Studia Semiotyczne, t. XXXI, nr 1 (2017), s. 23–43
ISSN 0137-6608
doi: 10.26333/sts.xxxi1.03
Rozprawa

Paweł Grabarczyk*

# DIRECTIVAL THEORY OF MEANING RESURRECTED[1]

Summary: The first aim of this paper is to remind the reader of a very original theory of meaning which in many aspects has not been surpassed by subsequent theories. The theory in question is Kazimierz Ajdukiewicz's Directival Theory of Meaning. In the first section I present a version of this theory which, I trust, retains the gist of the original but loses its outdated language. In the second section I analyze some problematic consequences of the directival theory (specifically Tarski's counterexample) and show how they can be addressed.

The second aim of this paper is exploiting some of the similarities between the directival theory and later theories of meaning. In the third section I argue that using the directival theory as an interpretative tool enables us to create explications of some of the notoriously vague notions which contemporary theories of meaning employ.

Keywords: Kazimierz Ajdukiewicz, directival theory of meaning, semantics, indeterminacy of translation, Wilfrid Sellars

There are two aims of this paper. The first aim is to remind the reader of a very original theory of meaning which in many aspects has not been surpassed by subsequent theories. The theory in

* University of Lodz, Faculty of Philosophy and History. E-mail: pagrab@gmail.com

[1] This paper is part of the project financed by the Polish National Science Centre, Grant number 2014/15/B/HS1/01928.

question is Kazimierz Ajdukiewicz's Directival Theory of Meaning (henceforth DTM). It was the world's first foray into functional role semantics, predating Wittgensteinian intuitions of "meaning as use" (Wittgenstein 1967) by almost 20 years. Despite this it has never been widely recognized or analyzed outside of Poland (apart from passing remarks by Carnap (Carnap 1959) and Quine (Quine 2013, p. 59)). There are two reasons for this. The first reason is that its original presentation leaves a lot to be desired when it comes to accessibility. In section 1 I present a version of the DTM which, I trust, will retain the gist of the original version but lose its outdated language and will simplify it as much as possible. The second reason is that it was quickly abandoned by Ajdukiewicz because of its counterintuitive consequences. In section 2 I will analyze these consequences (specifically Tarski's counterexample) and show how they can be addressed.

The second aim of this paper is to exploit some of the similarities between the DTM and later theories of meaning, specifically the theories of Wilfrid Sellars, Ned Block, Jerry Fodor and Willard Van Orman Quine. In section 3 I will show that apart from being a theory of meaning DTM can also be used as a pretty robust interpretative tool. I argue that using DTM in this manner not only helps us to understand these theories better but also enables us to create explications of some of the notoriously vague notions these theories employ.

## 1. DIRECTIVAL THEORY OF MEANING EXPLAINED

The directival theory of meaning was developed by Kazimierz Ajdukiewicz over two papers: *O znaczeniu wyrażeń* (*On The Meaning Of Expressions*)[2] and *Sprache und Sinn* (*Language and Meaning*)[3]. Although only the latter paper presents the full-blown version of the theory, it is important to remember about the former as it contains some

[2] The original Polish version has been published in (Ajdukiewicz 1985b) and can also be found in (Ajdukiewicz 1985b) the English translation can be found in (Ajdukiewicz 1978b).

[3] The original paper can be found in (Ajdukiewicz 1934), the Polish translation can be found in (Ajdukiewicz 1985a), the English translation can be found in (Ajdukiewicz 1978a).

preliminary considerations that have shed much needed light on assumptions which are crucial for understanding the theory[4].

It is worth starting with the central intuition that motivated DTM. It is so ubiquitous and common that it could be summed up in a popular slogan: "People do not argue over semantics". What this means is that sometimes the argument between two sides reaches a point where the sides start to suspect that the disagreement is merely verbal.

What happens next is interesting because of two reasons. The first interesting thing is that (for most of the time) people know how to test their suspicion. Contrary to what might seem to be the obvious solution they do not expect their interlocutors to provide a full definition of the problematic expression. Instead, they try to detect the suspected verbal difference by asking a few key questions about the expression. So, for example, if I was to discover that my interlocutor uses the term "idea" the same way I do, I may start by asking if "ideas" are mental entities. If the answer indicates a difference in usage, it might be enough to decide that the dispute was only verbal, that she meant something different – e.g. platonic ideas.

The second interesting thing is that the moment the two sides discover that the difference was only verbal the disagreement disappears[5]. Most of the time people do not have the motivation to fight with conventions because there is no right or wrong there and some of the conventions are mandatory: either you accept the convention and stay with the community that supports it, or you do not and you are automatically excluded from that community. Starting with these common sense observations Ajdukiewicz presumed that for every noncompound expression there are mandatory conventions and that they are adhered to in the act of confirming certain sentences. When someone knows the meaning of a given expression, and are then asked about it, they have to confirm certain sentences that this expression figures in. And if they refuse to do so, they are excluded from the community of users of this particular expression. Naturally,

[4] Some researchers consider both papers to be two different versions of the theory (Hanusek 2013).

