Conceptual Role Semantics Meets Default Logic

New Directival Theory of Meaning*

Paweł Grabarczyk[1,2] Michał Zawidzki[1]

pagrab@gmail.com zawidzki@filozof.uni.lodz.pl

[1]Institute of Philosophy, University of Łódź, Poland
[2]CCGR, IT University of Copenhagen, Denmark

In this note we present the Directival Theory of Meaning – a first semi formal theory of meaning introduced by a Polish logician Kazimierz Ajdukiewicz – and we show how it can be furthered developed by means of default logic.

1 Directival Theory of Meaning

The directival theory of meaning (which we are going from now on often call simply "the DTM") was a theory of meaning developed originally by Kazimierz Ajdukiewicz, a Polish logician and philosopher known especially for his seminal contribution to categorial grammars [1]. The theory was originally presented in two papers published in the 1930s [3, 2]. Both papers have been translated to English and published in [4]. The DTM has been a notable achievement for several reasons. It was the rst formal (or rather semi-formal) theory of meaning [5], it was the first theory of meaning that connected the notion of meaning with language use (20 years before Wittgenstein). Most importantly for us, it was the first conceptual (or functional) role semantics, that is semantics which explains meaning of expressions as the role they play in language.

Let us start with a quick, general description of some of the key features of the DTM. First of all, it is important to learn that the theory can be described as "semantics" only in a traditional, pre-Tarskian way. It is a theory of meanings understood as Fregean "senses", rather than a theory of reference. In fact, the theory is purely non-referential – it uses neither the notion of reference, nor the notion of truth. As we will see later, this is one of the reasons why default logic seems to be a natural fit for the DTM.

Second of all, the DTM construes meanings as environmentally narrow (i.e., not dependent on external grounding) but socially wide (i.e., dependent on the usage in a given community). As we are going to see in Sect. 5, the theory is very flexible as it can be interpreted as a theory of idiolect, sociolect or the whole language, yet not as a theory of private language.

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

The easiest way to understand the DTM is to start with a common observation about the way people react to substantial disagreements. When people use words as intended, communication is fluent; of course, it may be more or less fluent as people argue and disagree, but from our point of view it is interesting that sometimes communication comes to a complete halt. This means that sometimes an argument between two sides reaches a point where the sides start to suspect that the disagreement is, in fact, verbal. What users do in these cases is that they try to detect the suspected verbal difference by asking a few key questions about the expression in question. If the interlocutor does not answer as expected, she starts to be treated as if she used the expression with a different meaning.

Ajdukiewicz generalized this common-sense observation and assumed that every expression in a language figures in some sentences that the linguistic community expects the user to accept if she is to be treated as a member. Rules, or prescriptions, which codify these expectations are what the theory takes its name from: we call them "meaning directives". The additional constraint that we should not forget, is that the expression has to figure in the directives in an "essential manner". What it means in practice is that it cannot function in this expression just as a substitution of a variable.

2 Meaning directives

The general scheme of a meaning directive can be presented in a contemporary "input-output" format, where "input" refers to the circumstances the speaker encounters and "output" refers to the reaction to these circumstances. Using this format we can discern between four types of directives (the original Ajdukiewiczian account contained only the first three of them).

Axiomatic directives, which are the simplest type, should be understood as prescriptions which tell users of the language to accept (or reject) a certain sentence in any circumstances. In other words, they do not specify any input but require only a specific output – acceptance (or rejection) of a sentence. A very simple, non-controversial example of an axiomatic directive can be obtained by any substitution of an identity statement $x = x$. A sentence "a table is a piece of furniture" can serve as a more complex, natural example of a directive of this type. It is important to understand that particular directives are always relative to a given language and to the community of users. The same sentence can function as a directive in one community but it can be treated as a contingent sentence in another community. Its status is fully dependent on how the community treats it. The difference between a violation of a directive that assumes unconditioned acceptance of a sentence and a simple rejection of a sentence (which can be interpreted as a difference in opinion) is that the violator is not treated seriously by the community. Contrary to the standard rejection of a commonly held opinion, the violator will not be treated as having a different (even outlandish) belief. She will be seen as speaking a different language (or using some of the expressions figuring in this sentence in a different meaning) than the rest of the community. Her rejection will be treated not only as a false or absurd claim, but as completely incomprehensible.

