the adjective itself.

This is what we now turn to, namely a more mathematically oriented discussion of the properties of scales for moral adjectives, i.e. we venture into the properties of μ. We show results of two classes: first, boundedness properties are compatible both with ratio and interval scale and are, therefore, uninformative; second, we prove that

moral adjectives cannot always have a ratio scale because of some non-additive cases. These results, we submit, are a further argument for the multidimensionality hypothesis.

For concreteness, we consider a structure of individuals, an ordering, and a sum operation: $S = (S,\sqsupseteq,\sqcup)$. We seek to define an order-preserving mapping μ from $S = (S,\sqsupseteq,\sqcup)$ into $(\mathbb{R},\geq,*)$, where $*$ might be addition or a more complicated operation. Such a mapping is called a *representation* of S; to show there is such a function is to prove a representation theorem.

There are three kinds of scales used in empirical and social sciences (barring *absolute* or *nominal* "scales", which just label elements without any kind of quantitative ordering or measurement – see Duncan and Narens (1987) for a concise but high-level survey):[7] *Ordinal scales* represent the ordering among the elements to be measured with the usual ordering among (real) numbers, with no further properties assumed: we do not know anything about the respective distances between elements. Examples of ordinal scales are the scales that we come across in surveys, such as those ranging from "1 = very boring" to "5 = very exciting". More precisely, μ is an *ordinal scale* iff if for all $a,b \in S$ and $a \sqsupseteq b$, then $\mu(a) \geq \mu(b)$. *Interval scales* represent the ordering with the usual ordering among (real) numbers, but where differences are meaningful, i.e. differences represent actual distances between the elements to be measured. Examples of such scales are the Celsius or Fahrenheit scales. More formally, μ is an *interval scale* if the following conditions are met: (i) if $a,b \in S$ and $a \sqsupseteq b$, then $\mu(a) \geq \mu(b)$; (ii) $\mu(a \sqcup b) = k\mu(a) + p\mu(b) + q$, with $k,p,q \in \mathbb{R}^+$; (iii) for any μ' satisfying (i) and (ii), there are n, m, k', p', q' with $n,k',p',q' \in \mathbb{R}^+$, $m \in \mathbb{R}$ s.t. $\mu'(x) = n\mu(x) + m$, $q' = nq + m(1 - k - p)$, $k' = k$, $p' = p$, i.e. an interval scale is unique only up to positive affine transformation. The following known theorem lists the conditions on S for the existence of an interval scale (see Krantz et al. 1971, p. 294ss.):

Theorem 1 Let $S = (S,\sqsupseteq,\sqcup)$ be a structure such that for all a, b, c, d, e, f, it is monotonic ($a \sqsupseteq b$ iff $a \sqcup c \sqsupseteq b \sqcup c$), bisymmetric ($(a \sqcup c) \sqcup (c \sqcup d) \approx (a \sqcup c) \sqcup (b \sqcup d)$, restrictedly solvable, Archimedean and $\sqsupseteq$ is a weak ordering. Then μ is an *interval scale*.

[7] Although an infinite number of different scales can be characterized more precisely with reference to homogeneous and point-uniqueness features.

Finally, *ratio scales* represent the ordering with the usual ordering among real numbers, where difference and multiplication are meaningful so that ratios are preserved. Examples of such scales are the weight or height scales. More precisely, μ is a *ratio scale* if the following conditions are met: (i) if $a,b \in S$ and $a \sqsupseteq b$, then $\mu(a) \geq \mu(b)$; (ii) $\mu(a \sqcup b) = \mu(a) + \mu(b)$; (iii) for any μ' satisfying (i) and (ii), there's an $n \in \mathbb{R}^+$ s.t. $\mu'(x) = n\mu(x)$, i.e. a ratio scale is unique only up to linear transformation. The following known theorem lists the conditions on S for the existence of a ratio scale (see Krantz et al 1971):

Theorem 2 Let $S = (S,\sqsupseteq,\sqcup)$ be a structure such that it is *positive*, monotonic, solvable, Archimedean and $\sqsupseteq$ is a weak ordering. Then μ is a *ratio scale*.

We moreover say that if a scale obeys condition (ii) in the previous paragraph for all distinct $a,b \in S$, it is *additive*. Ratio scales are therefore additive. Interval scales are not additive, in that the "sum" of two values lies in the middle. Weighted averages are an example of such an operation.

One naturally wonders whether the scale of moral adjectives is among these options. At the very outset, we can exclude the idea that moral adjectives have an ordinal scale if we assume they are at least minimally more structured than a mere ordering. Lassiter (2016) offers an argument in this sense. Thus, in the following we consider three data points in favor of either ratio or interval scales: boundedness, modification with 'twice', and inferences about concatenation.

First, boundedness. In the previous section we considered the possibility that pairs of moral adjectives were partially closed. Can boundedness properties reveal something about scale structure? Unfortunately, no. The relationship of interval and ratio scales with boundedness is summarized in the following two theorems:

Theorem 3 Let μ be a *ratio scale*. Then it can be open; lower-bounded; upper-bounded; fully closed.

Proof (Sketch) We construct an example of each. Take a measure which respects the usual probability axioms. We check that it respects conditions (i)–(iii) of ratio scales, therefore it is a ratio scale. Normally, it is fully closed, i.e. its range is [0,1]. Remove one or both of the

endpoints and one gets the following cases: (0,1],[0,1),(0,1). We check that each measure, so modified, respects conditions (i)–(iii).

Theorem 4 Let μ be an *interval scale*. Then it can be open; lower-bounded; upper-bounded; fully closed.

Proof By analogy to the above.

Therefore, we can't exclude the possibility that moral adjectives have either a ratio or an interval scale depending on the hypothesis that they are open or partially or fully closed. Nonetheless, a second data point may be the admissibility of modifiers such as 'twice': when employed in certain contexts, 'twice' appears to point toward a ratio scale, since there is explicit talk of additions or multiplications (see Lassiter 2016). And indeed, 'twice' is acceptable as a modifier of moral adjectives; see the following examples in combination with 'cruel':

30) You're unwilling to buy a mousetrap, but you're happy to buy a sticky platter? [...] If anything, it's twice as cruel as a mousetrap. At least the mouse is killed instantly. Imagine the slow agonising death of being stuck to a plate.[8]
31) Well you've taken the best of our sailors, and, You've taken my love from me. [...] The sea is twice as cruel.[9]
32) If you are guilty or something, just break up. Telling is twice as cruel.[10]
33) If anger is cruel, then jealousy is (being) twice as cruel.[11]

However, the use of *twice*, by itself, is not conclusive evidence. For all these sentences show, speakers could be speaking loosely or metaphorically. Moreover, the lack of any standard measure of moral value impedes any precisification of what anyone could mean by describing something as 'twice as cruel' as something else.

[8] http://www.bangkokpost.com/print/1037653/

[9] Song *Oh Cruel Sea!*, by Jonny Nutt.

[10] http://whisper.sh/whisper/05338d292f99f2736bfbf12c4361ad8c88ecd1/If-you-are-guilty-or-something-just-break-up-Telling-is-twice-as-cruel

[11] Ashish Raichur, *Laying the Axe to the Root*, p. 33. Quite interestingly, this example opens up the possibility of investigating higher-order scales, i.e. scales introducing comparisons of properties (according to a higher-order property), rather than of individuals. For this reason we employ this example with some hesitation.

Finally, we should also take into account facts directly about concatenation. In particular, the following fact holds:

Theorem 5 If μ is not additive, then it is not a ratio scale.

It is enough to show that moral adjectives are not (always) additive to show that they do not have a ratio scale.

Consider the following example:

34) It is cruel to make him run in the heat.
35) It is cruel to make him run in the rain.
36) *It is twice as cruel to make him run in the heat and in the rain.

(36) is plausibly unacceptable as an inference from (34) and (35), for the concatenation of the two factors (heat and rain) might in fact make the run pleasant, or at any rate not nearly as cruel as if either factor was present singly, (provided we understand (36) as a fusion of the two situations, rather than a mere serial repetition),[12] so that even the following becomes unacceptable as a consequence of (34) and (35):

37) * It is cruel to make him run in the heat and in the rain.

The concatenation of the two factors, in this case, would even be less cruel, if at all, than either factor taken alone.

The choice of 'cruel' was somewhat arbitrary: a similar reasoning seems to apply to thin adjectives as well. We thus suggest that moral adjectives do not have ratio scales. However, we cannot conclude yet that moral adjectives have an interval scale, since in fact, further conditions need to be met for them to do so. Lassiter 2016 argues that 'good' has such a scale essentially for abductive reasons. But both the linguistic data and theorems 3, 4, and 5 are not enough to conclude that all moral adjectives have an interval scale. In fact, there are infinitely many different scales to choose from. At present, we have to leave open two other possibilities as well: *first*, moral adjectives do not

[12] A possible objection may identify the reason for the fact that (36) does not follow from (34) and (35) not in the scalar properties (or absence thereof) of cruelty but in the non-standard behavior of the conjunction in this particular case: "to make him run in the heat and in the rain" would not be equivalent to "to make him run in the heat and to make him run in the rain." While we don't think this is the case, such concerns are immaterial to our point to the extent that there is a mechanism to talk about the "fusion" of the two situations. We thank an anonymous referee for pressing on this point.

have an interval scale, but rather yet another scale, perhaps with a very complicated structure, or at any rate, a non-standard structure; *second*, moral adjectives may not have a unique scale at all (even up to the appropriate notion).

Both these possibilities would be explained by the hypothesis that moral adjectives are multidimensional, if their multidimensionality is constructed in a way compatible with associating possibly different scales to each dimension.

4. CONCLUSION

In this paper, we've (preliminarily) observed that moral adjectives are multidimensional, relative-standard adjectives. More tests, and especially, more experimental results should be obtained, in order to decide between the alternative features of scales presented here. We leave this for future work, but we note that getting clear on the scale of moral adjectives has important consequences for ethical theory, provided that moral language is somewhat indicative to ethics. Of course, one may argue that linguistic evidence is a poor indicator for philosophical analysis, and the structure of these concepts is ultimately a matter of normative philosophical theory, as Erich Rast (p.c.) notes, and that's why linguistic data may seem generally inconsistent. A natural way of reconciling the apparent inconsistency of the examples, however, is to allow for many different measures which are to be specified on a case-by-case basis. Thus, there would be different *kinds* of scales (as opposed to different scales equivalent up to some notion to be specified): sometimes the scale would be additive, and sometimes not, sometimes it would be ratio-like, sometimes interval-like. This scenario fits well with moral particularism. The links between multidimensional approaches and particularist approaches to ethical theory are, to the best of our knowledge, still unexplored.[13]

[13] The authors would like to thank the audience at the PhilLang 2017 conference, Natalia Karczewska, Erich Rast, Isidora Stojanovic, Matheus Valente and two anonymous referees. Andrés Soria Ruiz was supported by Obra Social La Caixa and grant numbers ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL.

REFERENCES

Barker, C. (2002). The Dynamics of Vagueness. *Linguistics and Philosophy*, *25*(1), 1–36.

Bierwisch, M. (1989). The Semantics of Gradation. In M. Bierwisch and E. Lang (Eds.), *Dimensional Adjectives* (pp. 71–261). Berlin: Springer-Verlag

Chang, R. (1997). Introduction. In R. Chang (Ed.), *Incommensurability, Incomparability and Practical Reason*. Cambridge, MA: Harvard University Press.

Cresswell, M. (1976). The Semantics of Degree. In B. Partee (Ed.), *Montague Grammar* (pp. 261–292). New York: Academic Press.

Darwall, S., Gibbard, A., Railton, P. (1992). Toward Fin de Siècle Ethics: Some Trends. *The Philosophical Review*, *101*(1), 115–189.

Fara, D. G. (2003). Gap Principles, Penumbral Consequence, and Infinitely Higher-Order Vagueness. In J. C. Beall (Ed.), *New Essays on the Semantics of Paradox*. Oxford University Press.

Kennedy, C. (2007). Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy*, *30*(1), 1–45.

Klein, E. (1980). A Semantics for Positive and Comparative Adjectives. *Linguistics and Philosophy*, *4*(1), 1–45.

Kölbel, M. (2003). Faultless Disagreement. *Proceedings of the Aristotelian Society*, *104*(1), 53-73.

Krantz, D. et al. (1971). *Foundations of Measurement*, Vol. 1. New York: Academic Press.

Kratzer, A. (2012). *Modals and Conditionals: New and revised perspectives*. Oxford: Oxford University Press.

Lasersohn, P. (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy*, *28*(6), 643–686.

Lassiter, D. (2016). *Graded Modality. Qualitative and Quantitative Perspectives*. Oxford University Press.

Lassiter, D., Goodman, N. D. (2013). Context, Scale Structure, and Statistics in the Interpretation of Positive-Form Adjectives. In Todd Snider (ed.), *Semantics & Linguistic Theory (SALT)* 23 (pp. 587–610). Ithaca: CLC Publications.

Lassiter, D., and Goodman, N. D. (2015). Adjectival Vagueness in a Bayesian Model of Tnterpretation. *Synthese*, 1–36.

Liao, S.-Y., Meskin, A. (2015). Aesthetic Adjectives: Experimental Semantics and Context-Sensitivity. *Philosophy and Phenomenological Research*, *94*(2), 371–398.

Liao, S.-Y., McNally, L., Meskin, A. (2016). Aesthetic Adjectives Lack Uniform Behavior. *Inquiry*, *59*(6), 618–631.

Luce, R. D., Narens, L. (1987). Measurement Scales on the Continuum. *Science*, *236*: 1527–1532.

MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.

Meier, C., van Wijnbergen-Huitink, J. (Eds.) (2016). *Subjective Meaning: Alternatives to Relativism*. Berlin: De Gruyter.

