Abductive Reasoning, Interpretation and Collaborative Processes

by
Claudia Arrighi and Roberta Ferrario

1. Introduction

In this paper we are going to investigate the role of abductive reasoning in the interpretation of natural language, arguing that the similarities go deeper than the ones shown in other works, and that we can derive from this parallel some enrichment of the notion of abductive reasoning. In talking about the interpretation of natural language, we refer specifically to the work of Donald Davidson, which stresses the impressive human ability to reach mutual understanding in a conversation, despite the numerous ambiguities, malapropisms, unusual expressions that, in everyday communication, we put in almost every sentence. How is it possible that people who, at the beginning of a conversation go from misunderstanding to misunderstanding, at the end of that same conversation can reach an agreement of any sort?

Davidson's position is that this agreement cannot be reached by simply following the rules of language, but we need extralinguistic abilities, namely the ability of forming provisional hypotheses on the interpretation of the words uttered by other speakers, and the ability of revising these hypotheses through the interaction with other speakers and the environment. This idea of provisional hypotheses is at the core of the connection with abductive reasoning, as it is shown in (Wirth 1999). However, we think that there is more to be said about this connection, in particular regarding the role played by the *interaction* with other *speakers* and the *environment.* We consider this aspect fundamental, because it can provide a framework to overcome the weakness imputed to abductive reasoning. Hence, we propose a way to integrate this aspect in a schematization of abductive reasoning, borrowing a structured way to represent interaction between speakers from the work of Herbert Clark.

From a wider point of view, this work is meant to be a contribution to a current of thought developing at the intersection of philosophy and computing, what has been called a form of "socially sensitive computing" (Addis et al. 2004). Inside this approach, the contrast between human cognitive abilities and formal models is mediated developing the idea of "irrational sets" as product of abductive inferences, in various contexts of empirical and social interaction (Addis, Billinge and Visscher 2004, Addis et al. 2005). As a consequence of some of our precedent work on meaning (Arrighi and Ferrario 2004), we think that this idea of 'not-well-defined' sets is clearly

needed in the study of language and in natural language processing, as is shown, from a technical point of view, in other in works, such as Vadera et al. (2005).

The paper is structured as follows: Section 2 presents the features of the problem of mutual understanding, while in section 3 we sketch Davidson's ideas on the process of interpretation. In section 4 we recall the parallelism that Wirth traced between Davidson and Peirce, and section 5 gives a presentation of the collaborative and interactive aspects of language interpretation, what we thought was missing in Wirth's paper. In section 6, based on Clark's work, we try to embed abductive inference in a collaborative framework; finally, section 7 contains the conclusions of the paper.

2. The problem of mutual understanding

Long-lasting debates, in philosophy as well as in other disciplines, have shown that it is clearly problematic to give a complete account of how we, human beings, communicate through language. The difficulty is also evident when we look at all the concrete applications produced by researchers in computer science and artificial intelligence in the field of Natural Language Processing, in order to simulate human communication. Many of the issues are now well known and thoroughly studied, nevertheless they still pose very serious obstacles to the technical mastery of human skills in using language. Why is explaining how language works so problematic?

One of the main issues is that of *ambiguity*, in interpreting single words, in the structure of sentences, and different uses of an expression. This ambiguity is manifest in everyday experience, it is usually the source of most misunderstandings, but it is also the source of many jokes, as is shown in this quotation from Groucho Marx:

> *Outside of a dog, a book is a man's best friend.*
> *Inside of a dog it's too dark to read.*

In this example, the joke results from two different - but in any case both ordinary - uses, of the term "outside". Things can get even more complicated with so-called *non-literal interpretations*, when speakers make rhetorical use of language, as in metaphorical, ironical, or sarcastic utterances.