[5] Or is vastly diminished. The point here is that it is significantly easier to achieve agreement, even if we have different views on which of the available dictionaries is to be treated as obligatory.

the model examples of these mandatory conventions are analytic sentences. For example, if you refuse to confirm a sentence "A circle is a figure" then you will be denied the knowledge of the meaning of the term "a circle"[6] and once it is revealed that there is a (admittedly unspecified) number of expressions you do not know the meaning of, you will not be treated as an English speaking person.

The novel idea Ajdukiewicz adds to these observations is his insistence that it should work both ways – if you accept a certain set of sentences which contain a given expression, you can be said to know its meaning. There is nothing more to it – to know the meaning of a word is to have a disposition to confirm its meaning directives (as the specified set of obligatory sentences are to be called). So, what do these meaning directives look like? In general a directive can be presented as a sentence in the form:

If $u$ is a user of a language $L$ and $u$ is in a situation $S$ then $u$ confirms a sentence $p$.

It is easy to see that the normativity of meaning is built into the directives from the start. Using a simple rule of contraposition we can derive the following consequence: if someone does not confirm sentence $p$ than either they are not in the situation $S$ or they are not a speaker of the language $L$. It means that if the user is allowed to disregard language directives they are automatically excluded from a given speaking community[7]. One thing to keep in mind is that what we talk about is the act of confirmation of a sentence and not the act of utterance. It is worth pointing this difference out because ignoring it may easily lead to a significant misinterpretation. The theory does not require the user to produce utterances automatically whenever they are in a given situation but only to react accordingly whenever they are asked to confirm the sentence $p$ in a proper situation. Again, analytic sentences are a good example here. We are not expected to walk around and whisper them to ourselves all the time. What is

[6] Of course you might as well be denied the knowledge of the meaning of the word "a figure" but it will be tested the same way – you will be asked to accept some other sentences the term "a figure" figures in.

[7] Needless to say it is an idealization. The forbidden behavior would have to be somewhat systematic for her to be really excluded. The important part is that the behavior would be treated as an error and not as an expression of their (even very peculiar) point of view.

expected of us instead is a constant, enduring disposition to confirm them when asked to.

We can now group the meaning directives into three sets depending on the type of situation *S*.

| **Directive type** | Axiomatic | Inferential | Empirical |
|---|---|---|---|
| **Character of *S*** | Anything | Set of sentences | Physical stimulus |
| **Example of *S*** | | "*p* implies *q*"<br>"*p* obtains" | The touching of a nerve |
| **Sentence to confirm** | "*A* is identical to *A*" | "*q* obtains" | "It hurts!" |

**Table 1:** Types of meaning directives

Now let me characterize the types of directives indicated above.

In the case of axiomatic directives there are no requirements specified as to what situation *S* has to be. It can be any possible stimulus – verbal, physical or a combination of both. There can be no stimulus at all. The point here is that in each and every situation the user is expected to confirm some of the sentences of her language (such as the identity statement used in the table).

The inferential directives seem to be another intuitive example of the idea of obligatory rules: after all, this is how most of us learn logic – we are told that whenever we confirm a given sentence we have to confirm another, subsequent sentence. If we do not follow those instructions we will not master logic because it is exactly what mastering logic boils down to. This normative aspect of logic works exactly the same way as it is supposed to work in the DTM.

Last but not least, we have empirical directives. It is important to note that the way I explain them here presents the most significant departure from Ajdukiewicz's version. As can be seen in Table 1, I have described the situation *S* which precedes the confirmation of the sentence *p* as a physical stimuli. Contrary to this, Ajdukiewicz referred to mental states rather than to their physical causes. But despite the psychological language that he was using most of his examples of empirical directives adhere to physical stimuli and not their mental correlates. Case in point: in the example I have used in the table above Ajdukiewicz talks about the expected confirmation of

the sentence "It hurts!" when a dentist touches the nerve of a patient's tooth and not about the feeling of pain[8].

There is an additional difficulty that most of the examples of empirical stimuli lead to. If I am presented with an object and asked to confirm the sentence "This object is red", I may refrain from doing so because I believe that the lighting in the room is so different from normal lighting that I am no longer sure of the object's color. It complicates matters because we have to expand the directive by a requirement that the user has a belief that the situation (understood as a state of the environment and the perception apparatus) is typical or normal. The addition of beliefs introduces a hybrid category of directives, a mix between the empirical and the inferential ones, one part of the situation *S* being a sentence expressing the belief and the other being a stimulus. Ajdukiewicz mentions this complication but does not elaborate on it (Ajdukiewicz 1934). I too am going to skip it in the present exposition of the DTM.