Sassoon, G. W. (2016). Multidimensionality in the Grammar of Gradability, MS.
Stephenson, T. (2007). Judge Dependence, Epistemic Modals, and Predicates of Personal Taste. *Linguistics and Philosophy*, *30*(4), 487–525.
Stojanovic, I. (2007). Talking About Taste: Disagreement, Implicit Arguments, and Relative Truth. *Linguistics and Philosophy*, *30*(6), 691–706.
Stojanovic, I. (2016). Expressing Aesthetic Judgments in Context. *Inquiry*, 1–23.
Temkin, L. (2012). *Re-thinking the Good*. Oxford: Oxford University Press.
Umbach, C. (2016). Evaluative Propositions and Subjective Judgments. In C. Meier and J. van Wijnbergen-Huitink (Eds.), *Subjective Meaning: Alternatives to Relativism* (pp. 127–168). Berlin: De Gruyter.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 179–199
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.09
Rozprawa

Jakub Gomułka*, Jan Wawrzyniak**

THE LIAR PARADOX FROM THE WITTGENSTEINIAN PERSPECTIVE

Summary: Our approach to the liar paradox is based on the Wittgensteinian approach to semantic and logical paradoxes. The main aim of this article is to point out that the liar sentence is only seemingly intelligible, and that it has not been given any sense. First, we will present the traditional solutions of the paradox, especially those which we call modificational. Then we will determine what the defects of these solutions are. Our main objection is that the modificational approaches assume that we can express in languages certain senses which are improper. Next, we will explain why we think that the liar sentence is a mere nonsense. This sentence does not have any role in any language game – it is completely useless. We will also respond to several objections to our approach. 1. That it is not consistent with the principle of compositionality of sense. 2. According to the Quineian philosophy of logic, paradoxical sentences can be conceived as false assumptions leading to crises of logical paradigms. 3. The liar sentence seems to be, contrary to our approach, intelligible.

Keywords: compositionality principle, context principle, intelligibility, liar paradox, nonsense, Wittgenstein.

* Pedagogical University, Cracow. E-mail: jakub.gomulka@up.krakow.pl
** Pedagogical University, Cracow. E-mail: jan.wawrzyniak@up.krakow.pl

THE PRESENTATION OF THE LIAR PARADOX

We think that the Wittgensteinian approach to the vast family of semantic and logical paradoxes may turn out to be fruitful. However, we will focus only on the liar paradox here. We will point out that, contrary to a certain well established opinion assumed by many approaches, the sentence which is assumed to generate this paradox, is only seemingly intelligible, and that it has not been given any sense (our conception of nonsense will be presented in the section "The liar sentence as mere nonsense"). We will determine what the defects of traditional solutions are, especially those which we call modificational – later we will explain what we mean by this term.

The liar paradox was formulated by Eubulides in the antiquity. The sentence generating the paradox is: "I am lying". The most famous contemporary exposition of the liar paradox was formulated by Alfred Tarski:

> To obtain this antinomy in a perspicuous form, consider the following sentence:
>
> The sentence printed in this paper on p. 347, 1. 31, is not true.
>
> For brevity we shall replace the sentence just stated by the letter 's.' According to our convention concerning the adequate usage of the term "true," we assert the following equivalence of the form (T):
>
> (1) 's' is true if, and only if, the sentence printed in this paper on p. 347, 1. 31, is not true.
>
> On the other hand, keeping in mind the meaning of the symbol 's,' we establish empirically the following fact:
>
> (2) 's' is identical with the sentence printed in this paper on p. 347, 1. 31.
>
> Now, by a familiar law from the theory of identity (Leibniz's law), it follows from (2) that we may replace in (1) the expression "the sentence printed in this paper on p. 347,1. 31" by the symbol "'s.'" We thus obtain what follows:
>
> (3) 's' is true if, and only if, 's' is not true.
>
> In this way we have arrived at an obvious contradiction. (Tarski 1944, p. 347)

Tarski's presentation of the paradox, as well as the formulations (not solutions) given by Graham Priest (1987), Bertrand Russell (1908), Wolfgang Stegmüller (1955), and Bas C. van Fraassen (1968),

presuppose that the liar sentence is intelligible and has a certain truth-value.

THE MODIFICATIONAL APPROACHES

Today, there is a prevailing opinion that the solution of the liar paradox (and other paradoxes) must consist in imposing certain constraints on language. The aim is to make the formulation of certain troublesome expressions impossible, or to block certain troublesome inferences. So, it can be said that the adherents of such solutions take a normative perspective, not a descriptive one – instead of describing how language works, they want to make its rules "stricter" or "better ordered". Their perspective can be then called regulative or modificational.

Zbigniew Tworak, a Polish scholar specializing in the problem of paradoxes, noticed that there are generally three different kinds of such modification or regulation of language (Tworak 2004, p. 126–128). The first of them concerns formation rules that determine how to form correct linguistic expressions. This modificational approach springs out of the belief that the source of our problems lies in the usual grammar which is far too liberal: it permits certain troublesome strings of symbols to become sentences of language. So, the proponents of this kind of solution suggest modifications to the formation rules and they also tend to disregard the whole natural language and postulate replacing it with a more precise "scientific" or "formal" language. One of the important contemporary exponents of this standpoint is Jan Woleński. He is a follower of the tradition commenced by Russell, Tarski, and Stanisław Leśniewski. The following quote from an article written by that last logician may serve as an exemplification of the view:

> Since, keeping to "natural intuitions" of language we get involved in irresolvable paradoxes, these "intuitions" seem to imply contradiction. The "*artificial*" frame of strict conventions is thus a far better instrument of reason than the language dissolving in the opaque contours of "*natural*" habits which often imply incurable contradictions – much as the "artificially" regulated Panama Canal is a better waterway than the "*natural*" rapids on the Dnieper. (Leśniewski 1991, p. 82)

According to the second form of the modificational approach, paradoxes are simply wrongly interpreted proofs of falsehood of certain commonly accepted non-logical principles. The proponents of this form of solution consider our troublesome formulas or utterances as a demonstration of the incorrectness of certain assumptions or definitions existing in both natural and artificial languages. This approach was presented among others by Leon Gumański (1990) and Stegmüller (1955). The former wrote about it as follows:

> Each antinomy which is not only apparent can be easily transformed into a valid proof by contradiction of a thesis saying that on the basis of a given theory or a given set of assumptions a term defined within the antinomy or used as if it was defined within it in a certain way does not refer to anything (its referent does not "exist"). (Gumański 1990, p. 270)

The third kind of modificational approach aims at a revision of received logical principles. It assumes that logic is not substantially different from other sciences and that it should be subject to improvement, just like physics and chemistry. According to the proponents of this standpoint, since there are many different systems of logic, we are free to decide which of these systems should be recognised as the right one. The most famous logician who held such a view on logic, Willard Van Orman Quine, wrote:

> Logic is in principle no less open to revision than quantum mechanics or the theory of relativity. The goal is, in each, a world system – in Newton's phrase – that is as smooth and simple as may be and that nicely accommodates observations around the edges. If revisions are seldom proposed that cut so deep as to touch logic, there is a clear enough reason for that: the maxim of minimum mutilation. (Quine 1986, p. 100)[1]

Interestingly, Quine himself admitted that revising logic in order to deal with paradoxes (e.g. by recognition of three-valued logic as the right one) "is not to my liking" (Quine 1986, p. 85). So, strictly speaking, Quine does not belong to the proponents of this kind of solution to the problem of paradoxes, although such an option is available within his general approach to logic. Among the thinkers who actually made use of such an option were Dmitri A. Bochvar (1938), Georg Henrik von Wright (1988), and Priest (1987). So, from

[1] At the end of this paper we present some arguments against Quine's holism.

this point of view, paradoxes may be considered as *sui generis* falsifiers of logical systems.

It is easy to notice that each of these three options focuses on a different aspect of what Ludwig Wittgenstein called "grammar". The author of *Philosophical Investigations* adopted this concept in his so-called middle period, when he gradually gave up the views presented earlier in the *Tractatus Logico-Philosophicus*. Initially he started using the term as a synonym of "logic in a broader sense" – it was to include rules which were to warrant non-classical, but still a priori, inferences. Further evolution of the concept of grammar was related to the idea of language as calculus, the idea of the arbitrariness of rules, and finally, the idea of language-games. Wittgenstein developed these ideas on the basis of the philosophy of mathematics proposed by the so-called older formalists, that is Eduard Heine and Carl J. Thomae, and refined – for critical purposes – by Gottlob Frege. According to Wittgenstein's conception, even in the initial language-as-calculus stage, grammar included both the rules governing internal structures of sentences and the rules responsible for inter-sentential relations (Gomułka 2016, p. 220–221, 254).

The pragmatic turn, taken by Wittgenstein in the middle of the 1930s, and thanks to which he arrived at the view known from his *Philosophical Investigations*, brought about the rejection of understanding of the grammar of a natural language as a set of strict rules. Still, grammar covered both kinds of relations: within sentences and among sentences. Anyway, the difference between these two groups of relations became less important after the pragmatic turn, because sentences appeared to be parts of greater wholes, namely language-games.[2]

Let us notice that the modificational approaches assume that the liar sentence is intelligible, i.e. it is a meaningful expression. If instead of being intelligible it were devoid of sense, we could not assign any truth-value to it. If one could not do this, one would not make the

[2] A good illustration of this approach can be found in §§19–20 of *Philosophical Investigations*, where Wittgenstein points out that the same expression can often be considered both as a sentence and as a single word. It should be noted on this occasion that the notion of "logic" appears in *Philosophical Investigations* both in the context of the truth-value calculus (see Wittgenstein 1974, §554) and as a synonym of Wittgensteinian grammar (see Wittgenstein 1974, §345).

inference ending with the formula "'s' is true iff 's' is not true," for the first premise of the reasoning would be senseless – a sentence assigning a truth-value to a senseless sentence lacks sense too.[3] The accounts of the paradox just described assume that either the liar sentence is incorrect or some logical or extra-logical principles which are seemingly obvious are false, for one infers from these principles and the assumption of the liar sentence having a truth-value the following contradiction: "*p* is true iff *p* is not true."

We would like to add that the reverse relation of entailment between the thesis of the intelligibility of the liar sentence and the modificational approach to the paradox does not hold. That is, although the assumption that the liar sentence is intelligible usually leads one to embrace the modificational approach to this paradox, the acceptance of this assumption does not have to lead to embracing this approach, i.e. the thesis that we must modify one or more principles belonging to the three aforementioned kinds. For example, according to Laurence

[3] It seems that a sentence ascribing truth to a senseless sentence should be false. However, if one assumes the meaningfulness of the following sentence "the sentence "colorless green ideas sleep furiously" is true" and accepts both classical logic and Convention T, one can infer the sentence "colorless green ideas sleep furiously", which is obviously senseless. This is unacceptable because one cannot infer a nonsense from meaningful sentences. So, it seems, one must either modify logic or abandon Convention T if one wants to claim that a sentence ascribing truth to a senseless sentence is false. If one modifies logic and retains Convention T, both sentences "the sentence "colorless green ideas sleep furiously" is false," "the sentence "colorless green ideas sleep furiously" is true" entail sentences which were assumed for the sake of the argument to be nonsensical, namely "colorless green ideas do not sleep furiously," "colorless green ideas sleep furiously." There are two objections to this proposal. Firstly, the principles of classical logic are much more intuitive than the "intuition" that a sentence ascribing a truth-value to a senseless sentence should be false. Secondly, as we have said, the idea that one can infer a nonsense from meaningful sentences is unacceptable. If one does not modify logic, but abandons Convention T, then one of the sentences "the sentence "colorless green ideas sleep furiously" is false," "the sentence "colorless green ideas sleep furiously" is true" must be true and the other one false. In this case, however, the words "true" and "false" must mean something other than in the case of applying them to meaningful sentences – if one predicates truth of a meaningful sentence, one asserts the same thing which is asserted by this sentence. So, one can say that "the sentence "colorless green ideas sleep furiously" is true" is false, but this statement does not assert that it is not the case that colorless green ideas sleep furiously – perhaps it can be understood as asserting that one does not ascribe truth to nonsensical strings of signs.

Goldstein (1988, 2000, 2009), the reasoning the conclusion of which is the sentence "*p* is true iff *p* is not true" shows that the liar sentence is not simply a negation of an atomic sentence but a biconditional of the form "*p* iff ~*p*". So – as it seems – the adherents of the modificational approach must accept one more assumption: the liar sentence is not a syntactic contradiction.

