Inferential Semantics as Argumentative Dialogues

Davide Catta, Luc Pellissier, Christian Retoré

HAL Id: hal-02922646

https://hal.science/hal-02922646

Submitted on 26 Aug 2020

Inferential Semantics as Argumentative Dialogues

Davide Catta[1], Luc Pellissier[2], and Christian Retoré[1]

[1] LIRMM, Univ Montpellier, CNRS, Montpellier
`davide.catta@lirmm.fr, christian.retore@lirmm.fr`
[2] Partout, INRIA Saclay Île de France
`luc.pellissier@inria.fr`

Abstract. This paper is at the same time a first step towards an "implementation" of the inferentialist view of meaning and a first proposal for a logical structure which describes an argumentation. According to inferentialism the meaning of a statement lies in its argumentative use, its justifications, its refutations and more generally its deductive relation to other statements. In this first step we design a simple notion of argumentative dialogue. Such dialogues can be either carried in purely logical terms or in natural language. Indeed, a sentence can be mapped to logical formulas representing the possible meanings of the sentence, as implemented with some categorial parsers. We then present our version of dialogical logic, which we recently proved complete for first order classical logic. Next we explain, through examples, how argumentative dialogues can be modeled within our version of dialogical logic.Finally, we discuss how this framework can be extended to real argumentative dialogues, in particular with a proper treatment of axioms.

Keywords: proof theory, dialogical logic, argumentation, natural language semantics

1 Introduction

A problem with the standard view of both natural language semantics and of logical interpretations of formulas is that the models or possible worlds in which a sentence is true cannot be computed or even enumerated [16]. As far as pure logic is concerned there is an alternative view of meaning called *inferentialism* [6,7,17,5,4]. Although initially inferentialism took place within a constructivist view of logic [6], there is no necessary conceptual connection between accepting an inferentialist position and refusing classical logic as explained in [5].

Inferentialisms takes the inferential activity of agents to be the primary semantic notion rather than denotation or reference in a given situation. According to inferentialism, knowing the meaning of a sentence means being able to recognize what are the conditions allowing the sentence to be correctly asserted: roughly speaking being able to recognize what counts as a justification of the sentence. From this it follows that there is a very tight link between inferentialism and argumentation theory: the study of arguments – reasonings, proofs, and

clues which make it possible to demonstrate or support an affirmation, a fact or a proposition. Thus for an inferentialist the study of the "good" arguments for a proposition is the study of its semantics. A very restricted view of inferentialism is proof theoretical semantics (see e.g. [9]); for instance, the proof-theoretic semantics for intuitionistic logic defines the interpretation of a formula by the set of its proofs: while it is satisfactory for provable formulas, it leaves unprovable sentences without meaning, and so is clearly inappropriate for natural language semantics. Another reason for the initial connection between inferentialism and intuitionistic logic is that proof theoretic semantics only makes sens for constructive logics, including intuitonistic logic, and not for classical logic.

An argument in favor of a statement is often developed when a critical audience, real or imaginary, doubts the truth, or the plausibility of the proposition. In this case, in order to successfully assert the statement, a speaker or proponent of it must be capable of providing all the justifications that the audience is entitled to demand. Taking this idea seriously, an approximation of the meaning of a sentence in a given situation can be obtained by studying the *argumentative dialogues* that arise once the sentence is asserted in front of such a critical audience. The aim of this paper is to sketch a frame in which this study can be carried forward. We use the dialogical logic paradigm [8,12] as a starting point: dialogical logic analyzes the concept of validity of a formula A through the concept of *winning strategy* in a particular type of two-player game. This type of game is nothing more than an argumentative dialogue between a player, called Proponent, which affirms a certain formula A and another player, called Opponent (called "teacher" in our version of dialogical logic), which contends its affirmation. The argumentative dialogue starts by Proponent affirming a certain formula A. Opponent takes its turn and attacks the claim made by Proponent according to the logical form of A. Proponent can, depending on his previous assertion and on the form of the attack made by the opponent, either defend his previous claim or counter-attack. The debate evolves following this back-and-forth pattern. Proponent wins the debate if he has the last word. An important characteristic of dialogic logic is that the concept of validity of a formula does not depend on a notion of model external to the dialogic game, contrary to what happens in his cousin theory game theoretical semantic [11,3]. We propose a new variant of dialogical logic for first order logic, for which we were able to give a clear and simple proof of the following completeness result: a first order statement F is provable in classical logic (e.g. in Gentzen **LK**) if and only if Proponent has a winning strategy argumentative dialogues starting with the formula F.