So, how is the notion of meaning to be derived from these three types of directives? Let us assume that we created a list of directives for every noncompound expression of the language. Once we have it, the next step would be to get the notion of s y n o n y m y. The intuitive formulation of the relation between the meaning directives and synonymy is this: expressions are synonymous when the meaning directives describe them identically. To present the notion of synonymy in a less metaphoric fashion we have to use an example of a very simple language. Let us say that it contains only the following axiomatic directives:[9]

$$P(a),\ S(c),\ R(c),\ R(d),\ Q(b),\ P(b),\ P(c),\ Q(a)$$

Now, focus on terms *a* and *b*. The interesting thing about them is that if you switch their places – replace every instance of *a* with *b* and *vice versa* you will end up with the same list of directives – the only difference being the order of the directives:

[8] The other important reason for preferring physical stimuli over mental states is that it will make our task in section 3 much easier.

[9] That these are axiomatic directives can be easily deduced from their syntactic structure. Only axiomatic directives can be presented as a single sentence.

$$P(b), S(c), R(c), R(d), Q(a), P(a), P(c), Q(b)$$

Using this observation Ajdukiewicz proposed to use this operation of systematic simultaneous replacement of terms to define the notion of synonymy:

Expressions $a$ and $b$ are s y n o n y m o u s iff they can be simultaneously replaced in all respective meaning directives without changing the sum of all the meaning directives of the language.

The obvious next step is to use abstraction to obtain the definition of meaning:

The m e a n i n g o f a n e x p r e s s i o n is the set of all the expressions which are synonymous with it.

It is easy to see that in most cases this definition yields rather disappointing results: in the case of expressions which are not synonymous with any other expression their meaning turns out to be a singleton consisting only of themselves. To counter this, Ajdukiewicz introduces a new (and at the time rather novel) idea: he proposed to define meaning by appealing to the notion of t r a n s l a t i o n. To present it, we will use another example of a simple language, let us call it $L$. Let $L$ contain the following terms: two one-place predicates: – $P(x)$, $Q(x)$; three constants – $a$, $b$, $c$; one zero-place predicate (a sentential constant) $Z$. Additionally we introduce three symbols which signify physical stimuli: $\alpha$, $\beta$, $\gamma$. It is important to stress that these symbols are not parts of $L$. They symbolize the extra-linguistic element in empirical directives. Now assume that $L$ contains the following directives:

Axiomatic directives:

1. $P(a)$[10]
2. $P(a)$ & $Q(b)$

Inferential directives:

1. $P(a) \models Q(b)$[11]

[10] Understood as: "in every situation confirm the sentence $P(a)$" and so on.

[11] Understood as: "If you confirm the sentence $P(a)$ you have to confirm the sentence $Q(b)$" and so on.

2. $P(a)$ & $Q(b)$ $\models Q(c)$[12]
3. $Q(b)$ $\models Z$

Empirical directives:

1. $\alpha$; $Z$[13]
2. $\beta$; $Q(b)$
3. $\gamma$; $Z$

Having all this we are ready to build something Ajdukiewicz called a language matrix[14]. A language matrix is divided into three sections corresponding to three types of directives. In our example they are designated by numerals 1, 2 and 3 in the first column of the table. Horizontally the table is divided into two parts indicating two parts of a directive: the first part (designated by the Roman numeral I) contains the situation specified by the directive (or the lack of a specified situation in the case of axiomatic directives), the second part (designated by the Roman numeral II) contains the sentence which the directive requires to be confirmed. Every sentence put into a language matrix is divided into its constituent parts using the following procedure: the first cell contains the sentence itself, the next cell contains its main connective or a predicate (in the case of an atomic sentence), the next cell contains the first argument of the connective (or an argument of the predicate). Then the same procedure applies to the first argument – we put its main connective first, then its first argument and so on. When we achieve the level of atomic parts we move on to the second argument of the main connective of the sentence we started with. The pattern is repeated for as long as there is nothing more to decompose. If we applied this procedure to our simple language we would end up

[12] You might be surprised that, given the existence of the axiomatic directive 2 and the inferential directive 2 the sentence $Q(c)$ is not an axiomatic directive as well. After all, it is a consequence of these directives taken together. The point of this example is to show that some of the consequences of the language rules are not by itself language rules and can be overlooked by the language user. This characteristic of the DTM will be used later in section 3.

[13] Understood as "In this situation (when the situation is $\alpha$) confirm the sentence $Z$" and so on (I use an indexical term to stress the extra-linguistic aspect of $\alpha$).

[14] This part is a substantially modified version of the original example. First of all, I use a modern predicate logic notation and secondly, I present the matrices in a more visual way which I believe makes the whole idea much easier to grasp.

with the following table (note the extra-linguistic part in the left bottom corner).