WITTGENSTEINIAN CRITIQUE OF THE MODIFICATIONAL APPROACHES

What are the main defects of the modificational approach? We would like to start our considerations on this question by drawing the reader's attention to the fact that the thesis that natural languages are defective because it is possible to formulate in them sentences which generate paradoxes can be interpreted in various ways. Does this thesis imply that the liar sentence is grammatically correct from the perspective of a natural language, or does it imply that this sentence is a meaningful expression of a natural language? These two questions should not be equated. In our opinion, this sentence is constructed according to the rules of school-grammar, but has no sense, because no meaning has been given to some parts of it. If our account of the role of this sentence were right, then natural languages would be no more defective than uninterpreted formal languages. So, as it seems, the defectiveness of natural languages must consist in something else according to the adherents of the modificational approach. Natural languages are defective because one can form in them sentences which have an i m p r o p e r sense. Considerations on the ambiguity of the thesis that natural languages are defective lead us to the uncovering of a hidden assumption lying at the bottom of the modificational approach: we can express in languages (e.g. in natural languages) certain senses which are improper. In our opinion this assumption is wrong, and it is our main objection to the modificational approach.[4]

[4] It is worth adding that we are not concerned here with the question "what features of natural languages allow us to formulate the liar paradox according to the adherents of the modificational approach to this paradox?" – but with the question "what does it mean when the sense of a sentence which generates a paradox is improper?" That is, we are not dealing with, e.g., the question whether the application of semantic terms of a given language to expressions

What could an improper sense be? One may try to answer this question in a few ways. According to one proposal, a sentence has an improper sense if it is formed out of expressions that do not fit together either semantically or syntactically – one will obtain such an improper sense if one concatenates the predicate of a language *L* "is false" with the name of a sentence belonging to this language (if the name of a sentence is "*s*" and the sentence designated by it is "*s* is false", we will obtain the liar sentence). An improper sense understood in this way is, in fact, a semantic or syntactic nonsense. In our opinion, the conceptions of semantic and syntactic nonsense are incoherent, not the very notion of nonsense. According to them, a sentence can be devoid of sense because the result of the concatenation of given expressions yields, semantically or syntactically, an incoherent whole. Such a whole consists of expressions which do not fit together, so, focusing on the conception of semantic nonsense, the whole is devoid of sense because the referents of its constituents cannot be connected in such a way as the expressions are connected in this nonsensical sentence (Diamond 1991). Let us consider, as an example, the following sentence "Julius Caesar is a prime number." This sentence is nonsensical, because Julius Caesar c a n n o t be a member of any set of numbers. The above explanation of nonsensicality of this sentence shows that according to the conception of semantic nonsense this sentence is a nonsense, because it represents (or quasi-represents) something impossible, namely the fact that Julius Caesar is a prime number. So, the conception of semantic nonsense treats some sentences as devoid of sense in virtue of their having a certain s p e c i f i c sense (Whitherspoon 2000). This shows that the conception of semantic nonsense is incoherent.[5] One can

of this language is responsible for the formulation of the paradox, but with the question of "what does it mean when a sentence, in which a semantic term of a given language is predicated of a certain expression of this language, has an improper sense or is ill formed?"

[5] As a response to the remark of the anonymous reviewer that we claim that "a nonsense lacks any sense, but some nonsenses seem to contain some sense, i.e. they express some impossible state of affairs" we would like to state that we accept the first statement and reject the second. We reject the conception of semantic nonsense. The argument presented above shows that it boils down to the incoherent "idea" that some nonsenses have sense, i.e. that they express some impossible states of affairs. Of course, the question whether a given sentence is nonsense or not can be controversial, but if one acknowledges that a given sentence is

draw an analogical conclusion as regards the conception of syntactic nonsense.

According to another proposal, a sentence has an improper sense, if on the one hand it can be treated as a certain kind of self-contradictory sentence, but on the other hand it seems to be a negation of a simple sentence which is unproblematic from the point of view of school grammar. For example, the liar sentence seems to be a negation of the following simple sentence "this sentence is true" and it can be interpreted as a contradiction of the form "*p* is false iff *p* is true" (one may say that the reasoning known as the liar paradox shows that it has such a form). The following sentence "this expression is not a sentence" is a slightly different example of the phenomenon under consideration, as it is of course a negation of a simple sentence, and can also be treated as somewhat self-contradictory – it is a case of a broadly understood pragmatic contradiction.[6] The second explanation of the idea of improper sense is also not satisfying. The very observation that certain negations of simple sentences can be grammatically correct and at the same time seem to be contradictory does not show that they have an improper sense. It shows, at most, that natural languages differ in this respect from standard formal languages. We think that instead of describing such sentences as having an improper sense, one can treat them either as real contradictions or as strings of signs devoid of meaning. The decision depends on the particular example and the context.

One can also criticize numerous modificational solutions of the liar paradox in the following way. All these solutions, which on the one hand claim that the liar sentence generates a contradiction and on the other hand postulate to exclude it from a language (Tarski's

---

nonsense then one does not ascribe to it any sense. Moreover, our approach does not have to assume that in every case it should be decidable without any doubt whether a given expression is meaningful or nonsensical. So, one of us is inclined to recognize "2 + 2 = 7" as a mere nonsense, the other is more hesitant as regards this question, someone may treat it as a meaningful expression, but all these facts do not undermine our position because we claim that if a certain sentence is nonsense, it does not have any semantic features and it is not a premise or a conclusion of any inference.

[6] We are inclined to treat the above sentences as examples of a degenerated kind of sentence. One uses them only in order to give examples of sentences which are false (or true) merely in virtue of the fact that they have been formulated.

solution may serve as a paradigm here), are not consistent (Wawrzyniak 2011). As we have mentioned before, the formulation of the liar paradox requires acknowledging that the liar sentence has a truth-value and, consequently, that it has a sense. Therefore, all solutions according to which the liar sentence is a nonsense cannot be reconciled with the conviction that an ascription of any truth-value to the liar sentence leads to a contradiction. Thus, to be consistent, one must either acknowledge that there is no liar paradox and there is only the liar sentence, which is devoid of sense, or present another solution of the paradox.[7]

Of course, Priest's solution of the paradox is not vulnerable to this objection. But we think that the costs of his solution are too high. It is a desperate move to claim that some sentences – the liar sentence, among others – are both true and false. We think that any consistent solution of the paradox is better than Priest's solution.

THE LIAR SENTENCE AS MERE NONSENSE

Our account of paradoxes is inspired by Wittgenstein's philosophy, especially by his approach to nonsense. We think that the adherents of the so-called "New Wittgenstein" line of interpretation have presented an adequate construal of Wittgenstein's conception of nonsense. Moreover, we think that this conception of nonsense is philosophically correct. We will present the short characteristic of such understood nonsense below. We will not give any substantial definition of nonsense for it would require a substantial definition of meaning. We are inclined to the view that it is impossible to formulate the so-called full-blooded theory of meaning (the question was discussed by Dummett 1987, McDowell 1998, Wawrzyniak 2015). This does not imply that one cannot say anything in general about meaning. We think that,

[7] The anonymous reviewer noted that such a description of Russell's and Tarski's approaches to the paradox is unjust because according to their solutions the liar sentence is devoid of meaning. We point out that if their approach were simple and unambiguous they would not claim that one can infer from the supposition that the sentence is true and the supposition that it is false a contradiction because in order to obtain this conclusion one must assume that the sentence is meaningful. (The following step of the derivation: "(1) 's' is true if, and only if, the sentence printed in this paper on p. 347, l. 31, is not true" would be nonsense if the liar sentence were nonsense.)

among others, an explanation of meaning of any expression consists in describing its use. However, this very general statement does not determine which aspects of use are essential to meaning. Thus, this statement allows one only to say that an expression is nonsense if it does not have any linguistic use. It should be added that the above remark does not exclude the possibility that a piece of nonsense can be used to achieve various purposes, e.g., to induce bafflement, it excludes only that a piece of nonsense is a linguistic expression of something, e.g., an expression of bafflement.

We believe that the liar sentence is mere nonsense.[8] That is to say, it is devoid of sense because no meaning has been ascribed to its parts: "The proposition is nonsensical because we have failed to make an arbitrary determination, and not because the symbol, in itself, would be illegitimate" (Wittgenstein 1961, 5.473). (The development of this idea can be found in: Diamond 1991 and Conant 2000.) Thus, we accept the following truisms about nonsense: an expression E is a nonsense iff it does not have any sense; if E is an assertion, a question, or an order or… (that is, it has not only the grammatical form of an assertion, but is an assertion, and so on), then E is not a nonsense. Because of the fact that it is not possible – as we think – to give a complete list of kinds of illocutionary acts one cannot present a full explanation of nonsense in terms of these acts.

So, there is no liar paradox, there is only a sequence of nonsensical sentences. Of course, such an approach may induce bafflement and even outright opposition: after all, we understand the liar sentence, after all, it is correctly constructed from the point of view of school-grammar, after all, the conclusion that the liar sentence is true if and only if it is false follows from the premises. Thus, we will present arguments supporting our approach to the liar paradox and consider the objections against it.

[8] Of course, this thesis is not a new one. However, our approach to the liar sentence differs from the majority of other approaches which also recognise it as a nonsense in this respect that it treats this sentence – according to the Wittgensteinian conception – as mere nonsense. (It is worth adding that the inspiration to treat the liar sentence as nonsense may have come from another aspect of Wittgenstein's thought, namely from his conception of truth which is quite often interpreted as deflationary. J Beall has pointed out that the deflationists have an independent reason to treat this sentence as a nonsense (Beall 2001).)

As we have mentioned before, we do not deny that a the sentence of the type "*s* is false" (where "*s*" is the name of any given sentence) is well formed. But we deny that the result of a grammatically correct connection of words, which have determinate meanings in other sentences, must be a meaningful whole. We deny this because we accept Frege's context principle (Frege 1959, p. x). Since the meaning of a word should be considered only in the context of the meaning of a sentence, words which are not constituents of meaningful sentences have no meaning. In order to explain in more detail our conception of nonsense, consider the following sentences: "all dogs bark loudly," "all dogs are prime numbers." According to our conception these two sentences contain the same word understood as an inscription, namely "dogs." However, they do not contain the same word understood as a meaningful expression (a logical part of a sentence).[9] Why? Because in the second case there is no context of a meaningful sentence which can determine the meaning of the word "dogs".[10] It seems that our conception is also supported by the fact that one can infer from the first sentence the sentence "all dogs bark," but one cannot infer from the second sentence "all dogs are numbers" because nonsense does not entail anything. Of course, one can transform the sentence "all dogs are prime numbers" into the sentence "all dogs are numbers," but this transformation would not be logical, but graphic. It is worth adding that the view according to which nonsensical expressions can be composed of meaningful parts, assumes the truth of a conception of either semantic or syntactic nonsense. So if our critique of these conceptions is right, there will be no obstacle to acknowledging that nonsensical expressions are not composed of meaningful parts. So, coming back to the liar sentence, we can conclude that if it is devoid of sense, then the expressions contained in it are devoid of meaning too.

Such an approach to nonsense and to the role of the liar sentence in a language may raise two important questions. How can one prove

[9] It is worth noticing that according to Peter Geach inscriptions which have the same shape do not have to be the same word (Geach 1971, p. 86–87), Blackburn does not agree with this view (Blackburn 1984, p. 18–26).

[10] The Fregean context principle was initially framed in terms of *Bedeutung*, but it also applies to *Sinn* – the distinction between *Bedeutung* and *Sinn* was made after the formulation of the context principle.

that the liar sentence is nonsensical? Is this approach to sense and nonsense consistent with the principle of compositionality of sense? As regards the first question, it should be underlined that, according to our conception, it is not possible to prove that any sentence, including the liar sentence, is in its essence a nonsense. (According to the Wittgensteinian view we accept, the very concept of an essence is just a shadow of the grammar of our language and the latter is arbitrary.) At most, one can show that no such sense can be given to the liar sentence that would be recognized by the adherents of the thesis of the intelligibility of the liar sentence as the intended sense. Why do we think that the liar sentence is a nonsense? First of all, this sentence does not have any role in any language game – it is completely useless:

> If the question is whether this is a statement at all, I reply: You may say that it's not a statement. Or you may say it *is* a statement, but a useless one. (Wittgenstein 1976, p. 209)

> Is there harm in the contradiction that arises when someone says: "I am lying. – So I am not lying. – So I am lying. – etc."? I mean: does it make our language less usable if in this case, according to the ordinary rules, a proposition yields its contradictory, and vice versa? – the proposition itself is unusable, and these inferences equally; but why should they not be made? – It is a profitless performance! – It is a language-game with some similarity to the game of thumb-catching. (Wittgenstein 1978, I, App. III, 12, p. 120)

Moreover, considerations on the liar sentence from other influential perspectives, like the Davidsonian and the Dummettian, reveal that the sentence has neither truth nor assertability conditions. So, it also turns out to be useless and hence senseless.

Secondly, one can notice that in constructions such as:

L: The sentence L is false.

the sign "L" does not refer to any determinate thing. It seems that it is to refer to a sentence. But to what sentence? The adherents of the thesis that the liar sentence has a sense will claim that the sign refers to "The sentence L is false." Of course, this sign can designate the uninterpreted inscription "The sentence L is false." This, however, does not allow one to defend the thesis that the liar sentence has a sense. What could the words mean: "a certain uninterpreted string of signs is false"?

So, the sign "L" should refer to a certain meaningful sentence. The formula "The sentence L is false" will acquire a sense, only if the sign "L" in this formula designates some a meaningful sentence, e.g. "All cats mew", but the adherents of the thesis of the intelligibility of the liar sentence do not accept any interpretation of this kind and claim that the sign "L" within the formula designates the sentence "The sentence L is false." So, as one can see, the attempt to ascribe a sense to the liar sentence ends with a failure – the result is either a meaningful sentence which is not recognized as the real liar sentence or *regressus ad infinitum*.[11]

Let us return to the second question: Is this approach to sense and nonsense consistent with the principle of compositionality of sense? Some authors point out that the principle of compositionality can be and is understood in various ways (Bronzo 2011, Pelletier 1994, Peregrin 2005). Our conception of nonsense is certainly not consistent with the conviction that the meaning of words is prior to the sense of sentences. To use a vivid picture, sentences are not like houses which are built of earlier existing elements (say, bricks). We think that the understanding of the principle of compositionality on the model of the relation between houses and their elements is only a possible, but not convincing, interpretation of this principle. According to the suggestions of Bronzo (2011) and Peregrin (2005), we assume that the principle of compositionality boils down to acknowledging that if a meaningful sentence is given, then its sense is a function of meanings of its parts and the mode of their combination. The principle does not entail that the meanings of words must be (conceptually or temporarily) prior to the senses of sentences and that they must be explained outside of the context of sentences. It is worth noticing that this principle does not say that the nonsensicality of a sentence is a function of meanings of its parts and the mode of their combination.

[11] This argument is similar to the line of thought presented by Alfred Gawroński (Gawroński 2004, see also Gawroński 2011, p. 109–156). It can be said that the argument shows that the quotation operator is not a logical operator (in the standard sense of logic). Therefore, there cannot be a function relating expressions in a language to expressions in its meta-language. We owe this remark to Wojciech Krysztofiak – our second reviewer. We are also thankful for his other inspiring comments. We did not address them here though – this would require substantial extensions of our present article. Indeed they are worth a separate paper.

