Can AI Models Capture Natural Language Argumentation?

Leila Amgoud Henri Prade

Institut de Recherche en Informatique de Toulouse
118, route de Narbonne
31062 Toulouse Cedex 4 France
amgoud@irit.fr, prade@irit.fr

Abstract

Formal AI models of argumentation define arguments as reasons that support claims (which may be beliefs, decisions, actions, ...). Such arguments may be attacked by other arguments. The main issue is then to identify the accepted ones. Several semantics were thus proposed for evaluating the arguments. Works in linguistics focus mainly on understanding the notion of argument, identifying its types, and describing different forms of counter-argumentation.

This paper advocates that such typologies are instrumental for capturing real argumentations. It shows that some of the forms cannot be handled properly by AI models. Finally, it shows that the use of square of oppositions (a very old logical device) illuminates the interrelations between the different forms of argumentation.

Keywords: Argumentation, Dialog, Non-Monotonic Reasoning.

Can AI Models Capture Natural Language Argumentation?

1 INTRODUCTION

Argumentation is a social activity of reason in which a proponent agent tries to convince an opponent one that a certain statement is true (or false) by putting forward *arguments*. While reasoning looks for the truth of a statement, argumentation looks only for persuading agents. Indeed, the proponent may succeed to persuade the opponent even if himself is not convinced by the statement.

Argumentation is an interdisciplinary topic. It is studied by philosophers like Hamblin [15], Rescher [25], Perelman and Olbrechts-Tyteca [19] and Toulmin [29]. Patterns of argumentation are studied in a pedagogical perspective for identifying fallacies in reasoning and avoiding them [10]. Argumentation also became an Artificial Intelligence keyword since early nineties. It is particularly used for non-monotonic reasoning (e.g. [13, 27]) and for modeling dialogues between agents (e.g. [2, 22]). See also [7, 8, 24] for descriptions of research on argumentation in AI. Whatever the application is, the same kind of argumentation model is considered. It consists of a set of arguments supporting statements and attacks among those arguments. Acceptability semantics are then used in order to evaluate the arguments and to decide on which statements to rely on. In all existing models, an argument has mainly three parts: a *conclusion*, a set of premises (called *support*) and a *link* between the support and the conclusion.

Besides, argumentation is largely studied by linguists like Salavastru [26] and Apothéloz [3, 4, 23]. The main focus here is on the notion of argument and its different types in real dialogues. In [3, 4], four argumentative types are defined. Two of them are arguments and two others are rejections of arguments. In addition, Apothéloz defined four modes of counter-argumentation. Each of them may be divided into at least two distinct cases.

Our aim in this paper is to analyze the typologies of arguments and the four modes of counter-argumentation proposed in [3, 4, 23], and to investigate whether they can be captured by the argumentation models developed in AI. Comparing research originating in the two communities (computer science and linguistics) is important since it allows a better understanding of work in both communities and may lead

to the development of richer models of argumentation.

The paper is organized as follows: We start by presenting and analyzing the notion of argument as defined by Apothéloz in [3]. In the definition, not only the reason and the conclusion of an argument are represented but also the functions of reason and conclusion are considered. We show how this may lead to four argumentative forms where only two of them are arguments. In a subsequent section, we present in detail the four modes of counter-argumentation proposed by Apothéloz in [3]. We analyze them through several examples. We show that the notion of a counter-argument in [3] takes into account the *intention* behind the counter-argument. The next section is devoted to AI formalizations of arguments and counter-arguments. It shows how arguments are defined using an underlying logic. In this paper, we do not focus on a particular logic. We assume a general and abstract logic in which negation is encoded. We show that the notion of argument is richer in linguistics than in AI. Then, we show that some of the modes of counter-argumentation cannot be handled properly by AI models. There are two reasons for that: The first one is due to the fact that in AI models, rejections of arguments are not modeled. The second reason is related to the fact that linguists encode intentions behind arguments when defining counter-arguments while this is not possible in AI models. Finally, we show that the use of square of oppositions (a very old logical device) illuminates the interrelations between the different forms of argumentation.

2 ARGUMENTATIVE FORMS IN LINGUISTICS

In [3], an argument is a pair $\mathcal{C}(x) : \mathcal{R}(y)$ where $\mathcal{C}$ is the *function of concluding* and x its content, $\mathcal{R}$ is the *function of reason* and y its content. The argument is read as follows: y is a reason for concluding x. We say that y is *argumentatively oriented* toward x. The contents x and y may either be premises (propositions) or arguments as we will see in the next section. Moreover, an argument is an *enthymeme*, i.e., an incomplete syllogism. Indeed, some generic rules relating y to x are left implicit. For instance, the argument "Mary will fail her exams (me) since she did not work hard (wh)" is written as $\mathcal{C}(me) : \mathcal{R}(\neg wh)$. Thus, the rule stating that "not working hard leads to failing exams" is not made explicit in the reason part of the argument. This is not surprising since linguists are concerned by natural language arguments, which are very often enthymemes (see [9, 12] for examples of works in AI on enthymemes).

