Building Monologue

Chris Reed
Department of Applied Computing
University of Dundee, Dundee DD1 4HN Scotland
Tel. (+44)(0) 1382 348083
e-mail Chris.Reed@computing.dundee.ac.uk

Abstract

To build an argument - and particularly an argument presented as a monologue - a writer must assemble and marshal a battery of supports for a claim. Some of those supports will be arranged in convergent structures, some as linked; some will be expressed, some will be left implicit; sometimes a support will need further support of its own - and sometimes, not. This paper explores the factors which lead a writer to make particular choices, the interactions between those factors, and the constraints on a writer's freedom in exercising her power, drawing on recent findings in computational modelling of the generation process.

Introduction

Within Artificial Intelligence (AI), there is growing interest in the field of Natural Language Generation (NLG), and the design of algorithms for rendering some communicative goal in an appropriate linguistic form. Given the intricate constraints at morphological, syntactic, semantic and pragmatic levels, the preoccupation in NLG has been upon producing text which is coherent. Yet it is becoming clear that this focus not only leads to wildly underspecified structural representations, but that it also ignores important aspects of the intentions behind a given discourse encounter. These problems are particularly pertinent to the generation of textual argumentation, where merely coherent text may very well fail to be persuasive (assuming, for the time being, that persuasion is the most common goal of argumentation), and may thus fail to service the communicator's aims at all.

In addition to the basic intention of a speaker[1] to create an argument which is textually coherent, she is also concerned with ensuring the coherency of the argument (a higher level constraint concerned with placement of premises and conclusions), deciding which material to include and which to omit (rather than the more usual approach, adopted in NLG, of communicating everything), and deciding - within the foregoing restrictions - how best to arrange the argument components. All of these interacting goals are derived from a small set of overarching intentions, typically including one of convincing the hearer that some conclusion holds.

In (Reed, 1998) a theoretical framework is presented which enables these various issues to be represented in a common language, and a computational implementation of this framework is demonstrated to be capable of designing a range of extended argument forms. The next section examines some of the factors taken into account by the framework, and how their influence is supported in the generation process, before discussion in the final section of how such a speaker-oriented view of the argumentation process might inform argument analysis and evaluation.

Determining Content

The most fundamental problem facing the designer of an argument is premise availability: do there exist premises which can support a given conclusion or which can rebut or undercut some counterargument? Classically, within studies of argumentation, this phase of invention is seen as lying outside the scope of analysis - and indeed of rhetoric. Hugh Blair explains:

> "Art cannot go so far, as to supply a speaker with arguments on every cause, and every subject; though it may be of considerable use in assisting him to arrange and express those, which his knowledge of the subject has discovered. For it is one thing to discover the reasons that are most proper to convince men, and another, to manage these reasons with the most advantage. The latter is all rhetoric can pretend to do." (Blair, 1838: 427)

Yet clearly, 'discovering the reasons that are most proper' is a crucial stage in argument design: how can appropriate premises be found? A traditional AI perspective on the problem suggests that some form of means ends analysis might be appropriate: define the situation we wish the system to reach, and then determine a series of operations which can move the system towards that goal. The goal state might be characterised using the mental attitudes of the audience - say, that a given hearer believes a given proposition (for the time being, it is more convenient to use a mental attitude such as belief, rather than the more flexible and precise notion of commitment; the discussion returns to this point below). How then might the 'operations' be characterised - what is it that changes the state, moving the system closer to the goal? A pragma-dialectician might conceivably suggest that such operations are in essence speech acts, and indeed, that such a characterisation would concur neatly with the functionalization component of the p-d methodology. Although this functionalization and subsequent computational operationalisation of speech acts forms a key component of the framework as a whole, its introduction at this stage is premature; speech acts themselves are too fine-grained for their role in large scale argumentation to be adequately captured in a computational model. Consider, for example, speech acts corresponding to a refutation in which some proposition is put forward and then shown (by the same interlocutor) to be false. Such constructions are perfectly common, and psychological evidence exists to demonstrate that they are highly effective (Hass and Linder, 1972). It is unclear, however, how a purely speech act based functional account of argument progression could adequately deal with the tabling of a proposition which is at odds with the intended conclusion. Some form of intermediate level structure - connecting speech acts with broad discourse aims - is required.

