Commentary on Cacean's Reliability of Argument Mapping

ROOSMARYN PILGRAM
Leiden University & University of Amsterdam
r.pilgram@hum.leidenuniv.nl & r.pilgram@uva.nl

1. INTRODUCTION

Cacean's paper concerns the interesting question of how argument mapping can be conducted in such a way that it generates generalizable results, despite the hermeneutic underdetermination of actual argumentative discourse. As Cacean points out, argumentative discourse is hermeneutically underdetermined in the sense that premises or conclusions can be left implicit and there are often various possible ways in which these elements can be made explicit. Even if no element is left implicit, underdetermination can arise from ambiguous linguistic elements in the discourse itself.

Hermeneutic underdetermination poses a difficulty for argument mapping: in argument mapping, reasoning structures within argumentative discourse are schematically depicted, but how can such structures be depicted if ambiguity allows for several interpretation of what these structures exactly amount to? Cacean argues that although underdetermination can lead to different interpretations of argumentative discourse, this does not mean that every interpretation is equally valid. He proposes context-dependent empirical reliability constraints to capture the margin of interpretation. By applying these constraints, the validity of argument maps can be determined.

In this commentary, I will deal with two questions about Cacean's paper, a more theoretical one and a more practical one. The theoretical question concerns the notion of 'validity', the practical question the reliability measures that Cacean uses to determine validity.

2. VALIDITY OF ARGUMENT MAPS

In his paper, Cacean takes care to demarcate the notion of validity: he does not deal with formulating criteria for evaluating whether an argument map constitutes a valid interpretation of a particular piece of argumentative discourse (which he calls 'validity$_1$'), but with the way in which we can check whether alternative argument maps of the same

argumentative discourse can still be regarded as valid when taking into account the interpretational margin ('validity$_2$').[1] However, I wonder whether Cacean's notion of validity$_2$ fully captures what he is after, since, based on this notion, it seems to be possible to call argument maps valid that contain internal contradictions or are completely blank, which does not seem to be very meaningful.

Cacean argues that argument maps are valid if they represent at least all the elements of a minimum argument map and not more elements than a maximum argument map. In other words, valid argument maps should at least include all the elements in the argumentation that are unambiguous (i.e., unique), but could also include additional ambiguous elements as long as they deal with the claims, arguments or relations between claims and arguments in the discourse.

Now imagine that there is a case of edge-type ambiguity (i.e., ambiguity about whether an argument is meant to support or attack a claim), because the exact relation between the argument and claim is left implicit in the discourse. Other than that, no ambiguities exist (i.e., the claim and argument are both unique). This would mean that the maximum argument map consists of a claim that is both supported and attacked by the same argument (see Figure 1). Following Cacean's reasoning, this map should be considered valid, since the maximum argument map is by definition a valid argument map. Yet, the map seems to suggest that a contradiction is present in the discourse itself, while that is not the case (only the relation between claim and argument is ambiguous). So, what does the validity of this map exactly signify?

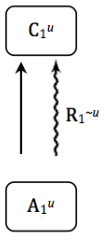

Figure 1 – Argument map of a claim of which it is ambiguous whether it is supported or attacked by the argument (i.e., edge-type ambiguity).

[1] It should thus be noted that Cacean deals with validity in a methodological sense (in that validity indicates whether a measurement measures that which it is supposed to measure), rather than validity in a logical or argumentation theoretical sense (in that validity indicates that a conclusion cannot be untrue if the premises are true – or any other variation of this).

Furthermore, in the case of argumentative discourse in which all elements (i.e., arguments, claims and relations) are ambiguous, the minimum argument map would actually not contain anything at all. Such a blank 'map' could nevertheless be regarded as a valid argument map based on the idea that a minimum argument map is by definition valid. This again raises the question what it exactly means to call an argument map 'valid' in the sense of Cacean's validity$_2$.

These examples suggest that an examination of validity$_2$ (the validity that indicates whether alternative argument maps of the same argumentative discourse are still acceptable) without some examination of validity$_1$ (the validity that indicates whether an argument map fulfils the normative acceptability criteria) might be undesirable.

3. RELIABILITY MEASURES

Apart from this theoretical issue, I would like to pose a more practical question about the way in which reliability is measured in the paper. As Cacean explains, the fact that people might interpret the same argumentative discourse differently because of ambiguities in the discourse means that a variety of argument maps can be valid. This variety could, however, pose a difficulty for empirical research: when using standard empirical reliability tests, ambiguities in the discourse result in lower inter-coder reliability. The context independency of these standard reliability tests therefore makes them an unsuitable measurement of discourse characteristics.

To be able to determine the reliability of argument maps, Cacean proposes an alternative way to deal with their reliability, namely by specifying reliability thresholds that are relative to margins of interpretation. A strong reliability threshold could, for example, be that coded argument maps should not exceed the mean distance of all valid argument maps. Weaker thresholds could take into account the distribution of argument maps.

A prerequisite for such context dependent reliability thresholds is that the distance between argument maps can be calculated. Indeed, Cacean does so by means of Hemming distances. The Hemming distance amounts to the number of elements that are not shared by two argument maps. Although this measurement is attractively simple, it does not seem to take into account important distinctions in the differences between argument maps.

Consider, for example, the argument maps in Figure 2. Each of these maps represents the same argumentative discourse. The difference between the maps in (a) and (b) is that in (b) one subordinate argument is lacking (argument A_2). The difference between the maps in (a) and (c) is that in (c) the claim is represented twice: once

as the main claim and once as an argument directly in support of this main claim.

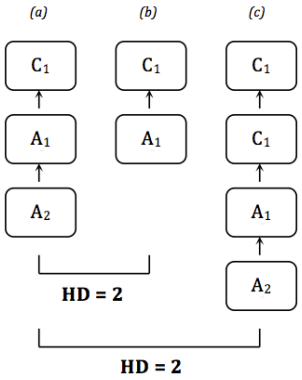

Figure 2 – Hypothetical argument maps with Hemming distances in bold.

Let us assume that the correct argument mapping of the discourse is the one in (a). The maps in (b) and (c) hence both contain mistakes. One could nonetheless argue that the more serious mistake occurs in (c): this is not just a matter of leaving out a subordinate argument (as in (b)); the discourse is mapped as though it contains circular reasoning (C_1 is depicted as supporting C_1), while that is not the case in the actual discourse. Yet, when calculating the Hemming distances between the maps (a) and (b), and (a) and (c), it is in both cases a distance of 2.[2] The question thus arises whether the Hemming distance is a suitable measure of differences between argument maps.

4. CONCLUSION

Despite the above questions, I think that Cacean's paper offers a very original way to evaluate argument maps. Given the increased importance of argument mapping techniques in combination with the hermeneutic underdetermination of language, a means to empirically determine the validity of argument maps is highly desirable. The idea of reliability thresholds is, in my view, an interesting way to tackle this matter. Thus, my questions should not be regarded as undermining this idea, but merely as sharpening it even further.

---

[2] The Hemming distance between the maps in figure 2 (b) and (c) is 4, so one could regard the map (c) as an invalid map after using a strong reliability threshold. Still, such evaluation would be *in spite of* the seriousness of the mistake in it, not *because of* this seriousness.