| | I | | | | | | | | II | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1.1 | | | | | | | | | *P*(*a*) | *P* | *a* | | | | | |
| 1.2 | | | | | | | | | *P*(*a*) & *Q*(*b*) | & | *P*(*a*) | *P* | *a* | *Q*(*b*) | *Q* | *b* |
| 2.1 | *P*(*a*) | *P* | *a* | | | | | | *Q*(*b*) | *Q* | *b* | | | | | |
| 2.2 | *P*(*a*) & *Q*(*b*) | & | *P*(*a*) | *P* | *a* | *Q*(*b*) | *Q* | *b* | *Q*(*c*) | *Q* | *c* | | | | | |
| 2.3 | *Q*(*b*) | *Q* | *b* | | | | | | Z | | | | | | | |
| 3.1 | *A* | | | | | | | | Z | | | | | | | |
| 3.2 | *B* | | | | | | | | *Q*(*b*) | *Q* | *b* | | | | | |
| 3.3 | *Γ* | | | | | | | | *Z* | | | | | | | |

**Table 2:** The language matrix of L

The main point about a language matrix is that it enables us to extract the structure of the language and abstract away from the actual expressions it uses. We could do that in a variety of ways but I find it the easiest to simply use some sort of visual indication. To extract the structure we are interested in we simply replace the symbols with graphical patterns; let us call it an expressionless language matrix.

| | I | | | | | | | | II | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1.1 | | | | | | | | | | | | | | | | |
| 1.2 | | | | | | | | | | | | | | | | |
| 2.1 | | | | | | | | | | | | | | | | |
| 2.2 | | | | | | | | | | | | | | | | |
| 2.3 | | | | | | | | | | | | | | | | |
| 3.1 | *A* | | | | | | | | | | | | | | | |
| 3.2 | *B* | | | | | | | | | | | | | | | |
| 3.3 | *Γ* | | | | | | | | | | | | | | | |

**Table 3:** The semantic structure of L
(expressionless language matrix)

Now you could fill this table anew using the following rules:

1. You do not change the $\alpha$, $\beta$, $\gamma$ records as they are extra-linguistic elements of the table.

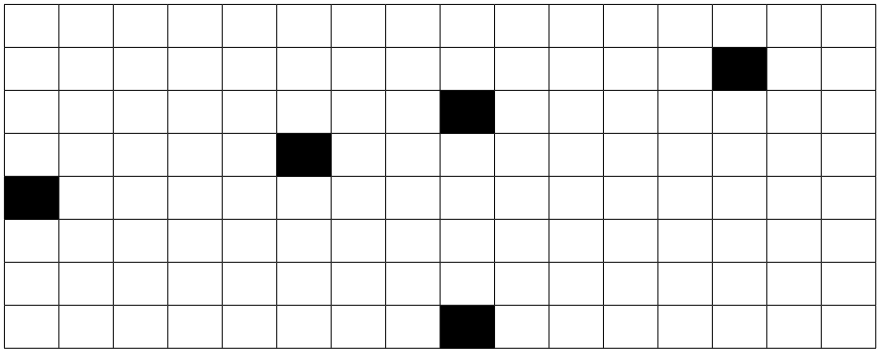
2. You do not fill the white records.

3. Whenever you put something in the record you have to repeat the same symbol in every record with the same pattern.

Every table obtained this way represents a language, which is translatable to the language we started with. Finally, the idea of a language matrix gives us the possibility to define meaning:

The meaning of a noncompound expression $t$ in the language $L$ is an ordered pair $\langle SL, P\rangle$ consisting of the structure of $L$ ($SL$) and the set of places $t$ occupies in this structure ($P$)[15].

As you have seen, the structure can be presented in the form of a language matrix and the set of places a given expression occupies can as well be shown visually. So, for example, the meaning of the expression $Q(b)$ from our table can be presented via the following diagram[16].



It shows that the DTM realizes the noble goal of a reductive, syntactic definition of meaning – the meaning can be literally represented as a shape, which makes it easy to handle mechanically. The fact that what we started with are the acts of confirmation of sentences just adds

[15] The relation of synonymy can still be defined using the notion of mutual exchangeability in meaning directives, just like we did on page 6.

[16] To stress the possibility of representing the meaning of the expression visually I omitted the extra-linguistic parts of the table. It is possible whenever a language matrix is fixed.

a dash of pragmatics to the definition. Because of this, the DTM could not be called a purely syntactic theory. The fact remains, though, that it is a theory in which no part uses any semantic notion. It is an idea entertained by many, but I guess that it is summed up most eloquently by Chomsky:

> It is possible that natural language has only syntax and pragmatics; it has a "semantics" only in the sense of "the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation is actually put to use in a speech community" (Chomsky 1995).

It is important to realize that even though the language matrix contains an extra-linguistic part, the theory does not stipulate that any of the expressions present in the matrix refer to these extra-linguistic elements. Moreover, even if the theory deals with the confirmation of sentences, in no part does it assume the sentences to be true. You might assume that they are held to be true by the users but it would be an additional assumption the theory does not depend on.