The principle of compositionality entails, at most, that any expression which has a certain meaning in a sentence may have the same meaning in some other sentences.[12] To simplify the matter a little, the principle does not say that someone who learns the full content of a comprehensive dictionary of a foreign language and a comprehensive grammar textbook of the same language will understand every sentence of this language, but rather it says that someone who fully understands a language will be able to indicate how the meanings of expressions and grammatical constructions determine the senses of sentences of this language. We would like to end this part of the text with a brief comparison between our approach to the liar paradox and Goldstein's approach to this paradox which is also inspired by Wittgenstein's philosophy. The main difference concerns the meaningfulness of the liar sentence. According to Goldstein, this sentence has a sense, but it cannot be used to make a statement (Goldstein 2009, p. 382). We think that we can acknowledge at most that the liar sentence is correct from the point of view of the ordinary grammar, but the same can be said about the sentence "colorless green ideas sleep furiously." Goldstein's argument for the thesis that it is meaningful is that that it can be "translated". We think that this argument is unconvincing because Chomsky's above-cited example can also be "translated", although it is obvious that this sentence is nonsensical.

QUINEIAN HOLISM VS. WITTGENSTEINIAN HOLISM

One may put forward the following argument which seems to refute our standpoint. If we take the Quineian understanding of logic, then there is no reason to give a special status to its propositions. Quine – followed by other thinkers including Richard Rorty – thought that the division between "empirical" and "analytical" judgements is purely dogmatic and mystifies the actual function of the propositions of logic in our system of knowledge. It is true that these propositions take central position in the structure of knowledge, but – according to Quine – this does not mean that they are invariants in the perpetual

[12] The practical learnability of a language requires that words and expressions should have the same meaning in a great majority of sentences in which they appear.

process of the self-revision of science. For logic also belongs to the body of the famous Neurath's boat (maybe it is the frame of the boat) and, therefore, it also undergoes the processes of partial reconstruction. If so, one can apply Kuhn's theory of crises of paradigms and their overcoming to logic as well, and therefore consider paradoxes and antinomies as symptoms of crises of logical paradigms.[13]

In response to this charge we have to answer that the Quineian view of science, and his understanding of the role of logic in particular, is simply wrong. We reject Quine's methodological holism, for it leads to the obliteration of the difference between what is false and what is nonsensical. But the difference is indispensable if we want to be able to speak of counterfactual possibilities at all. It may be blurred: presumably there are sentences about which we are not sure whether their negations are still understandable. So, even if logic (or grammar in the sense of the later Wittgenstein) changes due to the pressure of experience (or something else), the change does not go according to Quineian terms – it does not consist in the rejection of a certain theory. Grammar fixes the way in which we understand our theories, theorems, and reporting sentences, so its change cannot be conceived as a rational revision undertaken for some reasons, some arguments. Contrary to certain opinions, we are not willing to accept that some fundamental physical laws are quasi-logical. For example, we tend to think of the sentence "the principle of conservation is false" as understandable, because we may imagine a world, in which mass and energy emerge spontaneously. Indeed, the Steady State cosmological theory formulated in 1948 – now considered false by the majority of physicists – assumed such spontaneous emergence of particles.

Moreover, it is hard to say that the paradoxes and antinomies in logic emerge because of transformations of our knowledge about the facts. The liar paradox was presented for the first time in antiquity, when people's world-picture and science were vastly different from the present ones. It seems that the change of our empirical knowledge had no influence on the role of the paradox. Also, it is not above our intellectual abilities to imagine a scenario in which the ancient thinkers developed naïve set theory and formed the known set-theoretical

[13] Again, we must underline that Quine himself was not a proponent of such a solution to the problem of paradoxes.

antinomies including Russell's paradox, while their knowledge of natural facts remained at its actual level.

The later Wittgenstein is considered a holistic thinker. However, his holism is essentially different from the Quineian overall philosophy of science. The point of dispute is the principle of arbitrariness of syntax adapted by the author of *Philosophical Investigations* no later than at the beginning of the thirties (Gomułka 2016, p. 196). The principle assumes the difference between factual (empirical) and grammatical (conceptual) questions and propositions. However, in remarks published as *On Certainty* and written at the very end of his life, he seemed to realize that some empirical statements can "harden" up to the point in which they may function as "channels" for other, more "fluid" empirical propositions, but at the same time the distinction between a channel and something which flows through it remained:

> The mythology may change back into a state of flux, the river-bed of thoughts may shift. But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other. (Wittgenstein 1972, §97)

The author of *Philosophical Investigations* explicitly rejected the suggestion that logic is an empirical science (Wittgenstein 1972, §98). Wittgenstein's positive view can be illustrated by a fictional story about the king of a certain tribe brought up to the conviction that the world began with his own birth. The Austrian thinker pointed out that a possible change of the king's conviction cannot be understood as a usual shift of beliefs on a topic due to some rational arguments, but as something much more fundamental, that is a conversion (*Bekehrung*), for it would require the transition to a wholly new way of perceiving the world (Wittgenstein 1972, §92).[14]

[14] Nowadays it is pointed out that *On Certainty* belongs to the so-called third period of Wittgenstein's thought and is written in a somewhat different spirit from *Philosophical Investigations* (Moyal-Scharrock 2004). According to us, possible differences between these two works are rather irrelevant to the core of the problem illustrated by the tale of the king.

THE FINAL CHARGE

All our previous arguments can be rejected simply by referring to the fact that we seem to understand the paradoxical expression. How can we say that no one understands the so-called liar sentence when so many claim that they understand it very well. Does the fact that one has a sense of understanding of a sentence one utters not decide the question whether this sentence has a sense? Is not the question of meaningfulness of the liar sentence settled for good in this way?

No, it is not. The conviction according to which the understanding of a sentence or a word is like a pain – it just emerges for the consciousness – results from the mythical picture of the mind created by Rene Descartes. This mythical picture tells us that thoughts and sensations make up the two kinds of *cogitationes* – contents of consciousness – whose existence is beyond doubt. Accordingly, the sense of a sentence is to be roughly the same kind of "object" as pain, and a subject having a mind within which this "object" emerges is to have privileged and direct access to it. But the self-transparency of mind and its absolute sovereignty regarding its own content is not compatible with our scientific knowledge or even our common experience. For it happens many times that we only seem to understand something, we often admit that we do not know what we have just meant, and on the other hand we sometimes say something completely meaningful without any conscious or phenomenal "underlay". All these observations prompted Wittgenstein to formulate a standpoint that the criteria of understanding cannot be internal: whether one understands a rule or not depends on one's ability to use it (Wittgenstein 1974, §§146–155); for this very reason, as the author of *Philosophical Investigations* writes: "An «inner process» stands in need of outward criteria" (Wittgenstein 1972, §580).[15]

Wittgenstein's standpoint can be seen as a deepened Fregean anti-psychologism. As Frege underlined, we should not understand meanings of words as a kind of internal content available only to a mind that thinks or says these words. It should be noted that the philosopher from Jena used this conviction to justify his context principle, for he pointed out that if one does not observe the latter,

[15] Attacks on the Cartesian myth have been carried out by many other acknowledged thinkers, like Gilbert Ryle (2009) and Richard Rorty (1983).

"one is almost forced to take as the meanings of words mental pictures or acts of the individual mind," and, hereby, ends up in psychologism (Frege 1959, p. x).

So, if one can be wrong about the understanding of sentences, one can be equally wrong about whether a linguistic expression is understandable at all, and thus whether it is a meaningful sentence. When we apply pragmatic criteria we come to the conclusion that the so-called liar sentence cannot have any sense, for it has no role to play in our linguistic practice.

REFERENCES

Beall, J. C. (2001). A Neglected Deflationist Approach to the Liar. *Analysis*, *61*(2), 126–129.

Blackburn, S. (1984). *Spreading the Word*. Oxford: Oxford University Press.

Bochvar, D. A. (1938). Ob odnom trechznachnom ischislenii i yego primienienii k analizu paradoksov klassicheskogo passhirennogo funktsionalnogo ischiclenya (On a three-valued logical calculus and its application to the analysis of contradictions). *Matematicheskiy sbornik*, *4*(46):2, 287–308.

Bronzo, S. (2011). Context, Compositionality and Nonsense in Wittgenstein's *Tractatus*. In R. Read and M. Lavery (Eds.), *Beyond the Tractatus Wars* (pp. 84–111). London: Routledge.

Conant, J. (2000). Elucidation and Nonsense in Frege and Early Wittgenstein. In A. Crary and R. Read (Eds.), *The New Wittgenstein* (pp. 174–217). London–New York: Routledge.

Diamond, C. (1991). What Nonsense Might Be. In *The Realistic Spirit – Wittgenstein, Philosophy, and the Mind* (pp. 95–114). Cambridge, Mass–London: The MIT Press.

Dummett, M. (1987). Reply to John McDowell. In *Michael Dummett: Contributions to Philosophy*, B. Taylor (Ed.) (pp. 253–268), Dordrecht: Martinus Nijhoff.

Fraassen, B.C. van (1968). Presupposition, Implication and Self-Reference. *Journal of Philosophy*, *65*, 136–152.

Frege, G. (1959). *Foundations of Arithmetic: A Logical-Mathematical Enquiry Into the Concept of Number*. Transl. J. L. Austin. Evanston (IL): Northwestern University Press.

Gawroński A. (2004). Tzw. «Zdanie kłamcy» jako rekurencyjna funkcja zdaniowa. *Studia Semiotyczne*, *25*, 33–57.

Gawroński A. (2011). *Wizja i argumentacja w filozofii. Od lektury Homera do teorii metatekstu*. Kraków: Znak.

Geach P. (1971). *Mental Acts*: *Their Content and Their Objects*. London–New York: Routledge–Kegan Paul.

Goldstein, L. (1988). Wittgenstein's Late Views on Belief, Paradox and Contradiction. *Philosophical Investigations*, *11*(1), 49–73.

Goldstein, L. (2000). A Unified Solution to Some Paradoxes. *Proceedings of the Aristotelian Society*, *100*, 53–74.

Goldstein, L. (2009). A Consistent Way With Paradox. *Philosophical Studies*, *144*, 377–389.

Gomułka, J. (2016). *Rachunek. Filozofia nauk formalnych i jej związek z koncepcją podmiotu we wczesnym i średnim okresie twórczości Ludwiga Wittgensteina*. Kraków: WN UPJPII.

Gumański, L. (1990). *Wprowadzenie w logikę współczesną*. Warszawa: PWN.

Leśniewski, S. (1991). *The Critique of the Logical Principle of the Excluded Middle*. Transl. S. J. Surma, J. Wójcik. In *Collected Works*, Vol. 1, eds. S. J. Surma, J. T. Srzednicki, D. I. Barnett (pp. 47–85). Warsaw–Dordrecht: PWN, Kluwer.

Moyal-Scharrock, D. (2004). *The Third Wittgenstein: The Post-Investigations Works*. Aldershot: Ashgate.

McDowell J. (1998). Another Plea for Modesty. In *Meaning, Knowledge, and Reality* (pp. 108–131). Cambridge, Mass.: Harvard University Press.

Pelletier, F. J. (1994). The Principle of Semantic Compositionality. *Topoi*, *13*(1), 11–24.

Peregrin, J. (2005). Is Compositionality an Empirical Matter? In M. Werning, E. Machery, G. Schurz (Eds.), *The Compositionality of Meaning and of Content*, *Vol. 1 Foundational Issues* (pp. 231–246). Frankfurt: Ontos Verlag.

Priest, G., (1987). *Unstable Solutions to the Liar Paradox*. In S. J. Bartlett and P. Suber (Eds.), *Self Reference: Reflections and Reflexivity* (pp. 145–175), Dordrecht: Nijhoff.

Quine W. V. O. (1986). *Philosophy of Logic*, 2nd edition. Cambridge (MA), London: Harvard University Press.

Quine, W. V. O. (1980). *From a Logical Point of View*. Cambridge (MA), Harvard University Press.

Rorty, R. (1983). *Philosophy and the Mirror of the Nature*. Oxford: Blackwell.

Russell, B. (1908). Mathematical Logic as Based on the Theory of Types. *American Journal of Mathematics*, *30*, 222–262.

Ryle, G. (2009). *A Concept of Mind*. London–New York: Routledge.

Stegmüller, W. (1955). Die Antinomien und ihre Behandlung. *Insbrucker Beiträge zur Kutlurwissenschaft*, *3*, 27–40.

Tarski, A. (1944) The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, *4*(3), 341–376.

Tworak, Z. (2004). *Kłamstwo kłamcy i zbiór zbiorów. O problemie antynomii*. Poznań: WN UAM.

Wawrzyniak, J. (2011). Paradoks kłamcy. *Analiza i Egzystencja*, *15*, 161–179.

Wawrzyniak, J. (2015). *Znaczenie i wartość w filozofii Johna McDowella i Barry'ego Strouda. Przedmiot refleksji czy redukcji?* Kraków: Księgarnia Akademicka.

Witherspoon, E. (2000). Carnap and Wittgenstein. In A. Crary and R. Read (Eds.), *The New Wittgenstein* (pp. 315–349), London, New York: Routledge.

Wittgenstein, L. (1961) *Tractatus Logico-Philosophicus*. Transl. B. McGuinness and D. Pears. London: Routledge.

Wittgenstein, L. (1976). *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939*. Ed. C. Diamond. Chicago, London: The University of Chicago Press.

Wittgenstein, L. (1978). *Remarks on the Foundations of Mathematics*. Transl. G. E. M. Anscombe, eds. G. H. von Wright, R. Rhees, G. E. M. Anscombe, third Edition. Oxford: Blackwell.