Inferential directives relate two sentences: an input sentence (it can be also a conjunction of many sentences) and an output sentence. The main idea behind this type of directives is that once the user accepted a given sentence (which the directive does not necessarily prescribe her to, this might have been the user's choice), she has to accept the output sentence. A model example of this type of directives is the rule of introducing conjuction.

The third type of directives, **empirical directives**, introduce an extra-linguistic parameter to the mix. The idea behind them is that once the user finds herself in some specific empirical circumstances which manifest themselves in a specific internal state of the user (for example a visual perception of a color), she has to accept the prescribed output sentence. For example, the user looking at a big patch of red should accept the sentence “This is red”.

Last but not least, **promotive directives** relate an acceptance of a sentence (input) with a motor action of the user (output). To use an example – a user who accepts the command “Stop!” is expected to suspend her movement and will be said not to understand the meaning of the expression “Stop!” if she fails to do so.

3 Lanuage Matrix

Assuming that we managed to build a corpus of sentences enclosed in meaning directives, we can now proceed to the main part of the theory in which it shows us how the meaning of all non-compound expressions of a language can be extracted from the directives. In order to do that we have to build a “language matrix”. The matrix contains all of the sentences enclosed in meaning directives. Every sentence which is 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 reach 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 as long as there is nothing more to decompose. This enables us to provide the following definition of meaning.

Definition 1. The meaning of a non-compound expression e in a language $\mathscr{L}$ is an ordered pair $\langle \mathcal{M}, \mathcal{P} \rangle$ consisting of the matrix $\mathcal{M}$ of $\mathscr{L}$ and the set $\mathcal{P}$ of places e occupies in this structure.

There are two possible misunderstandings that we should avoid. First of all, it is important to differentiate two claims:

1. That for every expression e there is at least one sentence S, such that the user has to accept the sentence S in some circumstances C.
2. That for every circumstance C and every expression e, there is a sentence S containing an expression e, such that the user has to either accept or reject S in C.

In the setting represented by (2), language functions as a script of sorts which tells users which linguistic action they should perform in any situation. Under this description, users function as linguistic automata and language rules determine their verbal behavior completely. It is thus crucial to remember that the DTM subscribes only to the claim (1).

Second of all, even though the DTM can be described as a holistic theory of meaning, it is only weakly holistic or a “molecularist” theory. The reason for it is that the set of sentences which is used to define the meaning of all non-compound expressions is only a finite subset of all syntactically proper sentences of a given language.

What both of these caveats mean in practice is that the DTM is a “prohibitory semantics”. According to the DTM the meaning of an expression boils down to the constraints on the usage. The theory claims that the only role meaning plays in language is that it shows the users the boundaries of language, meaning helps the users avoid the misuse of words. As long as the users stay within these boundaries, they are free to say whatever they want.

By now, the biggest aw of the theory should be rather obvious to the reader. Up to this point the theory was able to inform us only about the meaning of non-compound expressions. In order to move further we need to add compositionality to language. Unfortunately, the original version presented by Ajdukiewicz does not contain this part.

4 Towards the formalization of directives

As mentioned in Sect. 2, we dinstinguish four types of directives: axiomatic, inferential, empirical, promotive. Note that we can draw another demarcation line dividing the four above-mentioned types into two categories: **stable** and **productive** directives. Indeed, axiomatic, empirical and promotive directives impose some rigid restrictions on language users, i.e., they oblige them to accept or reject an expression (axiomatic and empirical directives), or to act in a particular way (promotive directives), when they encounter this expression in given circumstances, and as such they do not expand the original set of directives. On the other hand, inferential directives force a language user to perform an action (of, e.g., accepting or rejecting a sentence) provided that another action was performed by them. In this sense they can produce new directives and thus make the original set expand.