## 2. DIRECTIVAL THEORY OF MEANING CHALLENGED

Ironically, this attractive feature of the DTM (its independence from reference) is exactly what killed it. It happened because of a very simple example that Alfred Tarski confronted Ajdukiewicz with (he did so in a conversation and it was reported many years later in (Ajdukiewicz 1978c)). Consider a very simple language of predicate logic (with identity) and add to it two new axiomatic directives:

$$A \neq B$$
$$B \neq A$$

$A$ and $B$ are extra-logical constants which appear only in these very directives. The problem is that the two expressions are mutually interchangeable in all the meaning directives of the language (because there are only two such directives and you can mutually replace them). On the other hand, we have to assume that both expressions do not refer to the same object, because it is precisely how we normally interpret the negation of the identity sign. It means that the DTM allows two expressions to have the same meaning but a different reference and it seems that we do not have any means within the

theory to block this unintuitive result because the theory does not say anything about the reference of the expressions[17].

It turns out that in spite of deliberately ignoring all the semantic notions Ajdukiewicz still wanted his theory to be Fregean – the meaning of the expression was supposed to determine its reference. It was so obvious to him that he did not even try to argue for it and remarked only that such a consequence was unacceptable (Ajdukiewicz 1978c). Fortunately, it is a sentiment we do not have to share today as there are at least three ways out of the trouble Tarski's example puts us in – ways which do not force us to abandon the reductive, non-semantic aspect of the DTM.

First of all, we can say that the objection works only because the example language does not contain any empirical directives. If it did, they would have differentiated the terms *A* and *B*. And in the case of uninterpreted languages there is no problem of reference anyway. This is the solution suggested by Ajdukiewicz himself (Ajdukiewicz 1978c). The question of whether this solution is effective is highly debatable though (see section 3).

The second thing we can modify is the simultaneous interchangeability requirement of the synonymy relation. It has been shown that we can modify this requirement and demand only when the terms *A* and *B* can be considered synonymous if and only if it is possible to replace *A* with *B* and then *B* with *A* (but not simultaneously) without changing the character of the directive we applied this procedure to. This means that if something has been an axiomatic directive, it remains an axiomatic directive after the replacement of the term (similarly for the other two types of directives). This solution has some disadvantages, but they will not be discussed here[18].

The third, and perhaps most interesting option, is that we could simply accept and embrace this surprising consequence of the theory

[17] It is worth noting that Tarski's example is very similar in spirit to Fodor and Lepore's objection against functional role semantics. As Fodor and Lepore rightly argue (Fodor, Lepore 1992, p. 170) the price hybrid theories pay for their flexibility is that there is nothing that prevents a given sentence having the inferential role of "4 is a prime number" but the truth conditions of "water is greenish" (as there is no necessary connection between inferential role and truth conditions).

[18] The results in question has been published only recently by (Nowaczyk 2006) and (Buszkowski 2010). Unfortunately both articles are only available in Polish.

– especially that it is not so surprising anymore. After all, this is what Putnam's Twin Earth thought experiment was set to do – it showed us that we do not have to hold to Fregean intuitions about the relation between meaning and reference (Putnam 1975). Could not we simply decide that a sensible strategy for a theory of meaning is to contain two parallel theories – a theory of reference and a separate theory of meaning which answers the questions about synonymy, translatability and meaningfulness of expressions?

Unfortunately the DTM has more issues than that. Specifically, there are two problematic theses it holds (one of them being an assumption, the other a consequence) which we have to analyze if the theory is to be useful for contemporary philosophers. We will refer to them later, so it might be convenient to label them:

(T1) The meaning of every word in the language changes whenever a new word is added to the vocabulary.

(T2) Syntaxes of all translatable languages have to be perfectly compatible.

(T1) is a direct consequence of definition (D2) presented above. If the meaning of a particular expression is the ordered pair of a language matrix and a set of places the expression figures in, then the meaning changes whenever the matrix changes, and the matrix changes whenever a new expression is added. It is so because the new expression has to have a set of new directives which regulate its usage and these directives have to be added to the language matrix.

(T2) follows on from the way the matrices are built and from the introduced notion of translatability. Whenever a given expression $A$ is to be a translation of some term $B$, both expressions have to figure in the same places in identical language matrices. Such a strict notion of translatability does not allow the translatable expression to differ syntactically. To see why it is so let us consider the opposite situation – let us say that we found two expressions which figure in exactly the same places of their respective language matrices, but one of them is atomic and the other is not. There would have to be a place in the second matrix where the second expression was decomposed into its atomic constituents but there would be no such place in the first matrix (because there was nothing to decompose there). But if the matrices are different then the expressions are by definition not translatable.

It is important to stress that neither of these claims present a serious challenge to the theory – they are simply counterintuitive. Nonetheless, I believe that it is worth pointing them out and analyzing ways of dealing with them because, as I hope to show, even small modifications to these claims produce interesting and useful variants of the theory.

In order to understand how we could deal with the thesis T1 we have to introduce an important requirement that Ajdukiewicz added to the theory. As he points out the directival theory can only be formulated for languages which are c o h e r e n t a n d c l o s e d.

A given language is c o h e r e n t if every expression it contains is connected to every other expression (directly or indirectly) *via* meaning directives[19].

In other words – if the language in question is coherent, we should be able to pick any expression and "reach" any other expression by "jumping" from a meaning directive to a meaning directive.