In AI works on argumentation, the functions of conclusion and reason are implicit in the formal definition of an argument. However, we will see that mak-

ing explicit these functions is of great importance in 'natural language' counter-argumentation. Besides, the two contents x and y are formally defined. They are generally propositions, except in [16, 31] where they may be arguments. Finally, in AI models the link between x and y is defined (as we will see in Section 4) whereas in the work of Apothéloz, it is not.

Due to the presence of functions and contents, Apothéloz argues that there are two forms of negation: one for refuting a function and one for refuting its content. Refuting a function does not mean that its content is also refuted. The difference between the two negations is similar to the difference between $\vdash \neg p$ and $\nvdash p$ (where p is a propositional formula and $\vdash$ stands for the classical consequence relation). In [3, 4] both types of negation are denoted by $-$. These double negations give birth to four basic argumentative forms:

c_1	$\mathcal{C}(x) : \mathcal{R}(y)$	y is a reason for concluding x
c_2	$\mathcal{C}(x) : -\mathcal{R}(y)$	y is not a reason for concluding x
c_3	$-\mathcal{C}(x) : \mathcal{R}(y)$	y is a reason against concluding x
c_4	$-\mathcal{C}(x) : -\mathcal{R}(y)$	y is not a reason against concluding x

The contents x and y can themselves be replaced by their negation, leading to a combinatorics of 16 distinct argumentative forms, which includes $\mathcal{C}(-x) : \mathcal{R}(y)$ (y is a reason for concluding 'not x'), or $\mathcal{C}(x) : \mathcal{R}(-y)$ ('not y' is a reason for concluding x). It is worth noticing that only the forms c_1 and c_3 are arguments. The forms c_2 and c_4 are rejections of arguments. The form c_1 allows the representation of two epistemic states: one in which x is true and one in which x is false (i.e., $-x$ is true). However, the form c_3 encodes *ignorance* wrt. x. It expresses the fact that the conclusion x cannot be made but this does not mean neither that $-x$ is true. Let us illustrate the four forms by a dialogue between agents A, B, C and D.

A: Clara is at home (h). There is light from her window (l).

B: The fact that there is light from the window does not mean that she is at home.

C: But, she is on vacation! (v)

D: The fact that she is on vacation does not mean that she cannot be at home.

Agent A presents the argument $\mathcal{C}(h) : \mathcal{R}(l)$ which is of form c_1. Agent B rejects this argument. Note that B is not refuting l (i.e., he is not saying that there

is no light from Clara's window). He is neither saying that the conclusion h is false, but he is refuting the fact that l may play the function of reason in favor of h. This move is written as $\mathcal{C}(h) : -\mathcal{R}(l)$, that is of the form c_2. Apothéloz argued that this rejection aims at refuting $\mathcal{C}(h)$, thus it can be considered as an argument, $-\mathcal{C}(h) : \mathcal{R}(\mathcal{C}(h) : -\mathcal{R}(l))$, which is read as follows: the fact of rejecting the argument $\mathcal{C}(h) : \mathcal{R}(l)$ gives a reason for suspending the conclusion $\mathcal{C}(x)$. The agent C does not know whether Clara is at home or not, but thinks that he has a good reason for suspending the conclusion h. Indeed, since Clara is on vacation, then one cannot confirm that she is at home. The argument of C is encoded as $-\mathcal{C}(h) : \mathcal{R}(v)$, i.e., it has the form c_3. Note that the negation is on the function $\mathcal{C}$ and not on the content h since $\neg h$ would mean that C thinks that Clara is not at home while this is not the case. Agent D thinks that the fact that Clara is on vacation is not a sufficient reason for suspending the conclusion h. This move is then encoded as $-\mathcal{C}(h) : -\mathcal{R}(v)$.

3 COUNTER-ARGUMENTATION IN LINGUISTICS

Some linguists studied the different ways of defining a counter-argumentation, i.e., how to attack a given argument. A prominent work was done by Apothéloz [3]. Indeed, Apothéloz identified four modes of arguing against a given argument $\mathcal{C}(x) : \mathcal{R}(y)$:

1. Disputing the *plausibility* or the truth of the propositions used in y.
2. Disputing the *completeness* of the reason y. This is done by providing a new reason that is anti-oriented to the conclusion x, and that is presented as being more decisive than the reason y.
3. Disputing the *relevance* of the reason with respect to the conclusion x.
4. Disputing the *argumentative orientation* of the reason, by stating that the reason considered is rather in favor of $-x$, or is at least not in favor of x.

Throughout the paper, $\mathcal{K}$ stands either for $\mathcal{C}(-x)$ or for $-\mathcal{C}(x)$.

3.1 Disputing the Plausibility of a Reason (DPR)

Disputing the plausibility of the reason of an argument $\mathcal{C}(x) : \mathcal{R}(y)$ amounts to prove that y is false. Apothéloz argued that there are three ways for doing that:

1. By asserting an argument of the form $\mathcal{K} : \mathcal{R}(-y)$. In this case, no reason is given in favor of $-y$. Let us consider the following example.

a_1: Clara will miss her exams (me). She did not work hard ($-wh$).

a_2: Clara? She did not stop working!