The link to natural language semantics and to argumentation in natural language is crystal clear in type logical grammars. Such grammars compute the possible meanings of a sentence viewed as logical formulae, as in the `Grail` platform a wide-scale categorial parser which maps French sentences to logical formulas [14,13,15]. Because of this close and computable connection between sentences and logical formulae [16], we can extend the logical notion of argumentative dialogue into an argumentative dialogue starting with a natural language sen-

tence, and even to an argumentative dialogue with natural language questions and answers, instead of logical formulas.

Admittedly, our first step in relating natural language semantics and argumentation is limited. In particular, argumentative dialogues in real-life situations can end without any of the two players winning: indeed, dialogues are never rooted in pure logic, but the players also have *axioms*, i.e. word meaning and assumptions about the world. These assumptions, especially the second ones, are not necessarily shared. For this reason, we focus our approach on situations where the definitions and postulates are known and precise, like in arguments between teacher and pupils in maths lessons (thanks to the data provided by maths education specialists) or in a future project on trial minutes.

Among the axioms some correspond to beliefs or postulates and are subjective and dependent on the speaker, but other axioms implement word meaning, which is, in a first approximation, objective and independent of the speaker in the language community. These latter axioms are fundamental to language understanding, natural language semantics, and argumentation. This is the lexical component of natural language semantics, and it is not so easy to properly integrate this aspect into compositional semantics which produces logical formulas describing the logical structure of the sentence without connecting the various predicates in the formulas like $book(x)$ and $read(u, v)$. But by now there exist several computable type theoretical formalisms that properly handle lexical semantics [1,18] parts of the later one being integrated in the Grail platform [15]. Our argumentative dialogues already handle lexical meaning as formulas, on a par with the main sentence under discussion which is also viewed as a formula. Nevertheless, we plan to have a specific hence more accurate treatment of those common axioms describing lexical semantics.

Another expected extension of argumentative dialogues is to extend our dialogues and their completeness result to first order modal logic, whose expressive power is needed for interpreting, e.g. belief verbs.

2   Dialogical Logic

We present here our variant of dialogical logic that deals with first order formulas in negation normal form i.e. in which the symbol of negation only appears as a prefix of atomic formulas — every first order formula is classically equivalent to a formula in negation normal form. As negation amounts to swapping the two players, by adopting this choice we obtain dialogical games in which the role of the two participants never change during the course of the game: the Proponent always asserts formulas while the Opponent always attacks the formulas. By defining dialogic games in this way, we are able to obtain a simpler definition of the notion of strategy than, for example, that given by Felscher [8]. As we will see, a strategy is a tree of dialogic games with binary branching, in which a branching is the consequence of the assertion of a conjunction.

The dialogue games that we are going to define have a reminiscent flavor of Socratic dialogues. Two characters participate in the dialogue-games: the pupil

and the teacher. The pupil claims that a certain proposition is true and therefore asserts it. The task of the teacher is to guide the pupil in the discovery of a *justification* for the asserted proposition. The teacher is a logic expert and whenever the student asserts a proposition, the teacher, in a somewhat pedantic way, reacts by questioning it by the following scheme.

conjunction If the student claims that "A_1 and A_2" the teacher will say "suppose that I claim that non A_i is justified, could you show me that I'm irrational by continuing the debate on A_i?"

disjunction If the student claims that "A_1 or A_2" the teacher will say "Suppose that I claim instead that both non A_1 and not A_2 are justified, could you show me that I'm irrational by choosing one of either A_1 or A_2 and continuing our debate on the proposition of your choice?"