Unfortunately, the structure to which reference is being made seems to be closely associated with the identification of argument units, a recurrent problem in the analysis of argumentation (witness, most recently, papers by Freeman and Vreen at the 1998 Amsterdam meeting). Following those concerned with practical analysis of argument, however, the current work adopts a pragmatic approach, and though acknowledging that there may be shortcomings and in some cases a lack of generality, the characterisation seems to be flexible enough to cover a wide range of argumentation.

It is proposed that the formal structure of an argument - Modus Ponens, Modus Tollens, Disjunctive Syllogism, etc. - is often useful both to characterise the relationships holding between speech acts, and also to define the extent of an argument unit. There are several obvious potential objections to this proposal. First, that it admits of deductive reasoning only; this can be countered simply by the inclusion of inductive forms, with structures such as Inductive Generalisation available explicitly. Second, that even with inductive reasoning embraced, the old deductive/inductive dichotomy is being maintained as both applicable and exhaustive. It does seem reasonable to claim that some arguments are truly inductive, so, from a generative point of view, the distinction is indeed applicable, as an argument designer may need to employ both deductive and inductive arguments. There is, however, no claim of exhaustiveness: although not implemented, alternative patterns such as abduction or conduction might also be included. Furthermore, argument forms which represent legitimate versions of fallacies (such as an appeal to a relevant authority) can also be captured. Third and fourth, that the deductive forms, with their reference to the basic components of formal logic, advocate a return to both the truth criterion and to a reliance on material implication, such that deductive soundness can be demonstrated. The proposal does not rest upon the canons of deductive logic: the requirement that premises be true is not only metaphysically problematic, but runs counter to the spirit of the project, which is above all speaker-oriented. Thus any notion of truth is eschewed in favour of a relativised modality, 'speaker-believes', and, at times, to something even weaker (and, perhaps, duplicitous) such as 'speaker-believes-that-hearer-believes'. Further, the implication in deductive forms such as Modus Ponens is also weakened to a semantic relationship of support. In so doing, the pseudo-deductive operators, and particularly Modus Ponens, can be seen as abstractions of Kienpointner argument schemes (Kienpointner, 1992). Indeed Kienpointner's *basic prototype of argumentative schemes*:

Warrant

Argument

Therefore: Conclusion

is, perforce, closely analogous with the Modus Ponens characterisation

get the hearer to believe and be aware of $P \supset Q$

get the hearer to believe and be aware of P

in order to get the hearer to believe Q

given that $\supset$ is read as shorthand for 'supports' (rather than implies). Thus the major premise of a Modus Ponens (when weakened to indicate a support relationship) is equivalent to Kienpointer's warrant. There are two extensions necessary: the first, identified by Kienpointner is to handle complex argumentation; the second is to handle argument which falls into other deductive patterns. This latter extension has already been alluded to: definitions of Modus Tollens and so on are analogous to the Modus Ponens given above. The former extension is also quite straightforward, bearing in mind how the process of argument construction proceeds.

Viewing the process as means ends analysis, and characterising the final goal as something of the form 'the hearer believes *x*', enables the Modus Ponens scheme, when appropriately instantiated (with the conclusion being *x*, and the premise being some other proposition, *y*), to represent the operation of moving from a state in which the hearer believes and is aware of *y* and $y \supset x$ into one in which the hearer believes *x*. Following the means ends analysis tradition, the task then is to arrange for the system to be in that earlier state such that the final operation can be performed. There are two components to the state which need to be addressed. First, getting the hearer to believe and be aware of *y*. It is easy to conceive that perhaps some third proposition *z* supports *y* in the same way that *y* supports *x*. A new instantiation of the Modus Ponens scheme will handle that nicely. It is also necessary to make the hearer *aware* of *y* - this is most easily achieved by performing a speech act corresponding to *y* (though there are reasonably straightforward rules governing when it is safe to assume that the hearer is already aware of a proposition, and utterance should be avoided for the sake of brevity). Let us assume at this point that the speaker believes that the hearer will accept all the remaining unsupported premises in the argument (namely, $y \supset x$, $z \supset y$, and *z*) so long as she makes him aware of them (through a series of speech acts). The result is a straightforward 'serial' argumentation with one premise supporting a conclusion which stands as a premise in a supervening argument. Notice that the problem of identifying argument units is discharged by permitting individual utterances to play roles in more than one argument unit: the unit is one instantiation of a scheme (i.e., one operation) which has (usually) two premises and (usually) one conclusion.