Another important source of problems is the role of *background* and *contextual information*: very often, in conversations a lot of elements that seem fundamental for a correct understanding of the message conveyed by a linguistic expression are left implicit, things like presuppositions, implicatures, chunks of common knowledge etc. For example, if someone asks us "Do you know the time?" it's likely that she would like to know what time it is, and not just if we are informed about

it or not. Nothing in her sentence literally asks for the time, but we simply know that this is what she *intends*. Let's give another simple real life example. Roberta calls Claudia in her office and Michael, her officemate, answers the phone:

Roberta: Hello, is Claudia there?
Michael: She is here, but not here
Roberta: Can I leave a message?
Michael: Shoot!

In this example, the answer "she is here but not here" seems to be informative for Roberta, instead of being perceived just as an overt contradiction, as suggested by a logical analysis of the sentence.

As a consequence of considerations like these, the debate on language has often turned around the question: Is the knowledge of words and rules of a certain language sufficient to explain how we understand that language? And if so, how many rules do we need, can they all be spelled out? One of the positions in the philosophy of language claims that speakers' mastery of a language cannot be represented by a set of words and rules of that same language. What is required is also an *extralinguistic* competence, i.e. a variety of skills that are not necessarily confined to the linguistic activity.

Amongst these extralinguistic skills which might be at the basis of language use, we are particularly interested in the ability of building plausible theories on the intentions of the other participants in the conversation. In other terms, each participant in the conversation forms some hypotheses aimed at giving a correct interpretation of the utterances of the others and the conversation is successful when such hypotheses fit sufficiently well what the interlocutors intended. Therefore, there is no exact criterion able to determine the success of a conversational interchange, only the participants can pragmatically guess if the information received was sufficiently similar to the one that was intended to be conveyed. Donald Davidson is one of the scholars who have suggested this view, and we want to spend few words in more details on his formulation.

3. Donald Davidson: interpreting as a process of formation and testing of contextual hypotheses

One of the central questions addressed by Donald Davidson in his work on philosophy of language is the so-called problem of radical *interpretation*, i.e. what makes it possible for two speakers of the

same language to share the meaning of a certain word or utterance? How do we know that it is the same meaning? Initially raised by Quine as a problem in dealing with different languages (as the problem of radical *translation*), it was soon extended to the broader question of how we know that some term means the same thing for two, or more, individuals.

Davidson absolutely rejects the common sensical explanation that we *share language,* arguing that the notion itself of "sharing a language" could not be defined inside the same language:

> *"[we] must give up the idea of a clearly defined shared structure which language users acquire and then apply to cases" (Davidson 1986).*

He argues instead for the necessity of a relation between linguistic meaning and extralinguistic context of an utterance, he strongly advocates a "contextualized" interpretation of language by means that are not part of the language itself. The importance of these extralinguistic means is particularly evident in observing what makes possible for the speakers to correctly react to ambiguous, unconventional or non codified uses of language and, at the same time, enables a creative or distorted use of language, as in metaphors or malapropisms.

In (Davidson 1986), Davidson gives an interesting example based on malapropisms in order to explain how understanding works in conversation. Davidson moves from the notion of first meaning, namely the first intended outcome of a speaker's utterance. The fact that often, in presence of a speaker who uses improperly some terms, a hearer appears nonetheless to grasp the first meaning comes as a surprise if we accept the theory that speaker and hearer share a common language where words mean – literally – something that as been *already fixed* and *predetermined*.

In the solution proposed by Davidson, the hearer comes to the conversation with a *prior theory* of how utterances should be interpreted; but during the conversation, if something doesn't fit this theory, she builds a *passing theory* to accommodate the discrepancy. This latter theory is built *ad hoc* to interpret this precise speaker in this precise conversation and can then be either abandoned or re-used in other conversations with the same speaker or with different interlocutors.

The ability to build a passing theory is anchored in human skills that enable us to retrieve an interpretation that makes sense of the speaker's utterance (in the case of Davidson's article, a similar sounding word has to be substituted to the one actually uttered by the speaker). These skills are something that we acquire and cannot be systematically described or taught.