A language is c l o s e d if for every new expression, which is to be introduced to it, it already contains an expression synonymous with it.

In other words – a closed language is a language that already contains all meanings which can be added to this specific language (as further enrichment would have produced either synonyms or an incoherent language).

The bad news is that Ajdukiewicz's requirement creates bigger problems than the problem we wanted to solve with its help (thesis T1). The second requirement is simply much too strong – there are no existing closed languages and, what is worse, we could not create a closed language even if we wanted to (see Buszkowski 2010).

## 3. DIRECTIVAL THEORY OF MEANING RESURRECTED

In the remaining part of this paper I am going to show how we can utilize DTM as an interpretative tool for other theories of meaning – theories which often lack the precision of Ajdukiewicz's account and

[19] Two expressions are directly connected if they figure together in a single meaning directive. Expressions $A$ and $B$ are indirectly connected if they are not directly connected but there exists an expression $C$ such that $A$ and $B$ are directly connected to $C$.

which can be seen as sketches DTM fleshes out. What Ajdukiewicz's theory can provide here is showing something which other theories only hint at.

Let us start with a suggestion, which, I hope, will be rather obvious for the reader – the possibility of treating DTM as a theory of narrow content. Let us use the example of Sellars-Block's account[20] because the similarity between it and the DTM is striking. Sellars introduced four types of language rules, depending on whether the character of the stimulus provided for the user and her response is linguistic or not (Sellars 1963). There are three obvious possibilities:

1. Extra-linguistic stimulus – linguistic response.
2. Linguistic stimulus – linguistic response.
3. Linguistic stimulus – extra-linguistic response.

There is also a fourth, less obvious option:

4. Any stimulus – linguistic response[21].

It is not hard to see that 1. can be understood as empirical directives, 2. as inferential directives and 4. as axiomatic directives. There is nothing similar to 3. in the DTM but what prevents us from adding a new type of directive to the theory[22]? This new category of directives could be called imperative directives – they instruct the speaker to perform a certain action whenever she acknowledges a certain sentence by confirming it.

Now, the idea Block adds to the mix is that language described this way can be understood as a network of inputs and outputs which in turn enables us to define the narrow content of an expression (or its "conceptual role", as Block prefers to call it) as a role the expression plays in this computational structure (Block 1986). The problem with this account is that, while attractive, it does not show us how exactly a set of user actions (sentence confirmations) translates into a network of interrelated expressions of the language. Is the network just a set of beliefs connected by their inferential roles? If so, which ones – all of them? Maybe they should be decomposed somehow or perhaps even translated into language of thought? It is precisely what language matrices can help us with. They start with a set of pragmatic

[20] This account is neatly summarized in (Putnam 1991).

[21] Sellar calls this type of rule a "free rule".

[22] In fact adding new directive types is a very natural way of extending the theory and deserves further inquiry.

phenomena and then break it down into syntactic constituents of expressions enabling us to see the mechanism that underlies the phenomenon of narrow content.

Speaking of the language of thought – arguably the biggest flaw of this hypothesis is the elusiveness of the language it postulates. What does it look like? What is the ontological status of its expressions? What exactly are its meanings and how can they determine the meanings of natural languages? To see how the DTM could help here let us modify the idea of closed languages and introduce a more liberal (and realistic) notion of s e m a n t i c a l l y p r e d e t e r m i n e d languages.

A language is s e m a n t i c a l l y p r e d e t e r m i n e d if every new expression introduced to the language is synonymous with a compound expression built from the expressions the language already contains.

What we mean by that is that even if the language does not contain a proper synonym for the new expression, its meaning can be construed out of the language's existing expressions and this is exactly what Fodor assumes (Fodor 1975). The other thing we have to change is (T2) – we have to decide which syntaxes of translatable languages do not have to be identical. Instead, we assume only that the syntaxes are compatible in a sense that the differences they demonstrate are only superficial and what is important is the identity of deep syntactic structures of both languages[23]. It is possible that it is a solution Ajdukiewicz tacitly assumed anyway. Consider the way we build language matrices. What we look for in sentences are connectives, their arguments, their ordering and nothing else. Ajdukiewicz was a pioneer of categorial grammars so it is possible that he assumed that a working theory of meaning presumes a developed theory of universal grammar. In other words – categorial grammar could be thought of as a description of the surface grammar of two languages that is general enough so it abstracts away from unimportant details and enables us to represent two superficially syntactically different expressions as expressions of the same type. What we end up with, then, is a theory which fits the language of thought hypothesis quite well because it gives us the answer it lacked – it shows how the semantic structure of the language can be

[23] One notable complication is that the relation between a given language and the language we use to show its deep structure could not be explained by the same notion of translation we use in the DTM, but it is a small price to pay.

construed out of its non-semantic aspects. Moreover, it gives us the much needed model of linguistic structure which contains no actual labels or sentences but is still compatible with many different sets of such labels and predetermines the relations between them. The result is a detailed functional model for LOT. We can postpone the question of what the expressions of this language actually are. Instead we point at an expressionless language matrix (similar to the one presented in Table 3) and say only that LOT is anything that works "like that". As a functional semantics the DTM is compatible with different answers to the question about actual expressions. They can turn out to be patterns of firing neurons or parts of the brain or whatever else.