It seems quite natural to interpret axiomatic and empirical directives as binary relations linking circumstances and expressions. In the same vein, promotive directives are simply binary relations between expressions and actions.

It is important to stress the difference between (1) a (stable) directive which forces a language user to perform an action in certain circumstances, and (2) an actual action taken by the user in given conditions. To avoid obscurity, let's consider an example. Assume that the theory of meaning for our language contains the following empirical directive: "When you see a red dress, accept the expression «Red!»." It can be represented as a binary relation of acceptance between the situation described as "seeing a red dress" and the expression "Red!". In (1) we deal with a binary relation that expresses obligation, in (2) – a binary relation that captures an actual action taken by a language user.

On the contrary, inference directives can be conceived as inference rules whose premises and conclusions comprise relations described in the previous paragraphs. As such, they are objects of a different form than relational formulas representing other kinds of directives. If X is a language user, she simply applies the inference rule representing an inference directive. The following implication represents an exemplary inferential rule: "If X accepted a sentence φ in a circumstance C, she has to reject a sentence ψ in a circumstance C'". This means that if the relation of (factual) acceptance holds between φ and C, then the complement of this relation holds between ψ and C'.

Among inferential directives we distinguish those that need to be applied **unconditionally** and those which ought to be applied only if it does not lead to a contradiction (**material**). The first kind contains all **logical** inferential directives, i.e., those which pertain to **logical constants** of a logic underpinning our language[1]

5 DTM+Default Logic

To formalize our set of directives we will use the **many-sorted language of first-order logic** which caters a good balance between expressive power and facility of use.

Let X be a metavariable denoting a language user. We introduce the following relational constants representing binary relations: A (for obligatory acceptance), R (for obligatory rejection), and P (for obligatory performance of an action). Additionally, for each language user

[1]We realize that the assumption that there exists a logical system "underpinning" a language is a bit sloppy (see, e.g., [7]). A formalization that we will present in the sequel of the paper does not, however, essentially rest upon this assumption, i.e., in the formal representation of a particular set of directives we can leave a set of logical inferential directives empty if we decide that in the language there exist no logical constants.

X, we introduce the relation of contingent acceptance, rejection and performance of an action, resp., A_X, R_X, and P_X. We distinguish three sets of individual variables and individual constants: $\mathcal{E}_{\text{var}} = \{x_{\mathcal{E}}, y_{\mathcal{E}}, z_{\mathcal{E}}, \ldots\}$ (expression-variables), $\mathcal{E}_{\text{cons}} = \{e_1, e_2, e_3, \ldots\}$ (expression-constants) $\mathcal{C}_{\text{var}} = \{x_{\mathcal{C}}, y_{\mathcal{C}}, z_{\mathcal{C}}, \ldots\}$ (circumstance-variables), $\mathcal{C}_{\text{cons}} = \{C_1, C_2, C_3, \ldots\}$ (circumstance-constants) $\mathcal{A}_{\text{var}} = \{x_{\mathcal{A}}, y_{\mathcal{A}}, z_{\mathcal{A}}, \ldots\}$ (for action variables), and $\mathcal{A}_{\text{cons}} = \{A_1, A_2, A_3, \ldots\}$ (action constants). Thus, in our setting expressions of the language, and descriptions of circumstances and of actions to be performed become *first-order terms*. The language of our formalization also contains the negation connective: $\sim$ and parentheses as auxiliary symbols.

As one can easily guess, axiomatic directives after formalization take the following form: $\forall x_{\mathcal{C}} A(e_i, x_{\mathcal{C}})$, $\forall x_{\mathcal{C}} R(e_i, x_{\mathcal{C}})$, where $e_i \in \mathcal{E}_{\text{cons}}$. Analogically, empirical (1 and 2) and promotive (3) directives will be formalized as follows: (1) $A(e_i, C_j)$, (2) $R(e_i, c_j)$, (3) $P(e_i, A_j)$, where $e_i \in \mathcal{E}_{\text{cons}}$, $C_j \in \mathcal{C}_{\text{cons}}$, and $A_j \in \mathcal{A}_{\text{cons}}$.