The teacher will have to concede that the pupil is able to justify the statement when the dubitative rational process that he has undergone, together with the answers offered by the student, force him to bet on asserting both a simple primitive proposition and its negation. Given the pupil's inexperience, the teacher is forgiving of him. If he realizes that the fact of *choosing* to assert a certain proposition A leads him to a deadlock, he can always retrace his step an choose to assert a different proposition B. This corresponds to the choice of a different subformula in an asserted disjunction or to the choice of a different term in an asserted existential formula. We now present the system more formally.

Formulas The set of first order terms and first order formulas are defined as usual. In our dialogical games each formula will be in negation normal form. In other words the negation symbol $\neg$ will appear only as a prefix of an atomic formula and the only binary connective appearing in the formula will be $\wedge$ and $\vee$. This is not a limitation since each formula is equivalent to at least one formula in negation normal form — although admittedly the dialogue on the negation normal form does not match the structure of the non negation normal form.

Moves Here we define the rules that permits to attack a formula and what counts as a defence against the attack. The symbol prefixed by ? in the central column of the table are called **attack symbols**

Assertion	Attack	Defence
$A_1 \wedge A_2$	$?\wedge_i$	A_i
$A_1 \vee A_2$	$?\vee$	A_i
$\forall x A$	$?\forall[y/x]$	$A(y)$
$\exists x A$	$?\exists$	$A(t)$

In the table $i \in \{1, 2\}$, x, y stand for arbitrary variables and t stands for an arbitrary term.

A **move** is a couple (i, s) where $i \in \mathbb{N} \cup \{\star\}$ and s is either a formula or an attack symbol. Given M_n one of the moves in the sequence of moves
$$M_0, M_1 \ldots M_j \ldots$$
we say that:

- M_n is an **assertion** if it is of the form (i, F) and we call F the formula asserted by the move, or the asserted formula;
- M_n is a question if is it of the form (i, s) where s is one of the attack symbol.

A question M_n of the form (i, s) is **justified** if $i < n$ and $M_i := (i', F)$ and s is an attack on F, and we say that M_i is the **enabler** of M_n. An assertion M_n of the form (i, F) is **justified** if $i < n$ and $M_i := (i', s)$ and M_i is a justified attack on $M_j := (j', F')$ where s attacks F' and F is a defence of s. M_i is the **enabler** of M_n

Games A **dialogical game** for a formula F is a sequence of moves

$$M_o, \ldots M_j \ldots$$

such that

1. M_o is $(\star, F)$ where F is a formula
2. $M_1 \ldots M_j \ldots$ is is a strictly alternated sequence of justified questions (odd index moves) and justified assertions (even index moves). Moreover the enabler of a question is the preceding assertion.
3. if M_n and M_n' asserts the same sub-formula occurrence of F and have the same enabler then $n = n'$
4. if the formula asserted by a move M_n is a literal then there is a question M_j $(j < n)$ such that M_j is of the form $(j - 1, s)$, the enabler of M_j is (k, F) , and the negation of the literal asserted by M_n is a direct subformula of F that is a defence against s.

Even numbered moves are called pupil moves or **P** moves while odd numbered moves are called teacher moves or **T** moves. A move M is **legal** for a dialogical game D if and only if the sequence D, M is a dialogical game. We say that a dialogical game D is won by the pupil when

1. D is finite and ends in a **P**-move
2. there is no **T**-move that is legal for D

It is easy to see that if D is won by **P** the last move of D asserts a literal.

Strategies We now define the concept of **strategy**. Strategies are the dialogical counterpart of formal derivations in a proof system. Informally speaking a strategy is something that tells the pupil how to react against any possible attack of the teacher. Formally, a strategy $\mathfrak{S}$ for a formula A is a tree of dialogical games that is rooted on $(\star, A)$ in which each **T** move has at most one **P** move as a child and where a **P** move has two children if and only if it is of the form $(j, A \wedge B)$. A strategy is **P**-winning if and only if each of its branches is a dialogical game won by **P**. We say that a formula is dialogically valid if and only if there is a winning strategy for the formula.