4. Uwe Wirth: parallelism between Davidson's idea and Peirce's abductive inference

In his interesting paper (Wirth 1999), Uwe Wirth has drawn a parallel between Davidson's ideas on *interpretation* as formulation of theories and the process of *abductive inference* presented by C.S. Peirce in many of his works.

Abduction is one of Peirce's four steps of scientific investigation; it has been called a "logic of discovery", even though it is not an argument for validating scientific hypotheses. These four steps are anomaly observation, abduction of hypothesis, inductive testing, and deductive confirmation. The idea is that, after having observed an unpredicted phenomenon, a reasoner starts to look for an explanation; this explanation is reached by abduction and is only tentative. Based on this explanation, the reasoner - through a deduction process - forms a new prediction and then, finally, performs some inductive test in order to verify the prediction.

Peirce writes:

> *Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis.*
>
> *Deduction proves that something must be; Induction shows that something actually is operative; Abduction merely suggests that something may be.*
>
> *(Peirce, CP 5.181)*

The abductive inference is described by Peirce himself in the following way:

> *[1] The surprising fact, C, is observed;*
>
> *[2] But if A were true, C would be a matter of course.*
>
> Hence,
>
> *[3] There is reason to suspect that A is true*
>
> *(Peirce, CP 5.189)*

Where C is the surprising fact and A a set of provisional hypotheses.

Peirce's "surprising fact" is the occurrence of an unexpected event, which is inconsistent with the assumptions and thus generates surprise and calls for an "accommodation" of the theory. In other words, the inconsistency yields the search for some unknown premises that should have been taken into account but weren't. Once that these new premises have been found (abduced), they change the context in which the observation of the surprising fact has taken place, therefore eliminating the inconsistency.

As explained by Wirth, something very similar happens in Davidson's views on conversational activities. When two agents A and B are engaged in a conversation, they make

some tacit assumptions about each other and about how a communicative exchange should work. These tacit assumptions include, for example, Grice's well known imperatives, such as expecting the speaker to tell the truth, to be as informative as required, to respond in a way that is relevant to the topic of discussion. Obviously, it is not always the case that these requirements are met; if a sentence uttered by A violates one of these assumptions, the violation qualifies as "surprising event" for B (in Peircean terms). Instead of rejecting the sentence, B performs an abduction of a premise that could make sense of A's contribution, thus enriching her interpretation of a new, previously neglected, element. We can give an example of this going back to the conversation on the phone, in order to analyze it under the new perspective we have acquired thanks to the suggestions given by Wirth's paper. In that conversation, the sentence under focus is the following:

Michael: She is here, but not here

If we paraphrase Peirce, we have:

[1] Michael's sentence is a contradiction, i.e. a surprising fact, because it violates common principles of communication;

[2] But if "here" referred to two different things (the room and the building), Michael's sentence would be informative.

Hence,

[3] There is reason to suspect that "here" refers to two different things.

Instead of being stuck in the apparent contradiction, Roberta performs an abductive process of reasoning, from which she concludes that Michael intended to communicate that Claudia is not in the same room where he is, but she is in the building, and this is informative for Roberta. From this conclusion, Roberta goes on in the conversation with this new assumption/interpretation about Michael's intended message, so she now thinks that she can leave a message for Claudia to Michael, which will be delivered to her as soon as she comes back in the room. Here is where Wirth finds the analogy between the two approaches: generally speaking, both these processes are seen as the *"provisional adoption of an explanatory hypothesis"* (Peirce, CP 4.451), when an inconsistency is found in the theory at hand. Even if Davidson is addressing his analysis to language, he argues that in order to correctly interpret a language, competent speakers make use of a skill that is pre-linguistic, namely the ability to create a plausible theory about the speaker's intention.

Wirth suggests another common point between the two theories, the fact that, in his opinion, in both approaches the abductive competence is obtained by the interplay of an "economy of discourse" with the well known *principle of charity*, introduced by Neil Wilson (Wilson 1959, p. 532) and advocated by Davidson in interpreting language. The principle of charity roughly says that

when we interpret a sentence in an unintelligible way, it is more likely that our interpretation is a bad one, rather than a good interpretation of an unintelligible view. Thus the principle encourages to revise the interpretation in case of a mismatch with the expectations and the revision should be directed to the maximization of the economy; in other words the solution that requires the minimum cognitive, physical and economic effort should be taken.