Another theory that could benefit from the DTM is Quine's behavioral theory of meaning. For the DTM to be useful here we would have to modify the requirement of coherency a bit. Let me digress for a second and say a few things about the notion of coherency I introduced earlier, because it proves to be even more useful than Ajdukiewicz had assumed. One disappointing aspect of the DTM I did not talk about is that although it provides the notions of translatability and synonymy, it does not give us any clue as to what it is for a given expression to simply "have a meaning" (as opposed to nonsense words). Does it suffice for an expression to simply be a part of a language matrix? This is where the notion of coherency can help: we can simply assume that an expression is meaningful if it is a part of a coherent language (which means that it is somehow connected to all the other expressions of the language). The problem with this idea is that it renders all the expressions meaningful. Consider the axiomatic directive of identity. The directive instructs the user to confirm every substitution of the formula $x = x$ regardless of the circumstances. What it means is that for every expression of the language there exists a meaning directive of the form $x = x$ where the expression is substituted for $x$. It follows that every expression is directly connected to the identity sign and the identity sign is directly connected to every other expression in the language. But it means that every expression is indirectly connected to any other expression.

Now, getting back to Quine's theory – instead of assuming, as we did above, that the meaningfulness of an expression depends on the number of connections to all the other expressions, we should focus only on selected connections, namely on the connections with the

non-linguistic, empirical parts of the table (that is the part specified in empirical directives, denoted by Greek letters in matrices). This way we could easily provide an explication for Quine's s t i m u l u s m e a n i n g. Note that by doing that we do not have to give up the non-semantic aspect of the theory because Quine's behavioral account does not imply that the expressions refer to stimuli.

If we allow for this modification of the DTM what we get in return is a theory which can be very well understood as a description of the manual constructed by Quine's radical translator. Remember that what the translator was supposed to do was to collect data on sentence confirmation. He collected the sentences which were confirmed in every situation, sentences which were confirmed after certain different sentences were confirmed and sentences which were confirmed whenever the empirical situation was such-and-such (Quine 2013). It is not hard to see that these three sets of data can be treated as our axiomatic, inferential and empirical meaning directives. Once again – the point here is that this convergence of theories goes far beyond a mere analogy. The DTM can be used to explain the idea which was originally rather vague – the idea of a translation manual (as created by a radical translator). The translation manual is a mapping of two different sets of expressions into a common language matrix.

Compatibility between the DTM and Quine's account is so great that we can easily recreate the infamous consequence of the latter theory, namely the indeterminacy of translation thesis. Let us get back to Tarski's counterexample. One way of looking at the problem it poses is that DTM allows for two expressions to be synonymous contrary to the beliefs of language users. If two expressions play the same role in a language (that is: figure in the same directives in the same places) they are synonymous no matter what. They are synonymous even if no one knows about it. They are synonymous even if the language users believe they are not synonymous! The latter happens when two expressions function the same way but one of the directives specifies that they are not identical (in these very words). A very well-known example of this is Putnam's elm/oak distinction (Putnam 1975, p. 226). If a given language[24] does not have tools to

[24] In Putnam's example they are, of course, idiolects. Although DTM has been originally construed as a theory of language and we retained this aspect of the

differentiate between two meanings, than an empty claim that they are different will not change anything. After all – the very claim still says the same thing about each expression. Let us call this peculiar type of synonymy t a c i t  s y n o n y m y.

To see how it generates the indeterminacy thesis consider two languages: $L_1$ and $L_2$. Let us say that they are translatable (in the specified sense) and that they both contain tacit synonyms: In $L_1$, $A_1$ and $B_1$ are synonymous and correspondingly in $L_2$, $A_2$ and $B_2$ are synonymous. Now the problem is that if you assume that $A_1$ is a translation of $A_2$, then, because $B_2$ is a synonym of $A_2$, $B_2$ is just as good a translation of $A_1$ as $A_2$ is. Of course we might rightfully say that it does not matter whether we translate $A_1$ to $A_2$ or to $B_2$ – all of these expressions are synonymous. The only thing we have to remember is that we keep the structure intact (so both languages still contain a pair of synonymous expressions). There is no "fact of the matter" as to which translation is better – both are just as good (unless you treat homonymy as a sufficient reason). But this is exactly how some of the researchers interpret the indeterminacy thesis (Field 2001, p. 282). Indeterminacy of translation does not present any genuine skeptical worry. The only thing it does is to subvert our expectations towards meaning.