3   Meaning, Proofs and Argumentation

We recently proved completeness of our argumentative dialogues for first order classical logic:

> *A formula is provable if an only if any of its prenex negation normal forms admit a winning stragegy.*

Proof. (sketch) The difficult direction is from Strategy to Proof. To do so, we first compute a derivation in the sequent calculus $G3$ [19] from a winning strategy for a quantifier free formula F (*). We then show that a strategy for a prenex formula can be divided into two parts (1) a part dealing only with (head) quantifiers (2) a part dealing with quantifier free formulas, solved by (*). Regarding (1) the difficulty (already noticed in [8] in the simpler intuitonistic case) is that some winning strategies do not respect the eigen-variable condition enjoyed by correct proofs. So we first show that (1.1) each quantifier part of a winning strategy for a formula F can be mapped to a correct quantifier structure (a kind of proof net for the quantifiers), (1.2) each correct quantifier structure for a formula F can be mapped to the quantifier part π' (the part below the *mid-sequent* obtained from the mid-sequent theorem) of a normal derivation π of F.

Such a result means that we can consider our dialogical system as a proof system for classical logic.

Consequently we can define the semantics of a sentence as the set of the argumentative dialogues starting with this formula. This set may or not include a winning strategy for the proponent ——- a tree may be viewed as the set of the prefixes of its branches. So this is a kind of proof theoretical semantics (see e.g. [9]) where formulas are interpreted by the set of their proofs. But there is a major difference with proof theoretical semantics: even non provable formulas have a non trivial interpretation (as in in Ludics see e.g. [10]), which is satisfying, since for most formulas neither F nor $\neg F$ are provable.

This also has an important consequence for computational semantics of natural language. While there is no way to compute or enumerate the possible worlds in which a given sentence is true, one can start argumentative dialogues from any formula, just by following the dialogue rules, and this actually produces some kind of enumeration. It is too early to elaborate on this aspect of the project, but the prospects seem promising.

4   A Toy Example of Natural Language Argumentation

Let us consider the following argumentative dialogue between **P** and **T**:

P David is a murderer
T Can you justify you assertion?
P Indeed, David killed Paul
T I see.

In the first step of the dialogue the student asserts a proposition. At the professor's request to justify his assertion, the student asserts another proposition i.e. "David killed Paul". This seems to indicate that the professor *believes* that it is sufficient to be able to assert "David killed Paul" to justify the claim "David is a murderer". If we were to model this mini argumentative dialogue faithfully through dialogical logic we would be embarrassed. Once the first statement has been translated into a a logical formula, we found ourselves with a literal i.e., $\mathsf{Murderer}(d)$ where d is a first order constant representing the individual David. Now condition 4 of the dialogical game definition prevents the pupil from asserting a literal unless there is a preceding question triggering the assertion of the literal negation. Moreover we do not see the deductive link between the proposition "David killed Paul" and the proposition "David is a murderer".

As the definition of dialogical games ensures logical correctness, we want to model this argumentative dialogue without a single change in the definition of dialogical games. The dialogue seems to suggest that the professor believes in the truth of a proposition, or of a set of propositions Γ, which, together with the proposition "David killed Paul" logically imply the proposition "David is a murderer". i.e. that there is a winning strategy $\mathfrak{S}$ for the formula $\Gamma \wedge \mathsf{Killed}(d,p) \Rightarrow \mathsf{Murderer}(d)$. Suppose that *in this case* Γ is $\forall x(\mathsf{Murderer}(x) \Leftrightarrow \exists z(\mathsf{Killed}(x,z)))$.

Figure 1, where the blue dotted arrow points back from **P** moves to the **T** move that enables them, shows a winning strategy for negation normal form of the formula:

$$F \equiv \forall x(\mathsf{Murderer}(x) \Leftrightarrow \exists z(\mathsf{Killed}(x,z))) \wedge \mathsf{Killed}(d,p) \Rightarrow \mathsf{Murderer}(d).$$

In the previous dialogue it is reasonable to suppose that the implicit proposition that allows the deductive link between the first and second statements made by the student is $\forall x(\mathsf{Murderer}(x) \Leftrightarrow \exists z(\mathsf{Killed}(x,z)))$. However, if we were to model more complex argumentative dialogues, it would be much more difficult to trace the set of hidden hypotheses that create the deductive link between the asserted propositions. To make the process more realistic we can impose that Γ contains *definitions* of words of open classes like nouns, verbs, adjectives, adverbs, etc. Those definitions are stipulated to be components of the sentences which are uttered by the dialogue proponent. In other words in order to check whether the thesis of an argumentative dialogue is implied by the sentence uttered by the person **P** who is asked for a justification one could use the following general procedure.