This principle is certainly present in Davidson's works and is useful in order to trace the similitude with Peirce's abduction, but it shows also that the two approaches share the same weakness: we have no clear indication of where to find alternative, charitable, interpretations, or what should be the parameter to evaluate the 'cost', in terms of cognitive effort, of each of these alternative options. A theory of language interpretation relying on such a fuzzy principle can be criticized for being too vague, because it does not indicate where to look to find a measure of what is charitable or economically advantageous. This criticism seems to have much in common with the criticisms traditionally moved against abduction as an instrument to give an account of reasoning and discovery. As well explained in (Paavola 2004), abduction has been criticized under two perspectives: on the one hand, it is seen as too permissive, since an infinite number of hypotheses can be generated in every process of abduction and this is because it doesn't provide a criterion of relevance for such hypotheses. On the other hand, it is not generative since the missing hypothesis we are looking for (A) is already present in the premise 2 (But if A were true, C would be a matter of course), i.e. it is supposed to be known in advance, as opposed to be generated by the abductive process.

The paper by Wirth has definitely the merit of pointing out some of the main analogies in these particular aspects of the works of Davidson and Peirce, but unfortunately does not provide articulated suggestions on how to overcome this weakness. This is exactly what we would like to do, underlining the intersubjective nature of language and the importance of interaction with the environment as something that provides a framework of reference in formulating hypotheses – as presented in the next section – and then look again at the abductive inference, to see if something similar is applicable.

5. The collaborative and interactive aspect, a fundamental feature for language interpretation

We want to argue that in Wirth's parallelism something has been left out, that could be a promising source of constraints to the charitable efforts in interpreting, or, on the other side of the parallelism, a source of missing premises to be tested in an abduction process. There is something in Davidson's ideas on interpretation that can be integrated with abductive reasoning, and which

Wirth does not point out sharply enough, namely the importance of interaction with other speakers and with the entities in the environment, beyond the mere analysis of the linguistic expression.

When dealing with natural language and acquisition of words, abductive reasoning takes place in an interactive setting, situated in a precise environment, or empirical context, where an essential part of this context is constituted by the speaker's and hearer's viewpoints, and another fundamental part is given by the fact that speaker and hearer share a world.

In (Davidson 1982) Davidson says:

> *To understand the speech of another, I must be able to think of the same things she does; I must share her world.*
>
> *[…] Communication depends on each communicator having, and correctly thinking that the other has, the concept of a shared world, an intersubjective world.*
>
> *(Davidson 1982)*

Elsewhere, Davidson uses the metaphor of the triangle to show how the formation of a concept (for instance of a table or a bell) has its roots in the interplay of (at least) two human beings and a shared world:

> *"It takes two points of view to give a location to the cause of a thought, and thus to define its content. We may think of it as a form of triangulation: each of two people is reacting differentially to sensory stimuli streaming in from a certain direction. [...] If two people now note each other's reaction (in the case of language, verbal reactions), each can correlate these observed reactions with his or her stimuli from the world. A common cause has been determined. The triangle which gives content to thought and speech is complete. But it takes two to triangulate"*
>
> *(Davidson 1991)*
>
> *[…] to have the concept of a table or a bell is to recognize the existence of a triangle, one apex of which is oneself, the second apex another creature similar to oneself, and the third an object (table or bell) located in a space thus made common.*
>
> *(Davidson 1992)*

This interaction must be conceived as something much more complex than just reception of stimuli. As argued by Dagfinn Føllesdal in (Føllesdal 1995) talking about the theory of meaning in Davidson, language has to be rooted on the "perception of publicly observable physical objects and events", something that is subject to social assessment. Our response to a certain stimulus is due also to the *anticipations* and *expectations* surrounding the reception of such stimulus:

> *"we group the stimuli and responses in equivalence classes in view of our tentative theory of what the subject means and does"*

(Føllesdal 1982).