Wittgenstein, L. (1972). *On Certainty*. Transl. D. Paul, G. E. M. Anscombe. New York: Harper Torchbook.

Wittgenstein, L. (1974). *Philosophical Investigations*. Transl. G. E. M. Anscombe. Oxford: Blackwell.

Wright, G.H. von, (1988). Comments and Reflections. In N. Incardona (Ed.), *Les forms actuelles du vrai* (pp. 185–199). Palermo: Enchridion.

Studia Semiotyczne, t. XXXI, nr 2 (2017), s. 201–226
ISSN 0137-6608
doi: 10.26333/sts.xxxi2.10
Rozprawa

Filippo Batisti*

LINGUISTIC RELATIVITY AND ITS RELATION TO ANALYTIC PHILOSOPHY

Summary: The history of so-called 'linguistic relativity' is an odd and multifaceted one. After knowing alternate fortunes and being treated by different academic branches, today there are some new ways of investigating the language-thought-reality problem that (i) put into dialogue the latest trends in language-related disciplines (ii) generate room for philosophical themes previously overlooked, (iii) reassess the very idea of linguistic relativity, despite its popularized versions which have circulated for decades and which have led an otherwise fruitful debate to extremes. It is argued that a multidisciplinary approach is desirable in order to broaden future research. In the last few years the opportunity to study this matter following a common trend in several disciplines has been created. Language, and cognition too, are now conceived as intrinsically social phenomena. It is argued that relativistic effects should be investigated in social realms, and that analytic philosophy could help with this task.

Keywords: linguistic relativity; Sapir-Whorf hypothesis; Philosophy of language; psycholinguistics; extended mind

INTRODUCTION[1]

This paper will address a single line of research within the many ways in which the language-thought relationship has been studied,

* Ca' Foscari University of Venice, Department of Philosophy and Cultural Heritage. E-mail: filippo.batisti@unive.it

[1] I would like to thank the two anonymous reviewers for their helpful comments which undoubtedly contributed to making this paper better. I also thank Maria Angela, Stefano, Roberto, and Benedetta for their precious support.

namely the so-called *linguistic relativity* principle (LR), also known as the 'Sapir-Whorf hypothesis'.[2] Let us define this idea in an expanded fashion:

> Linguistic relativity is the idea in accordance to which speakers of different specific varieties of natural languages, which differ in a number of respects studied by linguistics (such as phonetics, syntax, semantics, and pragmatics), could experience[3] the same objects and activities of the world (such as, but not limited to, physical objects perception, colour perception, space relationships, discourse interaction, calculus, shaping of categories, decision making) in different ways, on the grounds of that very linguistic diversity – and not because of other factors such as explicit cultural elaboration, or cognitive deficiencies or deviations.

More concisely, speakers of two languages that do not have similar linguistic structures in an identified respect could be affected by this asymmetry in the way they think of or experience that respect. As (1) shows, there is a wide variety in the domains supposedly interested by such LR effects.

But how many kinds of linguistic relativity exist? A very common historiographic solution is to pair a 'strong' and a 'weak' hypothesis. The former is used as a synonym for a more transparent 'linguistic determinism', *i.e.*, the conceptual system of a certain language is *incommensurable* to the others. Lakoff, among others, has analysed the "commensurability issue" in its composite meanings, with the result of highlighting that "there are several kinds of commensurability,

[2] This label, albeit popular, should be dispreferred because, as Lee (1996) stated, there is simply no such thing as a "hypothesis", formulated by Sapir or Whorf, let alone jointly. In Whorf's words, LR was a "principle", therefore a "conviction" (Dor 2015, p. 89–90). See also Everett (2013, p. 2), who notes that as contemporary research is in fact rapidly evolving, it is probably pointless to label the "hypothesis" as belonging to one or another scholar. Furthermore, it is worth underscoring that present studies in linguistic relativity are inspired by Whorf's work only in a broad sense. Criticism on Whorf's own positions does not automatically affect present-day researchers' claims, and *vice versa* (*ibid.* p. 22).

[3] Even if the word "experience" surely rings a phenomenological bell, the intent was to cover a vast number of aspects of human life (see *infra*) with one single term. It also takes into account Dor's (2015) complex proposal on considering language as a communication technology that constantly tries to overcome the *experiential gap* between individuals. This framework challenges a lot of mainstream assumptions and has implications for LR studies (see *ibid.*, chapter 5) as well, but for reasons of space it will not be discussed here.

and commentators are by no means clear about which kind is being discussed" (Lakoff 1987, p. 322). Even so, the 'weak' hypothesis – that linguistic structures affect in some non-dramatic, temporary, and reversible way our cognition – is the one that has caught the interest of scholars, especially in cognitive psychology.

However, more recent and in-depth definitions, such as Wolff and Holmes's, seem more useful in order to understand more clearly what we are referring to when we speak of linguistic relativity. LR is defined as a "'family' of related proposals that do not necessarily fall along a single strong-to-weak continuum" (Wolff, Holmes 2011, p. 253). The authors sketch a tree-diagram in which linguistic determinism (the 'strong hypothesis') has a premise that thought is indeed separate from language (*i.e*. language is not language-of-thought, in a Fodorian fashion), but nonetheless thought and language are considered to have parallel structures. From this assumption follows the incommensurability thesis discussed above. Thus, the contrary assumption, namely that thought and language differ structurally, corresponds to the 'weak hypothesis'. However, in Wolff and Holmes's account, this is not sufficient to single out the whole spectrum of specific manners in which language can affect thought: another seven classes and subclasses are individuated by the authors. That is to say, 'weak hypothesis' is too broad a label for scholarly purposes, albeit useful for differentiating that sub-family of hypotheses from the deterministic one. In fact, the *strong* v. *weak* account may have gained ground because it does not force the proponents of the 'weak' one to defend themselves from all the perilous ethical, and epistemological issues connected with the deterministic view.[4]

This paper aims to shed light on the relationship between philosophy and the study of LR in the last two centuries or so, especially analysing the last few years in which the whole branch has gained new vitality in its aims and methods, also – I argue – thanks to analytic philosophy. I mean to do so by a brief overview of the most interesting paths recently taken by scholars.

[4] See Lakoff 1987, p. 304 *ff*.

1.1 THEORETICAL PREMISE

Before engaging in sketching a history of the treatment of this line of work, I wish to explain the criteria upon which the following partition has been organised. As a premise, I need to state that I follow the opinion that language sciences, and philosophy of language as well, should try to treat their object of inquiry not as something abstract from its actual use in everyday contexts. Hypostatizing certain features of the linguistic structure may have the countereffect of making us stray from the ultimate scope of investigating language itself, namely to understand how and why humans *use* it. The concrete patterns of interaction – and of action in solitude as well, even if language arguably originated as a tool of communication (Tartabini 2011; *contra* Humboldt, see Koerner 2000, p. 10) – should be the starting point of an enquiry into its functioning, as well as its arrival point. Certainly, conceptual analysis and theoretical knowledge require some degree of abstraction, but, especially in psycholinguistic research, the output of scholarly elaboration should, eventually, describe the state of affairs without overlooking any of the actual situations in which language is used by (and among) individuals.

2. A BRIEF HISTORY OF LINGUISTIC RELATIVITY

	APPROXIMATE DATE	REPRESENTATIVE SCHOLARS	*RL* EXISTS / RELEVANT	*RL* AFFECTS ACTUAL BEHAVIOUR / ACTION	PHILOSOPHICAL INFLUENCES
1st WAVE	18–19th cent.	Hamann, Herder, Humboldt	Yes	Yes	Romantic Philosophy
	1890s–1950s	Boas, Sapir, Whorf	Yes	Yes	(Theosophism – Whorf[a])
2nd WAVE	1960s–1980s	Berlin, Kay, Rosch, Penn	No	No	/
	1990s–2000s	Lucy, Levinson, Slobin, Boroditsky	(Mostly) Yes	Mixed	/
3rd WAVE	2000s …	Michael, Enfield, Sidnell, Zinken	Yes	Yes	Analytic Philosophy

[a] See Whorf 2012, p. 23–25.

Figure 1.

Figure 1 summarises the criteria through which I look at the history of LR: first, a small number of representative scholars from each wave or sub-wave have been chosen; then it was noted if they supported the existence (and relevance) of LR. Thirdly, it was assessed if their approach was consistent with the idea that RL effects affect speakers in their everyday life and not only in artificial settings. Finally, the broad philosophical influences for each (sub-)wave were indicated. Let us now examine in greater detail each one of them.

2.1 THE 1ST WAVE: THE ORIGINS

The relationship between language and thought, broadly construed, has been a topic of philosophical elaboration since the Presocratics. Even the Bible offers much food for thought in this respect – just think of the myth of Tower of Babel. However, this particular line of study, namely the influence that *each* different language may have on thought sparked at a particular time. According to Dor (2015, p. 87–88), four historical and ideological factors decisively contributed to the outbreak of interest in such an approach:

> First, the rise of the nation-state as a political model, with its romantic ideology of nationalism […] brought along a vested interest in a view of language as both an exact reflection of the national spirit, the *Volksgeist*, and a major determining factor in its construction.

Secondly:

> Europeans, in the course of the project of colonialism, discovered more and more languages around the world that were ostensibly very different from the languages known to them at the time. Travelers, adventurers, and priests began to describe and analyze these languages, and suggest ideas as to the relationships between them and the cultures within which they emerged.

Thirdly:

> *secularization*: the question of linguistic relativity in its modern form could only begin to emerge with the weakening of the conviction that both human language and human thought, whichever way one thinks about them, are the divine creation of God.

And finally:

> Kant's philosophy of mind, was decidedly universalistic – the categories and intuitions are shared by all rational minds – but it immediately opened the door for

a relativistic re-formulation: what if we look at the world through the categorical lenses of our different languages?

Within this ideological environment, 19th century German scientist and intellectual Humboldt was the most eminent voice to offer some in-depth insights into the relationship between natural languages and the way in which one sees the world (*Weltansicht*). Humboldt wrote that "the world in which we live […] is exactly that into which the language we speak transplants us" (Humboldt 1904, p. 332), meaning that every language brings a world-view that, mostly unconsciously, "mirrors" the way in which language categories "construct the world" (see Koerner 2000, p. 10). Again, language is seen as something that strongly mediates the external world and the subject that afterwards gets to perceive it:

> [...] there resides in every language a characteristic *world-view*. As the individual sound stands between man and the object, so the entire language steps in between him and the nature that operates, both inwardly and outwardly, upon him [...] Man lives primarily with objects, indeed, since feeling and acting in him depend on his presentations, he actually does so exclusively, as language presents them to him. (Humboldt 1988, p. 6)

As Koerner has accurately shown, there exists a line of thought that unites German philosophers (Hamann, Herder, and Humboldt) and linguists and anthropologists based in North America (Boas, Sapir, and Whorf). Oddly enough, each one of these scholars had been the teacher of the next in line – or at least the two had been in contact for academic reasons. Sapir was the first, in 1924, to use the term 'relativity' to name the 'linguistic relativity hypothesis' as it was popularized by Whorf's papers, which also took advantage of the analogy with Einstein's theory of relativity[5] in physics.[6]

However, in the historical partition that I am trying to sketch, the German–North American circulation stage of LR still falls in the first of the three waves. This is due to the circumstance that the actual implications of linguistic diversity were described in terms of "action",

[5] This claim is consistent with the one made in footnote 1, as speaking of a "Sapir-Whorf Hypothesis" entails many factors that are not necessarily true, *e.g.*, that the two had the same view on the matter.

[6] See Zinken (2008) for a repertoire of the various metaphors used in the language-thought debate.

"behaviour" and "habits" – which are mostly alien to the second phase of the debate. But first let me clarify what Whorf, as the most prominent representative of the first phase, meant with the aforementioned notions. He wrote:

> [the grammar] of each language is [...] itself a shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock in trade. (Whorf 2012, p. 272)

Further, he adumbrated a definition of

> a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar [...]. (*ibid*., p. 274)

Language helps us "organize" the world as we perceive it, and then, on the basis of this mental organization, we get an already (at some level) elaborated blueprint for making decisions and acting. As Whorf clarified, he did not "wish to imply that language is the sole or even the leading factor in the types of behaviour mentioned [...] but that this is simply a coordinate factor along with others" (Lee 1996, p. 153). The point I want to make clear is that, in Whorf's view, language-driven perception is something that is linked in a causal chain to behaviour, that is, to action.

2.2 THE 2ND WAVE: CHOMSKIANISM AND THE WHORFIAN RENAISSANCE

This last link in the chain had been missing in the LR debate from, *grosso modo*, Whorf's posthumous publications in the 1950s until the last decade. So, phase two began as a consequence of the renovated *milieu* in psycholinguistic research due to the hegemony gained by Chomsky's Universal Grammar theory. Universalist interpretations of the language-thought problem were generally preferred over relativistic ones (Berlin, Kay 1969, Rosch 1972). Meanwhile, experimental cognitive psychology procedures and techniques were improved and fine-tuned, so that perceptual domains, conceptualisation or orienting in space were the dominant themes in LR research. Such a trend had the effect of lowering interest in the cognitive consequences of linguistic diversity, because if it was nothing but a superficial phenomenon and there existed a cognitive unity of mankind, then LR ought

to be false or, in the best case, irrelevant (see Penn 1972, p. I or Pinker 1994, p. 57).