- Translate each sentence uttered by **P** into a logical formula F obtaining a set of formulas Δ
- For each relational symbol R such that $R(\mathbf{t})$ is a subformula of some formula in Δ find a definition of the type $\forall x(R \Leftrightarrow B)$, obtaining a set of formulas Γ.
- Try to find a winning strategy for $\Gamma \wedge \Delta \Rightarrow A$ where A is the thesis of the dialogue.

It is reasonable to ask if there are *fragments of natural language* in which the meaning of words that belong to open classes is defined with sufficient precision.
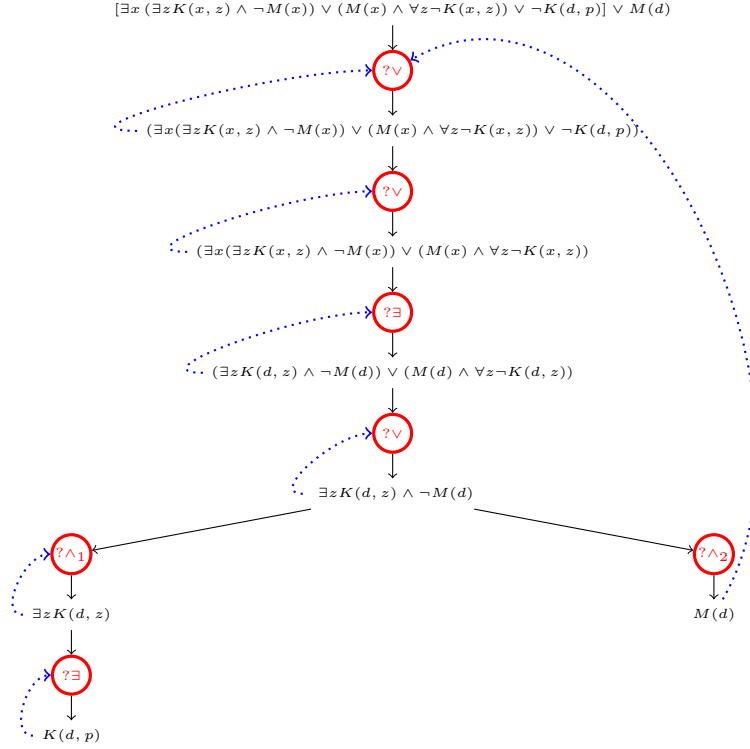
Fig. 1. A winning strategy for F. M stands for Murderer and K for Killed. Instead of integers we use blue dotted arrow that points back from **P** moves to the **T** move that enables them

Dialogical logic has already been applied in the area of mathematics pedagogy to model argumentative dialogues between pupils looking for a proof of a particular mathematical theorem (see e.g. [2]), and we think that another possible domain of application are (parts of) the minutes of a trial.

5   Conclusion and Perspective

In this first attempt to connect natural language semantics with argumentative dialogues, we proposed a notion of argumentative dialogues which enjoys a completeness theorem for first order classical logic. This notion can be used for defining the semantics of a sentences as the sets of all possible argumentative dialogues starting with this sentence.

When modelling natural language argumentation, we observe at least axioms representing lexical semantics are needed. Although those lexical axioms can be

treated as ordinary formulas as done in this paper, but it would be much more convenient to have a specific treatment of axioms. It would also open the door to a treatment of axioms that are not shared by the opponent and the proponent.

An even more interesting proposal would be to view unclosed dialogues as a way to trigger the discovery of axioms: what formula would close this dialogue? what other dialogue does this formula solves?

Regarding the purely logical aspects of our work, we would like to extend the transformation of a winning strategy to a proof to non necessarily prenex formulas (this is nearly done), to non necessary negation normal form formulas. Later on, we would like to encompass some standard modal logics (e.g. S4 which is proof theoretically well behaved) because modalities are needed when interpreting natural language sentences.