The argument a_1 is written as $\mathcal{C}(me) : \mathcal{R}(-wh)$. The counter-argument a_2 intends blocking the conclusion me and is thus encoded as $-\mathcal{C}(me) : \mathcal{R}(wh)$. Recall that this does not mean that $-me$ is true or even supported.

2. By asserting an argument $\mathcal{K} : \mathcal{R}(\mathcal{C}(-y) : \mathcal{R}(z))$, that is by providing a reason against y as illustrated below.

a_3: No, she worked hard. Her eyes are encircled (ee).

Here, not only the premise $-wh$ is denied but it is also supported by a reason, that is $\mathcal{C}(wh) : \mathcal{R}(ee)$. This argument gives a reason for not concluding me, thus the following argument: $-\mathcal{C}(me) : \mathcal{R}(\mathcal{C}(wh) : \mathcal{R}(ee))$.

3. By asserting an argument of the form $\mathcal{C}(\mathcal{C}(x) : -\mathcal{R}(y)) : \mathcal{R}(-y)$. Here, the fact of denying y is considered as a reason for rejecting the whole argument $\mathcal{C}(x) : \mathcal{R}(y)$. This is illustrated by the following example:

a_4: Clara works hard (wh) because she is ambitious (am).

a_5: It is not by ambition that Clara works hard. She is not ambitious.

The argument a_4 is written as $\mathcal{C}(wh) : \mathcal{R}(am)$. The intention behind a_5 is not to suspend (or to deny) the conclusion wh as in the two previous cases. The agent providing this argument seems agree on wh but not on am. His intention then, is to reject the whole argument a_4. Thus, a_5 is defined as $\mathcal{C}(\mathcal{C}(wh) : -\mathcal{R}(am)) : \mathcal{R}(-am)$. Note that the conclusion of a_5 is a rejection of an argument.

To sum up, by denying the reason y of an argument $\mathcal{C}(x) : \mathcal{R}(y)$, one intends either blocking the conclusion x (cases 1 and 2) or rejecting the whole argument $\mathcal{C}(x) : \mathcal{R}(y)$ (case 3). Moreover, $-y$ may be supported or not by another reason.

3.2 Disputing the Completeness of a Reason (DCR)

Unlike the previous case where the reason y of an argument $\mathcal{C}(x) : \mathcal{R}(y)$ is false, here it is accepted but it is not sufficient to conclude x. This is due to the existence of a stronger argument which is anti-oriented toward the conclusion x. In [3], it is argued that this task can be achieved in two ways:

1. By asserting an argument of the form $\mathcal{K} : \mathcal{R}(z)$ where z is anti-oriented toward x. The following example illustrates this case:

a_1: Clara will miss her exams (me). She did not work hard (wh).

a_6: Clara will not miss her exams. She is very smart (sm).

Here the agent who uttered the argument a_6 may agree that the premise $-wh$ is true, but thinks that it is *not sufficient* to conclude me. Indeed, there is a *stronger* reason which prevents this conclusion. Thus, the argument a_6 is given as $\mathcal{C}(-me) : \mathcal{R}(sm)$. Let us consider now the following alternative reply to a_1 in the previous dialogue.

a_7: But Clara is very smart.

In this case, the agent does not know whether Clara will miss or not her exams but he provides an argument against concluding that she will miss them. Thus, a_7 is as follows: $-\mathcal{C}(me) : \mathcal{R}(sm)$. It is worth noticing that this example is similar to the following one provided in [21].

a_8: This object is red (or) since it looks red (lr).

a_9: But the object is illuminated by a red light (irl).

The argument a_8 is written as $\mathcal{C}(or) : \mathcal{R}(lr)$ while the argument a_9 is defined as $-\mathcal{C}(or) : \mathcal{R}(irl)$ and its roles is to prevent concluding or.

2. The second possibility is more tricky. It consists of giving a reason that is in favor of y but which is anti-oriented toward the conclusion x. The counter-argument has the form: $\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$. Let us illustrate this form of counter-argumentation by a simple example:

a_{10}: Paul is in his office (of) because his car is in the carpark (pa).

a_{11}: But the car is in the carpark because it is broken down (br).

According to the argument a_{10}, written as $\mathcal{C}(of) : \mathcal{R}(pa)$, the fact that Paul's car is in the carpark is a reason to think that Paul is still in his office. The reply a_{11} gives an explanation why the car is in the carpark: thus an argument $\mathcal{C}(pa) : \mathcal{R}(br)$. However, this explanation is anti-oriented toward the conclusion of, i.e., it blocks this conclusion. The argument a_{11} is defined as $-\mathcal{C}(of) : \mathcal{R}(\mathcal{C}(pa) : \mathcal{R}(br))$.