These quotations should give an idea of how central, in Davidson's conception of language, the interactive and situated characters are. In our opinion, following the lines of the parallelism traced by Wirth, these aspects of language are very relevant also with respect to the discussion on the weakness of abduction. They could represent the missing ingredients to give a fuller account of abductive reasoning, providing a relevance criterion for hypotheses selection, thus showing that they are not already known *ab initio* by the reasoning subject, but they are the product of an interactive process, both with the environment and the others speakers.

In the case of language, different individuals enter a conversation with different prior theories about the subject matter and the opinions of the other participants. These prior theories shape expectations about how the conversation will proceed; every time one of the expectations of an individual is not met (i.e. a surprising event happens), this individual is naturally brought to revise her prior theory and to adopt a new passing theory (that includes the new hypothesis that accommodates the surprising fact). The new hypothesis is not created or selected at random, but it strongly depends on what was going on in the previous part of the conversation, or in the activity shared with the others speakers; the goal of the activity, the reactions of the other participants and/or the features and inputs of the shared environment are what shapes the passing theory of the speaker, what determines the adoption of a certain hypothesis among all possible ones.

Can we apply a similar idea to abductive reasoning? The application of this framework to the study of abduction should not sound as completely new, since other analyses of abduction have already suggested that the usefulness of this tool in reasoning can be fully appreciated only if embedded in a framework of interaction, either with an environment or with other agents. For example, in a recent collection of papers on abductive reasoning in science, Magnani (Magnani 2004) has presented a *manipulative framework*, according to which "*concrete manipulations of external objects influence the generation of hypotheses*" (p. 243), and Paavola (Paavola 2004) has argued in favor of a *strategic framework*, that further develops an idea found in Hintikka:

> *"the validity of an abductive inference is to be judged by strategic principles rather than by definitory rules"*
>
> *(Hintikka 1998, p. 513)*

> *"strategies are related to goal-directed activity, where the ability to anticipate things, and to assess or choose between different possibilities, are important"*
>
> *(Paavola 2004, p. 270).*

These views summarize well the two aspects that we have underlined as fundamental. In formulating theories, new hypotheses are not floating in a vacuum, waiting to be picked up, but

they can be suggested by our interaction with the present environment and the other subjects involved; and because this interaction is most of the times goal-oriented, the selection of the hypotheses worth to be tested is guided by strategic considerations, with respect to the goal to reach. We find very significant that in other works Paavola and colleagues have developed, along with the approach to abduction based on strategy, also the idea of a "trialogue", very much in the spirit of Davidsonian idea of triangulation (Paavola, Hakkarainen and Sintonen 2005).

To explain better these ideas, we would like to show in a more concrete fashion how an abductive inference can be embedded in an interactive and goal-oriented process, and what kinds of elements provide constraints to the number of provisional hypotheses that can be adopted. In order to do this, in the next section we introduce a basic structure of a linguistic exchange as represented by the psychologist Herbert Clark.

6. Clark's rendition of the collaborative framework of mutual understanding

The psychologist Herbert Clark has provided an interesting formulation of the interactive and strategic character of language in his book *Using language* (Clark and Wilkes-Gibbs 1996). He describes linguistic communication as a *joint activity*, i.e. an activity carried out by a group of people acting in *coordination* with each other, in order to reach a *common goal*, that is the mutual understanding of each other's utterances (Clark and Wilkes-Gibbs 1986). Usually this joint activity is embedded in other joint activities of higher level, for example the conversation we are having is in the context of preparing dinner together - building a hierarchy of different levels of activity. Coordination and the goal-oriented nature are very much central points in Clark's representation of linguistic exchanges, and he goes deep in trying to identify as many elements as possible which the coordinating process can rely on. Amongst these elements we would like to pick out two that we consider particularly important, the *common ground* and the *common construal:*