One of the champions of the Chomskian standpoint on the Whorfian hypothesis is former Boston MIT and now Stanford cognitive psychologist Steven Pinker. Pinker is well known not only in academic circles and in his books *The Language Instinct* (1994) and *The Stuff of Thought* (2007) he criticised in neither uncertain nor diplomatic terms Whorf's arguments and the idea of LR in general, which were, according to him, "wrong, all wrong" (Pinker 1994, p. 54). The problem with Pinker's criticism is that, in both books (*i.e.* even after a remarkable thirteen-year interval), he seems to obstinately equate linguistic relativity (as well as Whorf's hypotheses) with what should be properly called linguistic determinism (see *supra*). Linguist and anthropologist Pharao Hansen (2009) has noted that attacking the deterministic version of the issue most surely falls into the so-called straw man fallacy: since there is a consolidated and widespread consensus in psychology about the fact that language is nothing but *one* of the many factors contributing to the formation of thought, consequently linguistic determinism has long been removed from every serious research agenda, due to the untenability of the argument. So, arguing against a thesis that is not actually supported by anyone in academia and, on the other hand, misrepresenting the neo-Whorfian (see *infra*) has little use.

Apart from the specific case of Pinker's production, it remains true that for many years cognitive scientists and linguists have followed the innativist paradigm endorsed by Chomsky. Its non-relativistic basic assumption was that crosslinguistic variation should be treated as a "surface-level" feature. In Levinson's words, works such as Berlin and Kay's (1969) on colours wanted to demonstrate that "universals, or more exactly typological constraints, may lie behind the apparent semantic diversity of languages". The rejection of the relativity argument, then, was rooted in Chomsky's "conception of language as an autonomous formal system" (Dor 2015, p. 90), combined with the (Fodorian) thesis that "language and thought, so conceived, are separate modules, each with its own essence" (*ibid.*). In general, thus, it was assumed that taking seriously data which conveyed linguistic diversity was not as important and meaningful a task as concentrating on retrieving the common deep features that must have associated all

known languages.[7] Therefore, theoretical research in LR had suffered a "decades-long delay" before a number of previous proponents of Universal Grammar "became complete[ly] disenchanted" with such a universalist linguistic paradigm (see Everett 2013, p. 21).

So, in the early 1990s LR received new attention thanks to the seminal work by Lucy (1992b) and Gumperz, Levinson (1996) who represented a different stream in LR research, as they *confirmed* LR effects (e.g. Boroditsky 2001, on time and space, or Imai, Mazuka 2003, on objects and substances; Levinson 2003, on spatial frames of reference; see Casasanto 2008, 2016 and Everett 2013 for an overview).[8]

Slobin's proposal of "thinking for speaking" is an important one, but has a different history, since it refers to "online" effects. More clearly, Slobin (1996) holds that the words of the language we are using in a specific situation influence our cognition *only as long as* we use them, so that their constraints cease to be effective when the speaker stops talking. This kind of effect of language on thought is generally not considered a good representative of the relativity principle, since "offline" influences – *i.e*. when linguistic structures affect cognition even when speakers are *not* engaged in language-related tasks – would be less expected and much more interesting.

This whole movement has been tagged as "Neo-Whorfian" or as a "Whorfian Renaissance", but these names need clarification: virtually all work done under these labels is not strictly related to Whorf's, even though it is obviously inspired by his writings. As Everett (2013, p. 22) puts it, "[neo-Whorfian research] is very non-Whorfian methodologically", as Whorf's program does not meet the present standards in psycholinguistic research, so it is probably safer to use a more neutral label like 'linguistic relativity'.

Multiple perspectives were adopted in relation to a range of ontological domains, though "mostly nonsocial", as linguistic anthropologist Enfield points out:

---

[7] For example, Bloom and Keil (2001) offered an alternative explanation to Lucy's (1992a) empirical data bringing into play the causal role of *culture*, instead of language.

[8] Lakoff's chapter (1987, p. 304–337) on relativism too had helped to re-evaluate Whorf's ideas, paving the way for the "Rethinking Linguistic Relativity" conference held in Jamaica in 1991, which in turn led to the essential volume edited by Gumperz and Levinson (1996).

research that has been done [...] has covered only a thin slice of the possible scope of this topic because Neo-Whorfian work has been fairly consistent in its narrow interpretation of the three key concepts. *Reality* has been taken to mean the realm of objective, nonsocial facts: "concepts of 'time,' 'space,' and 'matter'". *Thought* or *mind* has been taken to mean general, nonsocial cognition: forms of categorization, reasoning, and memory about reality as perceived. And *language* has mostly been taken to refer to structural and semantic features, synchronically framed, with a focus on the referential functions of words […] Restricting the scope in this way has delivered valuable progress. But it is time to consider the larger space of things that could or should be regarded as instances of linguistic relativity. (Enfield 2015, p. 213)

Enfield has indeed good reasons to claim that the majority of research has gone in a certain direction, even if he slightly exaggerates the actual state of affairs. It is true, as pointed out in detail by Björk (2008), that in many experiments artificial settings have been employed and that the methodology is certainly "non-social"; at the same time, a different approach to the problem has not been absent, even since Gumperz, Levinson (1996, part IV where discourse-based approaches are considered). Moreover, volumes like *Grammars of Space*, edited by Levinson and Wilkins (2006), consider the semantic parameters involved in 'Where-' questions: many languages were taken into account and all the studies were based on fieldwork – such a methodology excluded, e.g., laboratory experiments. But, more importantly, its companion volume *Space in Language and Cognition* (Levinson 2003), where the linguistic data meet the study of crosslinguistic cognitive diversity, shows that relativity effects appear in everyday situations (see, for example, p. 216–244).

2.3 THE 3RD WAVE: THE EXPANSION PHASE

The last quotation by Enfield could be a starting point for a new generation of LR researchers as it represents the third phase; I propose to call it the 'expansion phase'. This choice of words is justified by a common trend shared by recent developments in many different disciplines, namely, the *extension* of their object of inquiry. This is happening in branches such as philosophy of mind, the so-called 4E-cognition in psychology, linguistic anthropology, linguistic pragmatics, and conversation analysis.

More precisely, the focus of their investigation is shifting from the individual, taken "in isolation", to the individual as an agent who

interacts with the environment she happens to inhabit; namely, when she deals both with other people, and with the so-called cognitive artefacts, that is, the artificial devices that affect human cognition (Norman 1993, Clark 2003, Heersmink 2013).

All the approaches of that kind seem to fit well with the theoretical concerns expressed *supra* (section 1.1). In order not to overlook the actual linguistic practices we are normally engaged in while doing research, it is useful to conceive language as a tool that primarily exists for communicating with other humans (Enfield 2010) and only secondarily for self-improving one's cognitive operations (Everett 2012). Consequently, a new wave in LR studies could find fruitful suggestions and notions apt to pursue the aforementioned goals. Let us see how.

4E-cognition relies on the assumption that every thought process is not entirely abstract, but is grounded on contextual axes related to the physical bonds on which mental characters are realised (DiFrancesco, Piredda 2012). Interaction with the environment, then, is a factor that contributes to defining the ongoing mental processes. From this perspective, among all the factors that affect cognitive processes, the first should be our body: in fact, low-level processes, such as perceptual and motor ones, seem to be in strict continuity with high-level ones, such as reasoning and cognition in general (Lupyan, Clark 2015). So, 4E-cognition employs a *situated* – and not abstract – notion of cognition, which conforms to the faithful picture of psycholinguistic processes sought here.

Linguistic anthropology too has always considered it crucial to study language in ordinary, daily contexts (Everett 2012, Lupyan 2012). Since language is intrinsically social (Enfield 2010), it seems clear that this proposition supports the notion of distributed cognition (Michael 2002), which serves as a *trait d'union* between research on the functioning of thought and on the nature of language. In this last vein, conversation analysis (Sidnell, Enfield 2012, Enfield, Sidnell 2015) falls within those approaches in philosophy of language and language sciences which study ordinary language and all its possible functions. The aim is not to lose the dynamic features which define the actual use of language.

Now my point should be clearer: there are new domains in which LR effects should be looked for. Nonetheless, LR researchers should be informed of the latest trends in all these disciplines which share

their core interests: language, cognition, what kind of relationship links these two elements of human life, and how speaking two different languages can affect this relationship.

This point is not that original *per se* but, in fact, previous attempts (see Enfield 2015, p. 214) apparently have not been adequately followed up by scholars – philosophers in particular. This new frame for research on LR should bring into play a plurality of disciplines. This is not a simple purpose and, it may seem rather more perilous than promising. Nonetheless, I argue that it is the matter involved itself that demands such a complex approach, without which our understanding of the language-thought problem is bound to remain incomplete. My further claim is that philosophy needs to regain a role in this expansion phase.

3. PHILOSOPHY IN PAST AND PRESENT LR RESEARCH

3.1. NEW PATHS IN RESEARCH

Let me illustrate a few examples of (future) LR research that could benefit from a philosophical contribution. Michael's attempt to reformulate LR may be a starting point: his is an example of empiric research that goes beyond the cognitivist paradigm, thanks to two "theoretical shifts":

> first, from a concern with grammar to a concern with discourse in the context of face-to-face interaction; and second, from individual, isolated cognition, to socially-distributed cognition among a group of individuals. (Michael 2002, p. 108)

This new paradigm unwraps many challenges. First, it is recognized that, so far, the conversational approach to culture has been tied to an individualist model of cognition – which should be integrated. Andy Clark's Extended Mind model (Clark, Chalmers 1998) fulfils this prescription, as it posits that, in Michael's words, cognition is "rarely, if ever, a process bounded by the skull" and "involves interaction with other individuals, and with semiotic artefacts such as texts and maps" (*ibid.*).

Indeed, according to Clark and others, humans inhabit a language-permeated environment (Clark 2003, Steffensen 2009, Enfield 2010). This has consequences for their epistemic access to the world, if we

acknowledge that cognitive artefacts play a critical role in enhancing cognition (Heersmink 2016, p. 78) and that language is "in many ways the ultimate artefact" (Clark 1997, p. 218). If we accept that, then language emerges as "central" for human cognition (Lupyan 2016), both for the high-level processes of abstract prediction and for the perceptual level, which is "cognitively penetrable" (Lupyan, Clark 2015).

It is clear that LR studies need to redefine the role of cognition in the light of this different paradigm. Should we look for LR effects in distributed cognition situations? The answer is yes. This kind of collective cognitive process will, to some extent, depend on the features of the means allowing such a communicative act. Therefore, the linguistic features of cognitive artefacts could be relevant: different languages may have different feedbacks from the artefacts involved, depending on the quality of linguistic diversity between the two, thus generating LR effects.

Following Michael's suggestions on the linguistic side of the problem, linguistic interaction should represent the basic scenario in which LR has to be studied. Scholars belonging to the first two waves were mostly concerned with grammatical structures (e.g. Lucy 1992) and tended to ignore the multiple ways in which they could have been *used* in linguistic interaction (see *supra* for a few exceptions). Language has many more functions than the referential one, which for many (contingent) reasons has been privileged (Enfield 2015, p. 215). This trend had the result of hiding one of its fundamental traits: language is a *social* tool for *action*, as well as for communication and for cognition. The distributed approach to cognition, then, seems a promising frame in which to investigate the nature of human language. Language turns out to be no longer an isolated or individual tool, but a situated and intrinsically social one, given that "human sociality is at the heart of language" (Enfield 2010). In conclusion, experimental research whose subjects are abstracted from "real-life contexts" (*ibid.*) in which everyday social action happens cannot claim to be depicting the actual state of affairs.[9]

For example, it has been shown that in situated social interaction, different languages may have different effects on the kinds of social

[9] See Björk (2008) for a detailed elaboration on LR empiric research and its "segregation" from real-life contexts.

actions that can be achieved, thanks to their different linguistic and pragmatic paths to construct the conversational schema. Here, relativity is about "the different rights and duties that speech acts […] can give you. [...] Language-specific side-effects on normative obligations in a next conversational move arise because of the unavoidable introduction of collateral effects when communicative tools have multiple functional features" (Enfield 2015, p. 218). Speakers of different languages are thus lead "to linguistically relative collateral effects, which lead in turn to differences in our very possibilities for social agency" (Sidnell, Enfield 2012, p. 320–321).

Crosslinguistic differences may have dramatic relevance in domains such as heuristics because decision making is often a less rational process than we may think (Gigerenzer et al. 2011). Since it must be efficient and quick, we rely on simple cues to take decisions, and language sometimes plays a role in this task, since "concepts are sieves" (Enfield 2015, p. 210) that filter what is brought to our attention. In fact, categorisation is one of the most powerful and frequently exploited functions of language. Categorisation is the means through which the concepts which are the basic units of many everyday actions are built up (Clark 1998, Diodato 2015, Enfield 2015).

Before I address in detail these new possibilities of interaction between philosophy and LR studies, let me first discuss *analytic* philosophy.

3.2 LANGUAGE AND REALITY IN ANALYTIC PHILOSOPHY

To be fair, the very notion of analytic philosophy has not a single univocal nor a universally accepted definition. Or, at least, even lengthy attempts at finding strict criteria to define it have somewhat failed (Glock 2008; see Marconi 2014, sec. II). According to Glock, some of the features of a typical analytic philosopher are the willingness to answer substantive questions rather than historical ones following "universally applicable standards of rationality"; the clarity and rigour of the argumentation (Beckermann 2004, p. 12); adhesion to the linguistic turn; rejection of speculative metaphysics[10]; just to name a few.

[10] It must be said, however, that since the second half of the 20th century, analytic philosophers have expanded their area of interest into other branches of philosophy, including metaphysics. Simons (2013, p. 709) states that the analytic

Defining the pure essence of analytic philosophy is clearly too vast a task for the present scope, if possible at all.[11] However, even if we restrict the area under examination, the analytic tradition shows some heterogeneity: different opinions coexist within the same philosophical area, of course. In fact, analytic philosophy is perhaps best defined by appealing to methodological features rather than to some sort of list of common beliefs. It is here argued that among the whole analytic area there is a line of externalist approaches to language and mind that is relevant in LR studies.

Returning to the relation with LR, there is a rich tradition of externalist approaches to meaning and mind in analytic philosophy that must be mentioned as an interesting source of inspiration for the empirical study of how languages affect cognition.