It is worth mentioning that in AI work on bipolar argumentation systems, namely the work [11], the authors consider the argument $\mathcal{C}(pa) : \mathcal{R}(br)$ as *supporting* the argument a_{10} (i.e., $\mathcal{C}(of) : \mathcal{R}(pa)$) since its conclusion is exactly a premise of a_{10}. Unfortunately, the previous dialogue shows clearly that this is not always the case.

3.3 Disputing the Relevance of a Reason (DRR)

The third way of attacking an argument $\mathcal{C}(x) : \mathcal{R}(y)$ is by disputing the relevance of the reason y with respect to the conclusion x. What is denied is neither x nor y but the fact that y may constitute a reason for x. This can be done in three ways:

1. By giving an argument of the form $\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$ showing that y is irrelevant for x. This is exactly the case of the previous dialogue where the fact that the car is broken down explains why the car being in a carpark is not a reason for concluding that Paul is in his office. Note that in this case it is both a matter of irrelevance and incompleteness of the reason.

2. By blocking the conclusion x via a rejection of the argument as follows: $-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$. Let us illustrate this case by considering the argument a_1 and with the reply a_{12}.

a_1: Clara will miss her exams (me). She did not work hard ($-wh$).

a_{12}: Indeed, she did not work hard, but not working hard is not a reason to necessarily miss her exams.

The intention behind such an argument is clearly to suspend the conclusion me by rejecting the fact that $-wh$ may play the role of a reason in favor of me. Note that in this reply, it is admitted that Clara does not work hard (i.e., the reason y is true).

3. By rejecting the argument, i.e., by uttering $\mathcal{C}(x) : -\mathcal{R}(y)$. An example would be:

a_{13}: She will not miss her exams because she did not work hard, but rather because of the stress (st).

In this example both x and y are recognized as true, but y is not the real reason for x being true. The real reason is st, that is $\mathcal{C}(me) : \mathcal{R}(st)$. Note that $\mathcal{C}(me) : \mathcal{R}(st)$ alone does not express the fact that the first argument is attacked or rejected. The rejection is expressed by $\mathcal{C}(me) : -\mathcal{R}(-wh)$.

3.4 Disputing the Argumentative Orientation of a Reason (DOR)

The fourth mode of counter-argumentation in [3] consists of disputing the argumentative orientation of the reason. The idea is that the reason y is not in favor of the conclusion x as stated in the argument $\mathcal{C}(x) : \mathcal{R}(y)$ but in favor of the opposite conclusion, that is $\mathcal{C}(-x) : \mathcal{R}(y)$. Let us illustrate this idea by the following example borrowed from [5].

a_{14}: 'A World Apart' is not a good film ($-gf$). It does not teach us anything new about apartheid ($-ta$).

a_{15}: That's precisely what makes it good.

The argument a_{14}, written as $\mathcal{C}(-gf) : \mathcal{R}(-ta)$, supports $-gf$ with the premise $-ta$. The counter-argument a_{15}, $\mathcal{C}(gf) : \mathcal{R}(-ta)$, supports the opposite conclusion with the same premise.

4 ARGUMENTATIVE FORMS IN AI

In the previous section, we have shown how arguments are defined by linguists. The definition is semi-formal since the link between the support and the conclusion is not specified, and the properties of the two functions are not clear. From the multiple examples given in [3], it seems that arguments are enthymemes. Thus, the content of the reason function leaves generic rules aside. For instance, the argument stating that Clara will miss her exams since she did not work hard ($\mathcal{C}(me) : \mathcal{R}(-wh)$) is based on an implicit generic rule which is 'not working hard leads to missing exams'. Finally, it is worth mentioning that Apothéloz did not study how arguments are evaluated, i.e., among the conflicting arguments, which ones win a given dialog.

Besides, in AI research has focused on formalizing nonmonotonic reasoning. Thus, various argumentation systems were developed for that purpose (see for instance [1, 13, 27]). The definition of an argumentation system follows several steps: constructing the arguments from a given knowledge base, identifying the attacks among them, evaluating the arguments and finally deciding which formulas to infer from the knowledge base on the basis of the accepted arguments. Thus, unlike in linguistics, the evaluation of arguments is largely studied in AI. See [6] for a description of existing semantics.

In this section, we focus only on the two first steps of an argumentation process since they correspond to the ones studied by Apothéloz. We show the type of

arguments that can be modeled, and analyze how to encode the different modes of counter-argumentation defined in [3].