- *Common ground:* the information that the speakers have in common
- *Common construal:* a model of the evolving present activity, including the identification of elements that are salient for this activity, mutual expectations on the roles played by every speaker and on the moves needed to reach the goal (mutual understanding)

The aim of the participants in a conversation (aside form the individual goals that led to the conversation itself) is to find an agreement on the meaning of what is said. In a very crude summarizing effort, we could say that *mutual understanding* is a process of verification of the common ground and of establishment of a common construal. This goal can be reached as it is in

many other joint activities: by acting together according to some intended principles, using salient objects or features in the environment and choosing our next move depending on which moves have already been made. Clark describes these steps in more details in his co-authored paper "Referring as a collaborative process" (Clark and Wilkes-Gibbs 1996, pp.20-24):

- *The steps:* Presentation, Evaluation, Rejection, Refashion, Acceptance, Follow-ups.

The best way to describe these steps is to give an example. The following is a slightly modified version of an example given in (Arrighi 2004), where the steps are used in a joint effort to verify a discrepancy in the assumed common ground, and to reach a common construal about the meaning of the word 'bread':

Dialogue	***Steps***
Sonal: "Can you bring the bread to the table, please?"	Sonal makes the presentation
Viola: "Bread? There isn't any bread"	Viola evaluates, then rejects. Sonal's move is not enough to reach the goal, to make herself understood
Sonal gives Viola a basket with naan bread.	Sonal refashions her presentation, with an ostensive act to make an object salient
Viola: "Oh, is this bread?"	Viola partially accepts, but asking for confirmation, a follow-up
Sonal: "Yes, it is naan bread"	Sonal refashions her ostensive act as an assertion
Viola: "I see... I didn't know"	Viola accepts

If we frame communication as a game, we can think of the speakers as players, guided by the goal of mutual understanding and constrained by many rules (the common ground), some about language, some about social interaction, some derived by physical constraints - exactly which rules apply depends on the stage of the game. In order to reach the goal, the players need to formulate a clear representation of the actual situation that is on the game-board (the common construal), and accordingly they think about a winning strategy, changing it as needed, depending on the moves of the other player.

The strategy is formulated with the final goal of the game in mind, and the elements of the context relevant to the game (for instance, an open door is not relevant to formulate a winning strategy in a chess game). When the other player makes a move that blocks our winning strategy,

abductive reasoning enters the picture: which other possibilities do I have to rebuild a winning strategy?

The way in which Clark presents the process of communication as a joint activity aimed at constructing a common ground for meaning seems to be a very accurate and plausible description of how things happen in linguistic communication. Now we would like to use Clark's schematization to compare the steps for reaching mutual understanding with the steps in the reasoning processes that characterize abduction.

7. Embedding abductive inference in a collaborative framework

Let's go back to the dialogue between Viola and Sonal, this time formulating the steps as an abductive inference. In this case we have a collaborative game with a common goal, the two players are supposed to reach mutual understanding – but nevertheless the moves of one player can be an obstacle for the other in some circumstances.

Dialogue	***Steps***
Sonal: "Can you bring the bread to the table, please?"	Viola`s assumptions include: Sonal is referring to something, the reference should be salient and available
Viola: "Bread? There isn't any bread"	Surprising fact: the reference is not available. Viola questions Sonal about this incongruence
Sonal gives Viola a basket with naan bread	Sonal performs an action, making an object salient. Given the general goal of their exchange, Viola has the assumption that Sonal's action must be relevant to the main goal
Viola: "Oh, is this bread?"	If this object was the intended reference, her request would be legitimate: Viola tests the new hypothesis
Sonal: "Yes, it is naan bread"	Sonal confirms the new common construal

In this example Viola is the player who needs to use abduction in order to reach the goal of interpreting Sonal's utterance. Given her assumptions about their common ground (for instance, they both speak English), and about the common construal (including that they are preparing dinner together), Viola understands that Sonal's utterance is a request, and if Viola wants to fulfill this request, she has to pick the reference of the word 'bread' and put it on the reference of the

word ‘table’. It is also assumed that Sonal wants to be understood by Viola, and therefore the references of the words must be available in the environment.