We also want to further explore the semantics in terms of sets of argumentative dialogues, both for formulas and for natural languages sentences — in this direction, a connection with ordinary models which have been more studied in linguistics would be much welcome.

References

1. Asher, N.: Lexical Meaning in context – a web of words. Cambridge University Press (2011). URL https://doi.org/10.1017/CBO9780511793936
2. Barrier, T.: L'élaboration des démonstrations mathématiques : une approche sémantique et dialogique. Recherches en Éducation **27**(27), 94–117 (2016)
3. van Benthem, J.: Logic in Games. MIT Press (2014)
4. Brandom, R.: Articulating Reasons: An Introduction to Inferentialism. Harvard University Press (2000)
5. Cozzo, C.: Meaning and Argument: A Theory of Meaning Centred on Immediate Argumental Role. Stockholm Studies in Philosophy. Almqvist & Wiksell International (1994)
6. Dummett, M.A.E.: What is a theory of meaning? In: S. Guttenplan (ed.) Mind and Language. Oxford University Press (1975)
7. Dummett, M.A.E.: The Logical Basis of Metaphysics. Harvard University Press (1991)
8. Felscher, W.: Dialogues as a foundation for intuitionistic logic. In: D.M. Gabbay, F. Guenthner (eds.) Handbook of Philosophical Logic, pp. 115–145. Springer Netherlands, Dordrecht (2002). DOI 10.1007/978-94-017-0458-8_2. URL http://dx.doi.org/10.1007/978-94-017-0458-8_2
9. Francez, N.: Proof Theoretical Semantics, *Studies in Logic*, vol. 57. College Publication (2015)
10. Girard, J.Y.: The blind spot – lectures on logic. European Mathematical Society (2011)

11. Hintikka, J.: Game-Theoretical Semantics as a Synthesis of Verificationist and Truth-Conditional Meaning Theories, pp. 250–273. Springer Netherlands, Dordrecht (1998). DOI 10.1007/978-94-017-2531-6_10. URL https://doi.org/10.1007/978-94-017-2531-6_10
12. Lorenzen, P., Lorenz, K.: Dialogische Logik. Wissenschaftliche Buchgesellschaft (1978). URL https://books.google.fr/books?id=pQ5sQgAACAAJ
13. Moot, R.: A type-logical treebank for french. Journal of Language Modelling **3**(1), 229–264 (2015). DOI http://dx.doi.org/10.15398/jlm.v3i1.92. URL http://dx.doi.org/10.15398/jlm.v3i1.92
14. Moot, R.: The Grail theorem prover: Type theory for syntax and semantics. In: S. Chatzikyriakidis, Z. Luo (eds.) Modern Perspectives in Type Theoretical Semantics, pp. 247–277. Springer (2017). DOI https://doi.org/10.1007/978-3-319-50422-3_10. URL https://doi.org/10.1007/978-3-319-50422-3_10
15. Moot, R.: The grail platform (syntactic and semantic parser) (2018). URL https://richardmoot.github.io/
16. Moot, R., Retoré, C.: Natural language semantics and computability. Journal of Logic, Language and Information **28**, 287–307 (2019). URL https://doi.org/10.1007/s10849-019-09290-7
17. Prawitz, D.: The epistemic significance of valid inference. Synthese **187**(3), 887–898 (2012). DOI 10.1007/s11229-011-9907-7. URL https://doi.org/10.1007/s11229-011-9907-7
18. Retoré, C.: The Montagovian Generative Lexicon ΛTy_n: a Type Theoretical Framework for Natural Language Semantics. In: R. Matthes, A. Schubert (eds.) 19th International Conference on Types for Proofs and Programs (TYPES 2013), *Leibniz International Proceedings in Informatics (LIPIcs)*, vol. 26, pp. 202–229. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2014). DOI 10.4230/LIPIcs.TYPES.2013.202. URL https://doi.org/10.4230/LIPIcs.TYPES.2013.202
19. Troelstra, A.S., Schwichtenberg, H.: Basic Proof Theory, 2 edn. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press (2000). DOI 10.1017/CBO9781139168717. URL https://doi.org/10.1017/CBO9781139168717