As mentioned before, inference directives are formalized as inference rules. Logical rules are represented as ordinary inference rules. For the sake of example, let's assume that in our language only two connectives have the status of logical constants, namely: negation ($\neg$) and conjunction ($\wedge$). Then the rule schemes expressing the appropriate inference directives containing these connectives could look as follows:

$$(\neg\gamma)\frac{\gamma(\neg e_i, C_j)^2}{\sim\gamma(e_i, C_k)} \qquad (\alpha\wedge)\frac{\alpha(e_i, C_k), \alpha(e_j, C_l)}{A(e_i \wedge e_j, C_m)} \qquad (\wedge\alpha)\frac{\alpha(e_i \wedge e_j, C_m)}{\alpha(e_i, C_k), A(e_j, C_l)}$$

$$(\wedge\beta)\frac{\beta(e_i \wedge e_j, C_k)}{\beta(e_i, C_l) \mid \beta(e_j, C_m)} \qquad (\beta\wedge_1)\frac{\beta(e_i, C_j)}{\beta(e_i \wedge e_j, C_k)} \qquad (\beta\wedge_2)\frac{\beta(e_j, C_j)}{\beta(e_i \wedge e_j, C_k)},$$

where $e_i, e_j \in \mathcal{E}_{\text{cons}}$, $C_j, C_k, C_l, C_m \in \mathcal{C}_{\text{cons}}$, $\gamma \in \{A, R, A_X, R_X\}$, $\alpha \in \{A, \sim R, A_X, \sim R_X\}$, and $\beta \in \{\sim A, R, \sim A_X, R_X\}$. The set of unconditioned inference rules must be expanded by adding to the logical rules the following **structural rules**:

$$(\mathsf{OR}_1)\frac{A(e_i, C_j)}{A_X(e_i, C_j)} \quad (\mathsf{OR}_2)\frac{R(e_i, C_j)}{R_X(e_i, C_j)} \quad (\mathsf{OR}_3)\frac{P(e_i, C_j)}{P_X(e_i, A_j)} \quad (\mathsf{CR}_1)\frac{A(e_i, C_j)}{\sim R(e_i, C_j)} \quad (\mathsf{CR}_2)\frac{R(e_i, C_j)}{\sim A(e_i, C_j)},$$

where $e_i \in \mathcal{E}_{\text{cons}}$, $C_j \in \mathcal{C}_{\text{cons}}$, $A_j \in \mathcal{A}_{\text{cons}}$, and X is a metavariable denoting a language user. (OR_1), (OR_2), and (OR_3) (**directive obedience rules**) guarantee that directives are obeyed, whereas (CR_1) and (CR_2) can be called **consistency rules** since they assure that A and R are disjoint (but not necessarily exhaustive) as relations. It is important to realize that all e- and C-expressions in the above schemes are, in fact, variables ranging over the sets of respective constants. Thus, each substitution of these variables with concrete constants (remaining in accordance with the original directives) constitutes a separate rule.

All **material inferential directives** are modelled by means of **defaults**, i.e., rules of inference introduced within **default logic**. Default logic is a nonmonotonic logic introduced by Reiter in [8] (see also [6]). It consists of a set of defaults added to an underlying logic as additional, nonmonotonic rules of inference. A scheme of a default has the following form: $\frac{\alpha:\beta_1,\ldots,\beta_n}{\gamma}$, where α is called **prerequisite**, $\beta_1, \ldots, \beta_n$ are called **justifications** and γ is called **conclusion** of the rule. It can be informally read as: "If α holds and nothing will <u>ever</u> contradict either of $\beta_1, \ldots, \beta_n$, then γ holds."