Here comes the surprising fact: there is no reference for the word ‘bread’. It is *surprising* because it is not consistent with the common ground or with the common construal that Viola thinks to be sharing with Sonal about how a conversation should work. But the interaction does not fail just because an inconsistency has been found – having a winning strategy in the game requires that difficulties must be overcome by making the proper move. In the present situation, the proper move is the simplest, acknowledging the problem and waiting for help: “Bread? There isn’t any bread”.

After the surprising fact, Viola does not stop formulating hypotheses on her own, but she chooses to interact with the environment, in this case to interact with Sonal. When Sonal responds with an action to Viola’s utterance, Viola assumes that Sonal’s action must be relevant to solve the “surprising fact” - the object shown must be the object Sonal intended. Viola, involved in a collaborative game, goal oriented and focusing on an effective winning strategy, does not stop considering all the plausible hypotheses, but just the one suggested in the present situation, immediately testing it: “Oh, is this bread?”

How does the join activity framework affect the selection of hypotheses? After Sonal’s request, is it possible that, looking around the kitchen, Viola sees a still nature on the wall featuring a loaf of bread. Can this have been taken as a plausible reference? Well, in many contexts it could, but in the present situation Viola judges this option not relevant for the common goal, for the joint activity in which this conversation is embedded. Viola is testing her assumption that Sonal is referring to a salient object in the room referred to by the word ‘bread’, but it must be salient in the context of the higher level assumption that they are engaged in the join activity of getting ready for a dinner. With her question, Viola expresses her difficulty with the assumption of lower level, but she still keeps the higher level assumption. This could be envisioned as a measure of the cognitive effort requested in revising the theory.

8. Conclusions

In this paper we have explored the common traits in interpreting language and in abductive reasoning. We started by considering a comparison between Davidson and Peirce put forward by Uwe Wirth, where an abduction process is compared to Davidson’s idea that linguistic interpretation relies on the speakers’ ability of building tentative hypotheses about the context and the intentions of the other speakers. These hypotheses can be revised when a new event does not fit the model built so far - but how do we choose new hypotheses?

The traditional weaknesses attributed to abduction have been summarized as the faults of being too *permissive* and *non generative*, and these faults seem to be the counterpart of the objections moved against the fuzziness of Davidson's ideas of the principle of charity and economy, in formulating new interpretations. Can these difficulties be overcome? We thought that we should look closer to some fundamental features of Davidson's account of interpretation, in particular, the interaction with both the environment and other speakers seems to be fundamental. Using a structured account of this interaction, as provided by the psychologist Herbert Clark, we have analyzed whether this 'structure' can be of any help in generating and selecting alternative hypotheses.

The first problem was permissivity: abduction doesn't provide a criterion to choose which assumptions were disregarded when formulating a failed explanation of a phenomenon. The "surprising event" tells us just that something is missing, but not *what* is missing, so any assumption can be taken in order to accommodate the anomaly. But, as happens in communication, the agent who performs abduction doesn't face the surprising event alone, she is embedded in an environment and in a social context that react to her attempts of accommodation, thus she can *negotiate* with the environment and with the other agents what was missing from her previous hypotheses. In other terms, the limits of the shared environment and social context, conceived as a common ground and a common construal, are a constraint for the range of hypotheses to evaluate.

The interaction with environment and social context also offers the solution to the other problem, namely non generativity. The problem emerges from the fact that in the classical formulation of abduction, the new hypothesis that should accommodate the surprising event seems to be forcedly known in advance by the subject who performs the abduction. But, again, this is consequence of the fact that we used to see the subject in isolation, while in real situations environment or the social context can *offer* new hypotheses, previously unknown to the reasoner. Moreover, the goal and the rules of the activity in which the reasoning is embedded drive in a certain direction the research for new hypotheses.