Throughout this section, we assume a logical language $\mathbb{L}$ in which two sets are distinguished: a set $\mathbb{F}$ of *facts* and a set $\mathbb{R}$ of *generic rules*. Facts concern particular instances, like 'Tweety is a bird', whereas generic rules concern classes of instances, like 'Generally birds fly'. This distinction is important for recovering some of the previous modes of counter-argumentation. Apart from this distinction, the only requirement that is imposed on $\mathbb{L}$ is that it contains a connector of negation, denoted by $-$. Let $\mathtt{CN}$ be a consequence operator, that is $\mathtt{CN} : 2^{\mathbb{L}} \rightarrow 2^{\mathbb{L}}$. It is assumed to be monotonic and for some $x \in \mathbb{L}$, $\mathtt{CN}(\{x\}) = \mathbb{L}$. Finally, from a logic $(\mathbb{L}, \mathtt{CN})$, a notion of *consistency* is defined as in [28], that is a set $X \subseteq \mathbb{L}$ is consistent iff $\mathtt{CN}(X) \neq \mathbb{L}$. Propositional logic is used in some places only to illustrate issues. An argument is defined as follows:

Definition 1 (Argument) *An* argument *is a pair* (x, y) *s.t.*

- $y \subseteq \mathbb{L}$
- y *is consistent*
- $x \in \mathtt{CN}(y)$
- $\nexists y' \subset y$ *s.t.* $x \in \mathtt{CN}(y')$

x *is the* conclusion *of the argument whereas* y *is its* reason/support.

In this definition, the function of reason and that of conclusion are not explicit. However, their contents are clearly defined. These contents cannot be arguments, thus arguments of the forms $\mathcal{K} : \mathcal{R}(\mathcal{C}(-y) : \mathcal{R}(z))$, or $\mathcal{C}(\mathcal{C}(x) : -\mathcal{R}(y)) : \mathcal{R}(-y)$ cannot be expressed in our formal setting. Another key difference with the definition of linguists is that arguments are not entymemes. Assume that $(\mathbb{L}, \mathtt{CN})$ is propositional logic, then the argument $a_1, \mathcal{C}(me) : \mathcal{R}(-wh)$, is written as follows in the previous definition: $(me, \{-wh, -wh \rightarrow me\})$. The generic rule $-wh \rightarrow me$ is left implicit in $\mathcal{C}(me) : \mathcal{R}(-wh)$. Finally, remember that Apothéloz defined four basic argumentative forms: $\mathcal{C}(x) : \mathcal{R}(y)$, $-\mathcal{C}(x) : \mathcal{R}(y)$, $\mathcal{C}(x) : -\mathcal{R}(y)$ and $-\mathcal{C}(x) : -\mathcal{R}(y)$. Only the two first ones are arguments and the two others are rejections of arguments. The above definition only captures one form of arguments: $\mathcal{C}(x) : \mathcal{R}(y)$. Indeed, it allows to provide a reason either for x or for $-x$, but it does not block conclusions, i.e., does not express *ignorance* wrt x. Thus, $-\mathcal{C}(x) : \mathcal{R}(y)$ cannot be expressed in Definition 1. Note that this drawback is shared by those argumentation systems that reason about arguments [16, 31], i.e., where arguments

may support other arguments. In AI work on argumentation, an argument is seen as a logical proof for a given formula x, thus one looks for a reason to conclude x or $-x$.

Let us now analyze how an argument (x, y) may be attacked. Four different ways are distinguished:

1. *By building a new argument in favor of the opposite conclusion*, i.e., $(-x, z)$. This relation is known as *rebuttal* in [14]. Indeed, an argument rebuts another iff they have opposite conclusions. Note that this form of counter-argumentation corresponds to the first way of disputing the completeness of a reason in [3]. Thus, the argument a_6 (written as $(-me, \{sm, sm \rightarrow -me\})$ under propositional logic) rebuts the argument a_1. This relation captures also the fourth mode of counter-argumentation, that is disputing the argumentative orientation of a reason. For instance, the arguments a_{14} and a_{15} are encoded respectively as $(-gf, \{-ta, -ta \rightarrow -gf\})$, $(gf, \{-ta, -ta \rightarrow gf\})$. Note that in this case, the disagreement comes from the generic rules. From the same information $-ta$, one of them leads to gf while the other concludes $-gf$. This situation may be more complicate. Imagine the two following arguments: $(x, \{y, y \rightarrow x\})$ and $(-x, \{y, y \rightarrow z, z \rightarrow -x\})$. From y and following different paths, the two arguments lead to opposite conclusions.

2. *By disputing a fact in the support* y. This amounts to build an argument (x', z) where x' is $-t$ and $t \in \mathbb{F} \cap y$. This relation is known in argumentation literature as *assumption attack* [14]. At a first glance, it seems to correspond exactly to disputing the plausibility of a reason in [3], especially since arguments are enthymemes in that work, thus the content of the reason is facts. However, this is not always the case. Indeed, since Definition 1 does not allow neither blocking conclusions nor supporting arguments, the intentions behind the three cases of disputing the plausibility of a reason cannot be encoded. Let us revisit the examples presented before. The two arguments a_1 and a_2 are encoded as follows: $a_1 = (me, \{-wh, -wh \rightarrow me\})$ and $a_2 = (wh, \{wh\})$ while in [3], $a_2 = -\mathcal{C}(me) : \mathcal{R}(wh)$. The reply a_3 is defined as $(wh, \{ee, ee \rightarrow wh\})$ while Apothéloz writes $-\mathcal{C}(me) : \mathcal{R}(\mathcal{C}(wh) : \mathcal{R}(ee))$. Finally, the two arguments a_4 and a_5 are defined respectively as: $(wh, \{am, am \rightarrow wh\})$, $(-am, \{-am\})$ while a_5 is written as $\mathcal{C}(\mathcal{C}(wh) : -\mathcal{R}(am)) : \mathcal{R}(-am)$ by Apothéloz.