Each material inferential directive, where α is the premise and β is the conclusion of the directive, will be formalized as a **normal default**, i.e., a default of the form: $\frac{\alpha:\beta}{\beta}$. For instance,

[2] Note that the rules for negation imply that in our example the bivalence assumption is suspended, i.e., that there can be expressions that are neither accepted nor rejected. Of course, we can easily restore bivalence by making those rules bidirectional.

if an inference directive takes the form: "For any circumstance C, if you (have to) accept a sentence «It's a human embryo.» in C, then you (have to) accept a sentece «It's a human being» in C ($\forall C \in \mathcal{C}_{\text{cons}}[A(\text{«It's a human embryo.»}, C) \to A(\text{«It's a human being.»}, C)]$), it will be formalized by the following default scheme:

$$\frac{A(\text{«It's a human embryo.»}, C) : A(\text{«It's a human being.»}, C)}{A(\text{«It's a human being.»}, C)}, \tag{†}$$

where $C \in \mathcal{C}_{\text{cons}}$.

If we want to model "the dynamics" of the set of directives, triggered by inference rules, we need to first determine the set of all axiomatic, empirical and promotive rules, as well as the sets of actual language behaviours. Then we alternatively "run" all unconditioned and material inference rules on our initial set and obtain a so-called **extension** as the limit of such a procedure. For a more formal definition of extension let D denote the set of all defaults and $\mathbf{Cn}$ – the consequence relation determined by all unconditioned inference rules. Then an extension is a **smallest** set $\mathcal{E}$ such that:

1. $\mathcal{E}_0 = W$.
2. $\mathcal{E}_{n+1} = \mathbf{Cn}(\mathcal{E}_n) \cup \left\{ \gamma \mid \frac{\alpha : \beta_1 \ldots, \beta_n}{\gamma} \in D \text{ and } \alpha \in \mathbf{Cn}(\mathcal{E}_n) \text{ and } \neg\beta_1, \ldots, \neg\beta_n \notin \mathcal{E} \right\}$.
3. $\mathcal{E} = \bigcup_{n=1}^{\infty} \mathcal{E}_n$.

In this setting multiple different extensions are allowed. If, for example, our set of defaults contains † together with the following default:

$$\frac{A(\text{«It doesn't have a fully mature nervous system.»}, C) : R(\text{«It's a human being.»}, C)}{R(\text{«It's a human being.»}, C)}, \tag{‡}$$

where $C \in \mathcal{C}_{\text{cons}}$, then our initial set of directives will yield (at least) two different extensions, one of which will comprise the empirical directive $A(\text{«It's a human being.»}, C)$, whereas the other one – the directive $R(\text{«It's a human being.»}, C)$.

Since the same initial set of directives can yield many different extensions, we can identify a(n extended) language matrix with the set of all directives occurring in the intersection of all yielded extensions. Directives that occur in subsequent extensions but are not elements of the intersection can possibly be interpreted as idiosyncrasies typical for idiolects, sociolects or different stages of the development of a language of interest. The last hypothesis, however, despite being a promising idea requires a more careful investigation which we intend to carry out in future research.

References

[1] K. Ajdukiewicz. Die syntaktische konnexitat. *Studia Philosophica*, 1:1–27, 1935.

[2] K. Ajdukiewicz. *Language and Meaning (1934)*, pages 35–66. In Giedymin [4], 1978.

[3] K. Ajdukiewicz. *On the Meaning of Expressions (1931)*, pages 1–34. In Giedymin [4], 1978.

[4] J. Giedymin, editor. *The Scientific World-Perspective and Other Essays, 1931–1963.* Springer Netherlands, Dordrecht, 1978.

[5] J. Hanusek. On a non-referential theory of meaning for simple names based on ajdukiewicz's theory of meaning. *Logic and Logical Philosophy*, 21(3):253–269.

[6] V. W. Marek and M. Truszczyński. *Nonmonotonic logic – context-dependent reasoning.* Artificial Intelligence. Springer, 1993.

[7] R. Montague. Universal grammar. *Theoria*, 36(3):373–398, 1970.

[8] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–137, 1980.