The final claim of the paper is then that an accurate account of the success in communication must be based on an integration of abductive with cooperative processes, and this is a direction of research definitely worthy of further exploration.

References

Addis, T., Billinge, D., Visscher, B. (2004). Socially Sensitive Computing: a Necessary Paradigm Shift for Computer Science. *Grand Challenges in Computing*, UK CPHC, Newcastle, 2004, pp. 1-6.

Addis, T., Townsend-Addis, J., Billinge, D., Gooding, D. C., Visscher B. (2005). Tracking Irrational Sets. *MBR 2004*, Pavia, Italy.

Arrighi, C., Ferrario, R. (2004). The Dynamic Nature of Meaning. *ECAP 2004*, Pavia, Italy. [year even if unpublished?]

Arrighi, C. (2004). *Meaning as a Product of Negotiation*. Master's Thesis, Philosophy Department, Stanford University.

Clark, H. H., Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22: 1-39.

Clark, H. H. (1996). *Using Language*. New York: Cambridge University Press.

Davidson, D. (1982). Rational Animals. *Dialectica* 36: 317-327.

Davidson, D. (1986). A Nice Derangement of Epitaphs. In R. Grandy and R. Warner (eds.) *Philosophical Grounds of Rationality*. Oxford University Press, pp. 433-446.

Davidson, D. (1991). Three Varieties of Knowledge. In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press, pp. 205-220.

Davidson, D. (1992). The Second Person. In *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press, pp. 107-122.

Føllesdal, D. (1982). Intentionality and Behaviorism. In L. J. Cohen, J. Los, H. Pfeiffer, and Klaus-Peter Podewski (eds.) *Proocedings of the 6th International Congress of Logic, Methodology and Philosophy of Science*, Hannover August 22-29, 1979. Amsterdam: North Holland.

Føllesdal, D. (1995). In What Sense Is Language Public? In P. Leonardi (Ed.), *On Quine: New Essays*. New York & Cambridge: Cambridge University Press.

Grice, H. P. (1969). Utterer's meaning and intentions. *The Philosophical Review,* 78: 147-177.

Hartshorne, C., Weiss P. (eds.) (1931-1958). *Peirce, Charles Sanders. Collected Papers*. Cambridge MA: Harvard University Press.

Hintikka, J. (1976). Language-Games. *Acta Philosophica Finnica,* XXVIII(1-3): 105-125.

Hintikka, J. (1998). What is Abduction? The Fundamental Problem of Contemporary Epistemology. *Transactions of the Charles S. Peirce Society*, 34(3): 503-533.

Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.

Magnani, L. (2004). Model-Based and Manipulative Abduction in Science. *Foundations of Science,* 9: 219-247.

Oltramari, A., Borgo S., et al. (2003). Negoziati di significato. *Sistemi Intelligenti*, 2003(3):419-428.

Paavola, S. (2004). Abduction as a Logic and Methodology of Discovery: the Importance of Strategy. *Foundations of Science*, 9(3): 267-283.

Paavola, S., Hakkarainen, K., Sintonen, M. (2005) Abduction with Dialogical and Trialogical Means, MBR 2004 Pavia, Italy.

Rapaport, W. J. (2003). What Did You Mean by That? Misunderstanding, Negotiation, and Syntactic Semantics. *Minds and Machines,* 13(3): 397-427.

Vadera, S., Rodriguez A., Succar E., Wu J. (2005). Using Wittgenstein's Family Resemblance Principle to Learn Exemplars. *MBR 2004*, Pavia, Italy.

Wilson, N. (1959). Substances without substrata. *Review of Metaphysics*, 12(4): 521-539.

Wirth, U. (1999). Abductive Reasoning in Peirce's and Davidson's Account of Interpretation. *Transactions of the Charles S. Peirce Society*, 35(1): 115-128.

Wittgenstein, L. (1958). *Philosophical Investigations*. New York: The Macmillan Company.