In his later philosophy Wittgenstein (2009) was concerned about the consequences of an internalist approach to thought. For example, in §52 of the *Big Typescript* Wittgenstein (2005) branded as "most dangerous" the idea of "thinking as a process in the head, in that completely closed-off space"; such a sentiment was confirmed by his well-known arguments against, respectively, private language and rule-following.[12] Putnam's renowned Twin Earth mental experiment maintained that in some cases (namely, natural kind terms, indexicals, and proper names) in order to determine the meaning and the reference of such terms definite descriptions or appeals to the subject's internal states are not sufficient, thus postulating the causal role of external factors. Burge took Putnam's intuition even further, claiming that the (at least partial) external determination of the semantic content applies to virtually every other part of language, i.e. not only to natural kind terms etc. In fact, in Burge's account, the relevant

anti-metaphysicism was not even the case at the beginnings of this tradition: "Among those with an outdated or partial conception of analytic philosophy, the whole movement is associated with the rejection of metaphysics. But such rejection, however motivated and justified, was never the sole prerogative of analytic philosophy, nor was it ever the majority view within that movement". In fact, "it was only during the "middle period" of the 1930s–1950s that, under the influence of logical positivism and ordinary language philosophy, metaphysics was first rejected and later marginalized."

[11] In Pietarinen's words, "Such a task will invariably be frustrating" (Pietarinen 2009).

[12] See Wittgenstein (2009) §258–271 and §143–155.

factors in determining the semantic content of certain intentional states are to be retrieved in the linguistic conventions, norms or rules of the given linguistic community.[13] Moreover, Davidson had defended a peculiar form of externalism of the mind, while rejecting Wittgenstein's, Putnam's and Burge's, even though he swung between a physicalist and a social version of externalism (De Caro 2011, p. 181 ff.).

Finally, Quine's *ontological relativity* thesis, based on the famous radical translation argument, bears a clear resonance with LR. Nonetheless, Quine has never explicitly confronted himself with Whorf's work, except for a very brief mention[14] by which we understand that – as was common at the time – he gave a strictly determinist interpretation of the linguistic relativity principle. Indeed, his thesis that linguistic reference or meaning cannot be determined outside the context of a given language (ontological relativity), thus leaving us with the unresolved question of what even the words of our language ultimately refer to,[15] can be related to a form of linguistic determinism. However, it is difficult to say if Quine would have fully endorsed Whorf's view that every culture "carries with it an implicit metaphysics, a model of the universe, composed of notions and assumptions organized into a harmonious system" (Whorf 2012, p. 361), as he argued for an "implicit metaphysics", nestled "in the very structure and grammar" of a given language, "as well as being observable in […] culture and behavior" (*ibid*., p. 75). There are linguists – more precisely, semanticists – who have tried to escape the burden of ontological commitment by analysing crosslinguistic structural differences dropping any claim whatsoever about the "metaphysical reality" embedded in different natural languages. Bach[16], for example, developed his Natural

[13] Putnam 1975 and Burge 1979. See also Amoretti (2013, p. 247–263) for an overview.

[14] Quine 2003, p. 61.

[15] It seems that the reference of the word 'rabbit' remains indeterminate, following Quine (1990, p. 50) where he comments that "'rabbit' refers to rabbits, whatever *they* are", which follows from the assumption that there is no naturalistic "matter of fact" as to what either "*gavagai*" or "rabbit" refer to. Davidson (1989) replied to this kind of *aporiai* deconstructing the "myth of the subjective": according to him, the (somehow quinean) idea that conceptual schemas are immanent to different natural languages or scientific theories is wrong. See Pavan, Sgaravatti (2015) for an overview.

[16] Bach 1986, Bach, Chao 2012. See also Pellettier 2011.

Language Metaphysics as a programme which wanted to answer the question "What do people talk *as if* there is?", as opposed to the "fundamental question of metaphysics 'What is there?'" (Bach, Chao 2012, p. 175). Bach professed modesty:

> Is there a natural language metaphysics? How could there not be? One of our main resources for coming to understand the world is, after all, language, a sort of tool box for doing whatever it is we want to do. Do the fundamental distinctions that are reflected in the overt and covert categories of natural language correspond in any way to the structure of the world? How could they not? But this is where linguistics stops. (Bach 1986, p. 597)

Further, he stated that it was "immoral" of a linguist to make claims whether grammatical objects corresponded to "real things in the world, perceptual or conceptual categories that are independent of language, or to nothing at all" (*ibid*., p. 592). Is this too pretentious an endeavour for philosophers as well? One way or another, empirical studies on LR will hopefully help address the dilemma. Please note that, in this fundamental respect, empirical *cognitive* research is crucially different from Wierzbicka and Goddard's Natural Semantic Metalanguage proposal. The authors, representing a vast number of field linguists, held that it is possible to empirically compile a metalanguage out of the "semantic primitives", i.e. "undecomposable meanings" which were eventually shared by a high number of diverse natural languages. They also maintained that "the simple propositions which can be expressed through the NSMs based on different languages will be fundamentally isomorphic."[17] In other words, according to the Natural Semantic Metalanguage programme, virtually all human languages (therefore all human cultures, and therefore all human beings) share a core set of semantic primitives, forming the common conceptual foundation of all cultures. However, this universalistic programme does not accept Bach's admonition about the danger lurking in inferring too much from merely linguistic data.

Still, the role of *past* philosophical contributions in assessing whether different languages affect the cognitive life of speakers remains uncertain. Humboldt and the other German romantic philosophers' interest in linguistics was linked to the idea that the "inner

[17] Goddard 1994, p. 10. See Goddard, Wierzbicka 1995; Pellettier 2011, p. 4–8.

form" of language of a community was an expression of a people's "national mind and unfolding, in line with the Romantic concept of history" (Koerner 2000, p. 1). So, perhaps, a (broadly construed) philosophical approach to LR is not helpful in terms of avoiding unsubstantiated exaggerations, but on the other hand even linguists are not exempt from such a temptation, from time to time. Let us consider, for example, this quotation from Goddard: "the comparatively muted quality of the English [emotion] words (except for joy, which is the least common of them) is consistent with the traditional Anglo-Saxon dislike of extreme emotions" (Goddard 1998, p. 94). Even if the latter came from a strongly universalist point of view, while the former expressed a clear relativistic attitude, the direction the third wave in LR studies is taking – and, most importantly, how the conceptual and methodological tools of analytic philosophy might be valuable – should be clearer.

3.3 A PHILOSOPHICAL STANCE ON A FEW PROBLEMS

Among the notions named so far, some are of obvious philosophical interest, e.g. 'cognitive artefact', which according to Heersmink (2016), needs to be better understood from a metaphysical point of view, integrating the existing literature in analytic philosophy of technology. Obviously, this notion is embedded in the Extended Mind (EM) paradigm, which has started one of the most interesting recent discussions in analytic philosophy of mind. However, without necessarily committing to the EM theory, the idea that language is a tool which shapes thoughts instead of merely communicating them has wide resonance in psychology (Lupyan 2012, 2016, Borghi et al. 2013, Gentner 2016), artificial intelligence (Mirolli, Parisi 2009) and, of course, philosophy of mind (Dennett 1993, Clark 1998). One argument of Vygotskian descent is central to this view:[18] the private speech of children (which later in development becomes internalized) is a symptom of the child experimenting with its capability of re-shaping the tasks and actions that it wants or is required to perform. Categories that language brings along help the speaker in finding commonalities between distinct objects and by such means simplifies and

[18] See Mirolli and Parisi 2009, p. 519–520 and Clark 1998.

accelerates the (cognitive) action. Thus, language is used as a scaffold in augmenting our cognitive skills – in fact self-referred speech represent "a significant portion of the child's linguistic production" (Mirolli, Parisi 2009, p. 523). It must be said that existing work on this matter concentrates mostly on the role of language in general and not on particular languages, i.e. in a crosslinguistic perspective. Nevertheless, here it is argued that such a line of research should be started. After all, once it is demonstrated that language influences at least some aspects of human cognitive life, the next natural step is to verify whether, other things being equal, different languages have peculiar features in the process of enhancing cognition.

Sidnell and Enfield (2012) considered social interaction as a new "locus" for LR (see *supra*). Within this framework, they note that the concept of "action" has received philosophical attention since Aristotle, although the most influential contribution in recent times is Austin's (1962), along with its followers, starting with Searle (1969). However, this notion needed further revision, at least in the opinion of the authors: the Austinian notion of illocutionary act has been judged insufficient to properly explain how interaction works, as in such a situation, "a person's primary task is to decide how to respond, not to label what someone just did" (Enfield, Sidnell 2017, *ii*). That this kind of philosophical-linguistic analysis may be labelled as "analytic" is argued, among others, by Glock (2008, p. 54), also considering Searle's and Grice's work. It is interesting to consider the authors' challenge to the classical philosophical approach to action:

> We suggest that philosophers and others have created a spurious (though both convenient and intuitive) category of things called actions that are distinct from, and causally related to, the specific practices of conduct and modes of inference through which these 'actions' are realized in interaction. (Enfield, Sidnell 2017, xii)

They criticize the standard account of the notion of "action", questioning the fictitious ontology assumed by scholars (also in conversation analysis and linguistic anthropology) who have argued that "a list or inventory of possible action types" is achievable, in principle, and that, therefore, if an individual wants to perform one of them, "they merely need to provide adequate cues as to which one of these possible actions they mean to be doing" (*ibid*.). So, further philosophical elaboration is needed if we are to understand how social interaction works,

complying with detailed ethnolinguistic data. Thereafter, crosslinguistic differences in social interaction could be better investigated.

Another philosophically relevant notion in analytic ontology is "social reality". According to Searle (2007), social reality is only created through language and it can be approached by linguistic means alone. In a crosslinguistic perspective, it should be investigated if different languages create different socio-institutional realities. As Enfield (2015, p. 216) puts it, "whenever language is used to create social reality […] it is never just language but always *a* language". Notions sensible to crosslinguistic variation would be – among many others, including money, property or corporate identity – that of "social self" and that of "accountability" (see also Sidnell 2017). For instance, public signs that verbally prohibit such and such behaviours may perform this illocutionary act with different nuances depending on the language used, generating correspondently different degrees of accountability for those who do not obey the prescription (in this perspective, multilingual situations would be of great interest).

More generally, the problem of individuality *versus* collectivity is being considered with increasing attention in analytic philosophy, from various perspectives. For example, the linguistic component in the *distribution of agency*, considered as an instance of social interaction, turned out to be crucial, according to Rossi and Zinken (2017). The authors analysed the ways in which Italian and Polish treat impersonal deontic declarative statements (such as the English "it is necessary to") and the relation between the grammatical means of bringing about a request for cooperation and the interactional negotiation of agency (namely, who has to do the required action). Rossi and Zinken concluded that "what may at first glance appear only subtle, differences of expression […] put constraints on what people can or should do in a given situation. Moreover, given the great diversity among languages, grammatical variation will be consequential also for social interaction across cultures" (*ibid*., p. 85). To conclude, the possibility that the fact that "the conceptual distinctions made available by different languages can differ radically […] implies diversity in the kinds of reality that language can create" (Enfield 2015, p. 216) must be taken seriously.

4. CONCLUSION

More than 15 years ago, Michael wrote:

> the long-standing controversy over linguistic relativity has been only modestly impacted by two significant developments in our modern understandings of language and cognition – namely, the now commonplace position that both language and cognition are fundamentally interactional and socially-situated practices that cannot be reduced to isolated, abstract knowledge structures. (Michael 2002, p. 107)

Unfortunately, little has been done since, either in terms of the amount of empirical research following this recent "theoretical shift", or in terms of the number of particular languages taken into consideration (Everett 2013, p. 267). Whatever the reasons, both linguistics and philosophy of language have limited themselves to regarding the referential function of language as its core function giving a biased view of the actual use of language and directing LR research only onto specific trails. This trend has been recently inverted, but another factor that can positively contribute to this change is a philosophical analysis of the notions involved in this paradigm shift. Thus, it will be possible to give an increasingly more accurate account of how language works in real-life contexts. An interdisciplinary approach is certainly needed, and analytic philosophy appears to be the most appropriate companion – perhaps not by referring to the arguments discussed in the past as much as by appealing to its conceptual tools and more recent debates on relevant topics.

In conclusion, let me return to the initial defining issue, and try to update the initial definition (1) with an even more lengthy but complete modified version:

> (2) Linguistic relativity is the idea in accordance to which speakers of different specific varieties of natural languages, which differ in a number of respects studied by linguistics (such as phonetics, syntax, semantics, and pragmatics), could experience the same objects and activities of the world (such as, but not limited to, physical objects perception, colour perception, space relationships, calculus, shaping of categories, decision making) in different ways, interacting with the environment (including external artefacts or other agents), on grounds of that very linguistic diversity – and not because of other factors such as explicit cultural elaboration, or cognitive deficiencies or deviations. Moreover, some forms of linguistic relativity involve domains that exceed individual experience, such as patterns of language-mediated social interaction, or the by-products of social reality, which is created and accessible only through language.

NOTE

Different versions of this paper were presented orally at the 2nd International Philosophy Students' Symposium held at the University of Maribor (Slovenia) and at the 5th International Conference on Philosophy of Language and Linguistics (PhiLang 2017) held at the University of Łódź (Poland).

REFERENCES

Amoretti M. C., (2013). L'esternalismo del contenuto. *Aphex. Portale italiano di filosofia analitica*, 7, 247–279.

Austin, J. L., (1962). *How to Do Things with Words*. Cambridge, MA: Harvard University Press.

Bach, E. (1986). Natural Language Metaphysics. In Marcus, R. Barcan, G. J. W. Dorn, P. Weingartner (Eds.) *Logic, Methodology, and Philosophy of Science, VII* (pp. 573–595). Amsterdam: North Holland.