Additionally, we have to realize that that the language we analyze might contain synonyms on various levels. It can have one-on-one mappings that we have just discussed but it could also contain surprising mappings where a single expression figures in the same set of places as a different, syntactically complex expression. Let us call it s y n t a c t i c a l  t a c i t  s y n o n y m y. This is something Ajdukiewicz did not foresee but there are no reasons as to why such a situation could not happen. For example, we could realize that for every sentence which figures in language directives and contains the word "a rabbit" there is an analogous sentence which contains a complex expression "an organized set of rabbit parts". The important difference between the current situation and the case of simple tacit synonymy we discussed in the preceding paragraph is that a syntactical tacit synonymy is much more bewildering to us than a normal tacit synonymy. We believe that

---

theory in this paper, it is worth pointing out that the theory can be easily tailored to function as a theory of idiolects.

there is a difference between rabbits and organized rabbit parts but try as we might, we cannot find directives to support this assumption. Tacit synonymy of syntactically different expressions can then be understood as the main mechanism of indeterminacy of translation.

What I wanted to show in this paper (apart from presenting a version of the DTM better suited for contemporary readers) is that the main reason why DTM had been abandoned by its creator and largely forgotten, namely because of the Tarski objection, is not very serious anymore, because we learned a few lessons along the way and no longer expect the theory of meaning to be also a theory of reference. The other problematic aspects of the DTM can be interpreted as challenges to be met by different, enhanced versions of the theory. Such enhanced versions can then be used as explications of existing ideas, like the notion of n a r r o w c o n t e n t (understood as a product of language rules), or as a framework for the behavioral theory of meaning or as a model for the language of thought hypothesis. The Directival Theory of Meaning is worth resurrecting because it can provide much needed details for propositions which are notoriously presented as sketches or outlines of possible future theories rather than as full-blown accounts.

## REFERENCES

Ajdukiewicz, K. (1934), "Sprache und Sinn", *Erkenntnis* 4(1), pp. 100–138.

Ajdukiewicz, K. (1978a), *Language and Meaning*, in: J. Giedymin (ed.), *The Scientific World-Perspective and Other Essays, 1931–1963*, Dordrecht: Springer, pp. 35–66.

Ajdukiewicz, K. (1978b), *On The Meaning of Expressions*, in: J. Giedymin (ed.), *The Scientific World-Perspective and Other Essays, 1931–1963*, Dordrecht: Springer, pp. 1–34.

Ajdukiewicz, K. (1978c), *The Problem of Empiricism and the Concept of Meaning*, in: J. Giedymin (ed.), *The Scientific World-Perspective and Other Essays, 1931–1963*, Dordrecht: Springer, pp. 306–319.

Ajdukiewicz, K. (1985a), *Język i znaczenie*, in: *Język i znaczenie*, vol. 1, Warszawa: PWN, pp. 145–173.

Ajdukiewicz, K. (1985b), *O znaczeniu wyrażeń*, in: *Język i znaczenie*, vol. 2, Warszawa: PWN, pp. 31–77.

Block, N. (1986), "Advertisement for a semantics for psychology", *Midwest Studies in Philosophy* 10(1), pp. 615–678.

Buszkowski, W. (2010), *O równoznaczności wyrażeń w ujęciu Ajdukiewicza*, in: J. Grad, J. Sójka, A. Zaporowski (eds.), *Nauka – Kultura – Społeczeństwo. Księga jubile-*

*uszowa dedykowana Profesor Krystynie Zamiarze*, Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Carnap, R. (1959), *Introduction to Semantics, and Formalization of Logic*, Cambridge: Harvard University Press.

Chomsky, N. (1995), "Language and Nature", *Mind* 104(413), pp. 1–61.

Field, H. (2001), *Truth and the Absence of Fact*, New York: Oxford University Press.

Fodor, J.A. (1975), *The Language of Thought*, New York: Crowell.

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

Hanusek, J. (2013), "On a Non-Referential Theory of Meaning for Simple Names Based on Ajdukiewicz's Theory of Meaning", *Logic and Logical Philosophy* 21(3), pp. 253–269.

Nowaczyk, A. (2006), *Dyrektywalna teoria znaczenia, czyli dramat Filozofa*, in: J. Pelc (ed.), *Sens, prawda, wartość: filozofia języka i nauki w dziełach Kazimierza Ajdukiewicza, Witolda Doroszewskiego, Tadeusza i Janiny Kotarbińskich, Romana Suszki, Władysława Tatarkiewicza*, Warszawa: Biblioteka Myśli Semiotycznej.

Plato (1998), *Cratylus*, Indianapolis: Hackett Pub. Co.

Putnam, H. (1975), *The Meaning of "Meaning"*, in: *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge, Mass.: Cambridge University Press, pp. 215–271.

Putnam, H. (1991), *Representation and reality*, Cambridge, Mass.: MIT Press.

Quine, W.V.O. (1969), *Ontological Relativity & Other Essays*, New York: Columbia University Press.

Quine, W.V.O. (2013), *Word and Object*, Cambridge, Mass.: MIT Press.

Sellars, W. (1963), *Science, Perception and Reality*, London; New York: Routledge and Kegan Paul; Humanities Press.

Wittgenstein, L. (1967), *Philosophical Investigations*, Oxford: Blackwell.