3. *By disputing the applicability of a generic rule* t *in the support* y, i.e., $t \in y \cap \mathbb{R}$. The idea is that the rule t is true in general but not applicable in a certain situation. This relation, called *undercut*, was defined in [20, 21]. Several

cases discussed by Apothéloz fall into this relation. The first way of disputing the completeness of a reason can be captured by this relation. Indeed, the argument $a_7 = -\mathcal{C}(me) : \mathcal{R}(sm)$ is against applying the generic rule $-wh \rightarrow me$ when a person is smart (sm). The argument $a_9 = -\mathcal{C}(or) : \mathcal{R}(irl)$ aims at blocking the application of the rule ('when an object looks red the it is red' ($lr \rightarrow or$)) when the object is illuminated by a red light (irl). Similarly, the argument a_{11} blocks the applicability of the generic rule saying that if Paul's car is in the carpark, then Paul is in his office ($pa \rightarrow of$). It is important to notice that the phenomenon of blocking a generic rule raises in *default reasoning*. Indeed, a rule is blocked in presence of an *exception*.

4. *By disputing a generic rule*, that is by asserting that it is false. This is typically what happens in the second way of refuting the relevance of a reason. Let us consider the argument a_{12}. It says that just because Clara did not work hard is not a reason to miss her exams'. Here the agent recognizes that Clara did not work hard. So what is disputed is the plausibility of the rule $-wh \rightarrow me$. This is again captured by assumption attack which consists of undermining an element of the support of an argument.

The following table summarizes the four modes of attacking an argument $\mathcal{C}(x) : \mathcal{R}(y)$ as defined in [3] as well as the ways of capturing them in an AI model.

DPR1	$\mathcal{K} : \mathcal{R}(-y)$	Assumption attack on facts
DPR2	$\mathcal{K} : \mathcal{R}(\mathcal{C}(-y) : \mathcal{R}(z))$	Assumption attack on facts
DPR3	$\mathcal{C}(\mathcal{C}(x) : -\mathcal{R}(y)) : \mathcal{R}(-y)$	Assumption attack on facts
DCR1	$\mathcal{C}(-x) : \mathcal{R}(z)$	Rebut
DCR2	$-\mathcal{C}(x) : \mathcal{R}(z)$	Undercut
DCR3	$\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$	Undercut
DRR1	$\mathcal{K} : \mathcal{R}(\mathcal{C}(y) : \mathcal{R}(z))$	Undercut
DRR2	$-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$	Assumption attack on rules
DRR3	$\mathcal{C}(x) : -\mathcal{R}(y)$	?
DOR	$\mathcal{C}(-x) : \mathcal{R}(y)$	Rebut

The table shows that most of the modes of counter-argumentation are only partially modeled in our logical formalism. Indeed, the intention behind each attack is not captured. Moreover, at a formal level we do not make any difference between the four cases of applying assumption attack. Similar comment holds for undercut and rebut. While the differences may be crucial for evaluating arguments. Indeed, disputing a fact is not like disputing a generic rule and refuting a fact by providing a new reason is not like rejecting the fact without justification. Moreover, from

a dialogical point of view, it is important to be able to represent accurately the moves of the agents. In our formalism, the rejection of an argument (DRR3) is not possible while such a move is very common in dialogues.

5 ORGANIZING ARGUMENTATIVE STATEMENTS IN A SQUARE OF OPPOSITION

A key point in the categorization introduced by Apothéloz in [3] is the presence of two kinds of negation, one pertaining to the contents x or y, and the other to the functions $\mathcal{R}$ or $\mathcal{C}$. It has been observed that such a double system of negations gives birth to a formal logical structure called *square of opposition*, which dates back Aristotle's time (see, e.g., [18] for a historical and philosophical account). We first briefly recall what this object is, since it has been somewhat neglected in modern logic.

5.1 Classical Squares of Opposition

It has been noticed for a long time that a statement (A) of the form "every a is p" is negated by the statement (O) "some a is not p", while a statement like (E) "no a is p" is clearly in even stronger opposition to the first statement (A). These three statements, together with the negation of the last statement, namely (I) "some a is p", give birth to the square of opposition in terms of quantifiers $A : \forall a\ p(a)$, $E : \forall a\ \neg p(a)$, $I : \exists a\ p(a)$, $O : \exists a\ \neg p(a)$, pictured in Figure 1. Such a square is usually denoted by the letters A, I (affirmative half) and E, O (negative half). The names of the vertices comes from a traditional Latin reading: **A**ff**I**rmo, n**E**g**O**). Another standard example of the square of opposition is in terms of modalities: $A : \Box r$, $E : \Box \neg r$, $I : \Diamond r$, $O : \Diamond \neg r$. As can be seen from these two examples, different relations hold between the vertices, which gives birth to the following definition:

Definition 2 (Square of opposition) *Four statements A, E, O, I make a* square of opposition *if and only if the following relations hold:*

1. *A and O are the negation of each other, as well as E and I;*
2. *A entails I, and E entails O;*
3. *A and E cannot be true together, but may be false together, while*
4. *I and O cannot be false together, but may be true together.*

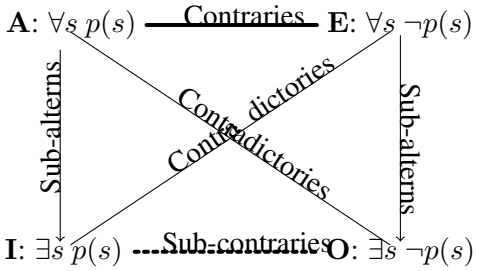

Figure 1: Square of opposition

Note that A entails I pressupposes in the example of Figure 1 that $\{s \mid p(s)\ is\ true\} \neq \emptyset$, otherwise A cannot entail I since there is no s. Similarly $r \not\equiv \bot$ is assumed in the modal logic case.

5.2 A Square of Opposition for Argumentation

The observation that two negations are at work in the argumentative statements classified by Apothéloz [3] has recently led Constantin Salavastru [26] to propose to organize the four basic statements into a square of opposition; see also [17]. However, his proposal may be discussed on one point, as we are going to see. Indeed, taking $\mathcal{C}(x) : \mathcal{R}(y)$ for vertex A, leads to take its negation $\mathcal{C}(x) : -\mathcal{R}(y)$ for O. Can we take $-\mathcal{C}(x) : \mathcal{R}(y)$ for E? This first supposes that A and E are mutually exclusive, which is clearly the case. Then, we have to take the negation of E for I, i.e., $-\mathcal{C}(x) : -\mathcal{R}(y)$. We have still to check that A entails I and E entails O, as well as condition (4) above. If y is a reason for not concluding x, then certainly y is not a reason for concluding x, so E entails O; similarly y is a reason for concluding x entails that y is not a reason for not concluding x, i.e., A entails I. Finally, y may be a reason neither for concluding x nor for not concluding x. This gives birth to the argumentative square of opposition of Figure 2. It can be checked that the contradiction relation (1) holds, as well as the relations (2), (3), and (4) of Definition 2.

Proposition 1 *The four argumentative forms $A = \mathcal{C}(x) : \mathcal{R}(y)$, $E = -\mathcal{C}(x) : \mathcal{R}(y)$, $O = \mathcal{C}(x) : -\mathcal{R}(y)$, $I = -\mathcal{C}(x) : -\mathcal{R}(y)$ make a square of opposition.*

Note that we should assume that $\mathcal{C}(x) : \mathcal{R}(y)$ is not self-contradictory (or self-attacking) in order that the square of opposition makes sense. In propositional logic, this would mean that $x \wedge y \neq \bot$.

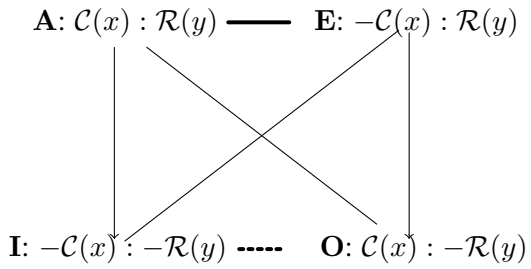

Figure 2: An argumentative square of opposition

This square departs from the one obtained by Salavastru in [26] where vertices A and I as well as E and O are exchanged: in other words the entailments (2) are put in the wrong way. This may come from a misunderstanding of the remark made in [3] that the rejection $\mathcal{C}(x) : -\mathcal{R}(y)$ is itself a reason for not concluding x, which can be written $-\mathcal{C}(x) : \mathcal{R}(\mathcal{C}(x) : -\mathcal{R}(y))$. But this does not mean that $\mathcal{C}(x) : -\mathcal{R}(y)$ entails $-\mathcal{C}(x) : \mathcal{R}(y)$ since it may be the case, for instance, that $\mathcal{C}(-x) : \mathcal{R}(y)$. Salavastru made another mistake regarding the link between A and I. He assumed that I entails A. Let us show through a simple example that this implication is false, but it is rather in the other way around.

a_{16}: The fact that Paul is a French citizen fr is not a reason to not conclude that he is smart st.

This is clearly a statement of form c_4, i.e., $-\mathcal{C}(sm) : -\mathcal{R}(fr)$. The question now is: does this statement entails the argument $\mathcal{C}(sm) : \mathcal{R}(fr)$ (i.e., the fact that Paul is french is a reason to conclude that he is smart)? The answer is certainly no. However, the converse is true. That is $\mathcal{C}(sm) : \mathcal{R}(fr)$ implies $-\mathcal{C}(sm) : -\mathcal{R}(fr)$.