Bach, E., Chao, W. (2012). The Metaphysics of Natural Language(s). In R. M. Kempson, T. Fernando, N. Asher (Eds.), *Philosophy of Linguistics, 14* (pp. 175–196). Amsterdam: North Holland, Elsevier.

Beckermann, A. (2004). Einleitung. In P. Prechtl (Ed.), *Grundbegriffe der analytischen Philosophie* (pp. 1–12). Stuttgart: Metzler.

Berlin B., Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Björk, I. (2008). *Relativizing Linguistic Relativity: Investigating Underlying Assumptions about Language in the Neo-Whorfian Literature*. Studia Linguistica Upsaliensia 6. Uppsala: Uppsala University.

Bloom, P., Keil, F. C. (2001). Thinking Through Language. *Mind & Language*, *16*, 351–367.

Borghi, A. M., Scorolli, C., Caligiore, D., Baldassarre, G., Tummolini, L. (2013). The Embodied Mind Extended: Using Words as Social Tools. *Frontiers in Psychology*, *4*.

Boroditsky, L. (2001). "Does Language Shape Thought? English and Mandarin Speakers' Conceptions of Time". *Cognitive Psychology*, 43(1), 1–22.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*, 73–122.

Casasanto, D. (2008). Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought. *Language Learning*, *58*(suppl. 1), 63–79.

Casasanto, D. (2016). Linguistic Relativity. In N. Riemer (Ed.), *Routledge Handbook of Semantics* (pp. 158–174). New York: Routledge.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass. – London: MIT Press/Bradford Books.

Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P. Carruthers, J. Boucher (Eds.), *Language and Thought: Interdisciplinary Themes* (pp. 162–183). Cambridge: Cambridge University Press.

Clark, A. (2003) *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*. New York: Oxford University Press.

Clark, A., Chalmers, D. (1998). The Extended Mind. *Analysis*, *58*(1), 7–19.

Davidson, D. (1984). On the Very Idea of a Conceptual Scheme. In *Inquire into Truth and Interpretation* (pp. 183–198). Oxford: Clarendon Press.

Davidson, D. (1989). The Myth of the Subjective. In M. Krausz (Ed.), *Relativism: Interpretations and Confrontations* (pp. 159–172). Bloomington: Indiana University Press.

De Caro, M., (2011). The Short Happy Life of the Swampman: Davidson and Social Externalism. In Amoretti, C., Preyer, G. (Eds.), *Triangulation: From an Epistemological Point of View* (pp. 179–196). Frankfurt: Ontos.

Dennett, D. C., (1993). Learning and Labelling. *Mind and Language*, *8*(4), 540–547.

Di Francesco, M., Piredda F. (2012). *La mente estesa*. Milan: Mondadori Università.

Diodato, F. (2015). *Teorie della categorizzazione*. Naples: Liguori.

Dor, D. (2015). *The Instruction of Imagination. Language as a Social Communication Technology*. Oxford: Oxford University Press.

Enfield, N. J. (2010). *Human Sociality at the Heart of Language*, prolusion. Nijmegen: Radboud University Nijmegen.

Enfield, N. J. (2015). Linguistic Relativity from Reference to Agency. *Annual Review of Anthropology*, *44*(1), 207–224.

Enfield, N. J., Kockelman, P., (Eds.) (2017). *Distributed Agency*. Oxford: Oxford University Press.

Enfield N. J, Sidnell J. (2015). Language Structure and Social Agency: Confirming Polar Questions in Conversation. *Linguistics Vanguard*, *1*(1), 131–143.

Enfield, N. J., Sidnell, J. (2017). *The Concept of Action*. Cambridge: Cambridge University Press.

Everett, C. (2013). *Linguistic Relativity. Evidence Across Languages and Cognitive Domains*. Berlin, Boston: De Gruyter Mouton.

Everett, D. L. (2012). *Language: The Cultural Tool*. London: Profile.

Gentner, D. (2016). Language as Cognitive Tool Kit: How Language Supports Relational Thought. *American Psychologist*, *71*(8), 650–657.

Gigerenzer, G., Hertwig, R., Pachur, T. (Eds.) (2011). *Heuristics: The Foundations of Adaptive Behavior*. New York: Oxford Univ. Press.

Glock H.-J., (2008). *What Is Analytic Philosophy?* Cambridge: Cambridge University Press.

Goddard, C. (1994). Semantic Theory and Semantic Universals. In *Semantic and Lexical Universals* (pp. 7–29). Philadelphia: J. Benjamins.

Goddard, C. (1998). *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.

Goddard, C., Wierzbicka, A. (1995). Key Worlds, Culture and Cognition. *Philosophica*, *55*, 37–77.

Gumperz. J. J., Levinson, S. C. (Eds.) (1996). *Rethinking Linguistic Relativity*. Cambridge: Cambridge University Press.

Heersmink, R. (2013). A Taxonomy of Cognitive Artefacts: Function, Information, and Categories. *Review of Philosophy and Psychology*, *4*(3), 465–481.

Heersmink, R. (2016). The Metaphysics of Cognitive Artefacts. *Philosophical Explorations*, *19*(1), 78–93.

Humboldt, W. v. (1904). Essai sur les langues du Nouveau Continent. *Gesammelte Schriften*, III, 300–341.

Humboldt, W. v. (1988). *On Language: The Diversity of Human Language-Structure and Its Influence on the Mental Development of Mankind*. P. Heath (Ed.). Cambridge: Cambridge University Press.

Imai, M., Mazuka, R. (2003). Re-evaluation of Linguistic Relativity: Language--Specific Categories and the Role of Universal Ontological Knowledge in the Construal of Individuation. In D. Gentner, S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Issues of Language and Thought* (pp. 430–464). Cambridge, MA: MIT Press.

Koerner, E. F. K. (2000). Towards a 'Full Pedigree' of the 'Sapir-Whorf Hypothesis'. From Locke to Lucy. In M. Pütz, M. Verspoor (Eds.), *Explorations in Linguistic Relativity* (pp. 1–24). Amsterdam-Philadelphia: John Benjamins.

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

Lee, P. (1996). *The Whorf Theory Complex: A Critical Reconstruction*. Amsterdam–New York: John Benjamins.

Levinson, S. C. (2003). *Space in Language and Cognition*. Cambridge: Cambridge University Press.

Lucy, J. A. (1992a). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. A. (1992b). *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.

Lupyan, G. (2012). What Do Words Do? Toward a Theory of Language-Augmented Thought. *Psychology of Learning and Motivation*, *57*, 255–297.

Lupyan, G. (2016). The Centrality of Language in Human Cognition. *Language Learning*, *66*(3), 516–553.

Lupyan, G., Clark, A. (2015). Words and the World Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279–284.

Marconi, D. (2014). *Il mestiere di pensare*. Torino: Einaudi.

Michael, L. (2002) Reformulating the Sapir-Whorf Hypothesis: Discourse, Interaction, and Distributed Cognition. *Proceedings of the Tenth Annual Symposium about Language and Society – Austin*. Citeseer.

Mirolli, M., Parisi, D. (2009). Language as a Cognitive Tool. *Minds and Machines*, *19*(4), 517–528.

Norman, D. (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. New York: Basic Books.

Pavan S., Sgaravatti D., (2015). *Willard Van Orman Quine*. Aphex. Portale italiano di filosofia analitica, nr. 11.

Pelletier, F. J. (2010). Descriptive Metaphysics, Natural Language Metaphysics, Sapir-Whorf, and All That Stuff: Evidence From the Mass-Count Distinction. *The Baltic International Yearbook of Cognition, Logic and Communication*, *6*(1), 1–46.

Penn, J. (1972). *Linguistic Relativity versus Innate Ideas: The Origins of the Sapir-Whorf Hypothesis in German Thought*. Paris: Mouton.

Pharao Hansen, M. (2009). *A Critique of the Treatment of Linguistic Relativity in Steven Pinker's* 'The Stuff of Thought'. Accessed December 6, 2017, https://www.academia.edu/1610936/A_critique_of_the_treatment_of_Linguistic_Relativity_in_Steven_Pinkers_The_Stuff_of_Thought

Pietarinen, A. (2009). Significs and the Origins of Analytic Philosophy. *Journal of the History of Ideas*, *70*(3), 467–490.

Pinker, S. (1994). *The Language Instinct*. New York, NY: Harper Perennial Modern Classics.

Pinker, S. (2007). *The Stuff of Thought: Language as a Window Into Human Nature*. New York, NY: Viking.

Putnam, H. (1975). The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.

Quine, W. V. O. (1990). *Pursuit of Truth*. Cambridge, MA: Harvard University Press.

Quine, W. V. O. (2003) [1953]. *From a Logical Point of View: 9 Logico-Philosophical Essays. Second Edition Revised*. Harvard: Harvard University Press.

Rosch, E. (1972) Universals in Color Naming and Memory. *Journal of experimental Psychology*, *93*, 10–20.

Rossi, G., Zinken J. (2017). Social Agency and Grammar. In N. J. Enfield, P. Kockelman (Eds.), *Distributed Agency* (pp. 79–86). Oxford: Oxford University Press.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. London: Cambridge University Press.

Searle, J. R. (2007). Social Ontology and the Philosophy of Society. E. Margolis, S. Laurence (Eds.), *Creations of the Mind*. Oxford, New York: Oxford University Press.

Sidnell, J. (2017). Distributed Agency and Action under the Radar of Accountability. In N. J. Enfield, P. Kockelman (Eds.), *Distributed Agency* (pp. 87-86). Oxford: Oxford University Press.

Sidnell, J., Enfield, N. J. (2012). Language Diversity and Social Action: A Third Locus of Linguistic Relativity. *Current Anthropology*, *53*(3), 302–333.

Simons, P. (2013). Metaphysics in Analytic Philosophy. In Michael Beaney, *The Oxford Handbook of the History of Analytic Philosophy* (pp. 709–728). Oxford: Oxford University Press.

Slobin, D. I. (1996). *From "Thought and Language" to "Thinking for Speaking"*. In J. J. Gumperz, S. C. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Steffensen, S. V. (2009). Language, Languaging, and the Extended Mind Hypothesis. *Pragmatics & Cognition*, *17*(3), 677–697.

Tartabini, A. (2011). Aspetti evolutivi della comunicazione animale ed umana. In N. Grandi (Ed.), *Dialoghi sulle lingue e sul linguaggio* (pp. 127–134). Bologna: Pàtron Editore.

Wittgenstein, L. (2005). *The Big Typescript, Ts. 213*. Hoboken: Blackwell.

Wittgenstein, L. (2009). *Philosophical Investigations*. 4th edition, transl. P. M. S. Hacker and J. Schulte. Oxford: Wiley-Blackwell.

Whorf, B. L. (2012). *Language, Thought, and Reality*. Cambridge, MA: MIT Press.

Wolff P., Holmes K. J., (2011). Linguistic Relativity. *Wiley Interdisciplinary Review of Cognition, Sci.*, *2*(3), 253–265.

Zinken, J. (2008). The Metaphor of 'Linguistic Relativity'. *History & Philosophy of Psychology*, *10*(2), 1–10.

NOTES ABOUT AUTHORS

Mieszko Tałasiewicz – dr hab., prof. UW, Instytut Filozofii, Zakład Semiotyki Logicznej, Uniwersytet Warszawski, ul. Krakowskie Przedmieście 3, 00-927 Warszawa, Poland

Hsiang-Yun Chen – Ph.D., Assistant Research Fellow, Institute of European and American Studies, Academia Sinica, No. 128, Academia Road, Secttion 2, Nangang, Taipei 115, Taiwan

Filip Kawczyński – Ph.D., Assistant Professor, Instytut Filozofii, Zakład Semiotyki Logicznej, Uniwersytet Warszawski, Krakowskie Przedmieście 3, 00-927 Warszawa, Poland

Adriano Marques da Silva – Ph.D., Assistant Professor, Departamento de Filosofia, Instituto Federal da Paraiba, Cidade Universitária, s/n – Castelo Branco III, PB, 58051-085 João Pessoa, Brazil

Joanna Klimczyk – Ph.D., Assistant Professor, Instytut Filozofii i Socjologii, Zakład Logiki i Kognitywistyki, Polska Akademia Nauk, ul. Nowy Świat 72, 00-330 Warszawa, Poland

Erich Rast – Ph.D., Senior Research Fellow, IFILNOVA Institute of Philosophy, Universidade Nova de Lisboa, Av. de Berna 26, 1069-061 Lisboa, Portugal

Federico L. G. Faroldi – Ph.D., chargé de recherches, postdoctoraal onderzoeker, Research Foundation, Flanders (FWO), Centre for Logic and Philosophy of Science, Ghent University, Blandijnberg 2, 9000 Gent, Belgium

Andrés Soria Ruiz – Ph.D. Candidate, Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL Research University, CNRS, 29 rue d'Ulm, 75005, Paris, France

Jakub Gomułka – dr hab., Assistant Professor, Instytut Filozofii i Socjologii, Katedra Epistemologii, Logiki i Metodologii Nauk, Uniwersytet Pedagogiczny im. Komisji Edukacji Narodowej, ul. Podchorążych 2, 30-084 Kraków, Poland

Jan Wawrzyniak – dr hab., Assistant Professor, Instytut Filozofii i Socjologii, Katedra Filozofii Współczesnej, Uniwersytet Pedagogiczny im. KEN, ul. Podchorążych 2, 30-084 Kraków, Poland

Filippo Batisti – Ph.D. Student, Department of Philosophy and Cultural Heritage, Ca' Foscari University of Venice, Palazzo Malcanton Marcorà, Dorsoduro 3484/D, 30123 Venice, Italy

STUDIA SEMIOTYCZNE

Tom XXXI • Numer 2 • 2017

W numerze:

ISSN 0137-6608