6 CONCLUSION

This paper reported an interesting work by a linguist on argumentation theory, and analyzed it from an AI perspective. We have shown how Apothéloz defines the notion of argument by making explicit two functions: a function of conclusion and a function of reason. This allows also to have two types of negation: one for refuting a function and another one for disputing its content. These double negations give birth to four argumentative forms: two of which are arguments and two others are only rejections of arguments. We have shown through examples that the four forms are meaningful and very frequent in natural language dialogues. We

have then shown the four modes of counter-argumentation proposed by Apothéloz in [3]. Each mode can itself have various cases. We have then defined the notion of argument and counter-argument in a more formal way as it is done in AI. We have shown that the formal definition captures only one argumentative form among the four proposed by Apothéloz. As a side effect, the different modes of counter-argumentation cannot all be captured. Moreover, the ones which are captured are only encoded partially. The last contribution of this paper consists of showing that the proposal of Apothéloz makes sense since it obeys the properties of a square of opposition. Indeed, we have shown that the four argumentative forms constitute a square of opposition.

A future work would be to develop a rich argumentation system that captures the various modes of argumentation and counter-argumentation. Another idea consists of comparing the arguments schemes developed by Walton (in [30]) with the argumentative forms of Apothéloz.

References

[1] L. Amgoud and Ph. Besnard. Bridging the gap between abstract argumentation systems and logic. In *International Conference on Scalable Uncertainty Management (SUM'09)*, pages 12–27, 2009.

[2] L. Amgoud, Y. Dimopoulos, and P. Moraitis. A unified and general framework for argumentation-based negotiation. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07)*, pages 963–970. ACM Press, 2007.

[3] D. Apothéloz. Esquisse d'un catalogue des formes de la contre-argumentation. *Travaux du Centre de Recherches Sémiologiques*, 57:69–86, 1989.

[4] D. Apothéloz. The function of negation in argumentation. *Journal of Pragmatics*, pages 23–38, 1993.

[5] D. Apothéloz, P. Brandt, and G. Quiroz. Champ et effets de la négation argumentative : contre-argumentation et mise en cause. *Argumentation*, 6:99–113, 1992.

[6] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.

[7] T. Bench-Capon and P. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[8] Ph. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.

[9] E. Black and A. Hunter. A relevance-theoretic framework for constructing and deconstructing enthymemes. *Journal of Logic and Computation*, 22(1):55–78, 2012.

[10] P. Blackburn. *Logique de l'Argumentation*. Editions du Renouveau Pédagogique, Saint-Laurent, Québec, 1989.

[11] C. Cayrol and M. Lagasquie. On the acceptability of arguments in bipolar argumentation frameworks. In *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'2005)*, pages 378–389, 2005.

[12] F. Dupin de Saint-Cyr. Handling enthymemes in time-limited persuasion dialogs. In *5th International Conference on Scalable Uncertainty Management*, pages 149–162, 2011.

[13] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence Journal*, 77:321–357, 1995.

[14] M. Elvang-Goransson, J. Fox, and P. Krause. Dialectic reasoning with inconsistent information. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 114–121, 1993.

[15] C. L. Hamblin. *Fallacies*. Methuen, London, UK, 1970.

[16] S. Modgil and T. Bench-Capon. Metalevel argumentation. *Journal Logic and Computation*, 21(6):959–1003, 2011.

[17] A. Moretti. Argumentation theory and the geometry of opposition (abstract). In *7th Conference of the Inter. Soc. for the Study of Argumentation (ISSA'10)*, 2010.

[18] T. Parsons. The traditional square of opposition. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, 2008.

[19] C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: a Treatise on Argumentation.* Notre Dame Press, University of Notre Dame, 1969.

[20] J. Pollock. Defeasible reasoning. *Cognitive Science*, 11(3):481Ű518, 1987.

[21] J. Pollock. How to reason defeasibly. *Artificial Intelligence Journal*, 57:1–42, 1992.

[22] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[23] G. Quiroz, D. Apothéloz, and P. Brandt. How counter-argumentation works. *Argumentation Illuminated, (F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds)*, pages 172–177, 1992.

[24] I. Rahwan and G. Simari (eds). *Argumentation in Artificial Intelligence*. Springer Verlag, 2009.

[25] N. Rescher. *Dialectics: A controversy-oriented approach to the theory of knowledge*. State University of New York Press, 1977.

[26] C. Salavastru. *Logique, Argumentation, Interprétation*. L'Harmattan, Paris, 2007.

[27] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence Journal*, 53:125–157, 1992.

[28] A. Tarski. *On Some Fundamental Concepts of Metamathematics*. Logic, Semantics, Metamathematic. Edited and translated by J. H. Woodger, Oxford Uni. Press, 1956.

[29] S. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.

[30] D. Walton, C. Reed, and F. Macagno. *Argumentation schemes*. Cambridge University Press, 2008.

[31] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In *ArgMAS*, pages 42–56, 2005.