Other arrangements of argument units into larger scale complex argumentation are also easily accommodated. Returning to the point at which a single instantiation of Modus Ponens (with y supporting x) has been selected, let us consider the result of achieving explicitly the second part of the state - that the hearer believe and be aware of $y \supset x$. Suppose there is some proposition b which can be used to support the hearer's belief in $y \supset x$; a new instantiation of the Modus Ponens scheme can be added to the argumentation. In supporting the warrant $y \supset x$, the proposition b is functioning in much the same way as Toulmin's 'backing' (Toulmin, 1958). It is, however, just another premise - the framework follows Freeman's (1991) comment at the close of his discussion of Toulmin's model, "The basic elements in arguments are good, old-fashioned premises and conclusions" (p88). Identifying this correspondence also leads naturally to drawing an analogy between the process of adding a new operation (such as the Modus Ponens with b supporting the warrant $y \supset x$) and the implicit posing of a challenger's questions: the argument from b can be seen to have been added in response to Freeman's *relevance question*: "Why is that reason (y) relevant to the claim (x)?" Freeman's account can then be followed further to view such an arrangement as linked, but again, the issue of defining whether it is arguments or parts of argument that are linked is side-stepped by continuing to see instantiated argument schemes as units and an 'argument' or 'argumentation', simply as the totality of such units.

Finally, convergent argumentation is constructed in a similar way. Returning again to the point in the means ends analysis process at which a single Modus Ponens scheme has been instantiated (with y supporting x), it may be possible for the speaker to find another premise, y' which also supports x. For although it is convenient to see the conclusion of the Modus Ponens scheme as 'hearer believes P' (particularly considering the similarity in form with deductive Modus Ponens), because the major premise does not involve material implication, it is more accurate to see the conclusion as 'hearer is *more likely* to believe P'. With this in mind, it is easy to see that further instantiated schemes may be added to support a given conclusion at the discretion of the speaker. Again, such additions can be seen as being performed in response to implicit challenges, this time of the *ground adequacy* type: "Can you give me another reason?".

As well as admitting serial, linked and convergent argument structures based on acceptability, relevance and ground adequacy questions respectively, the framework also handles several other problematic phenomena in an intuitive manner. One problem with the Toulminian model of argumentation that Freeman spends some time analysing is how 'metatheoretical' rules of argument (such as disjunctive syllogism and conjunction) can be integrated into the object level system; the approach in the current work is simply to support the panoply of argumentation schemes such that a Modus Tollens, a Disjunctive Syllogism, an Inductive Generalisation or even an Argumentum ad Populum can be used in much the same way as the Modus Ponens schemes have been used in the foregoing example - given that certain conditions are met.

The framework also offers a consistent means of handling examples such as

> We shouldn't go. John would drive, and not only is his car uncomfortable, but his wife will travel with us, and she is really horrible.

In some structural theories, this would pose something of a problem. The conclusion 'we shouldn't go' is being supported by the premise 'John would drive', and then, using Freeman's account of the distinction between linked and convergent support, it seems that 'John's car is uncomfortable' and 'John's wife will travel with us' are independent of each other (that is, convergent - obtained by asking the ground adequacy question), but both provide linked support with 'John would drive'. The picture is then further complicated by the linkage of 'John's wife will travel with us' with 'John's wife is horrible' (since it is the relevance question which has been asked between these two statements). This analysis can be justified by consideration of the implicit questions:

We shouldn't go.

    Why should I believe that premise? (Acceptability)

John would drive.

    Why is that relevant? (Relevance)

His car is uncomfortable.

    Can you give me another reason? (Ground adequacy)

His wife will travel with us.

    Why is that relevant? (Relevance)

She is really horrible.

Although Freeman's analysis is a great help in determining how premises inter-relate, it is not clear that his theory could represent the overall structure, which might be diagrammed as in Figure 1. Arrows are here permitted to support arrows; this is intended to capture the underlying representation of the argument schemes in which one instantiation of a scheme may support either the minor or major premise in another (for the sake of clarity, the examples here employ only Modus Ponens schemes).

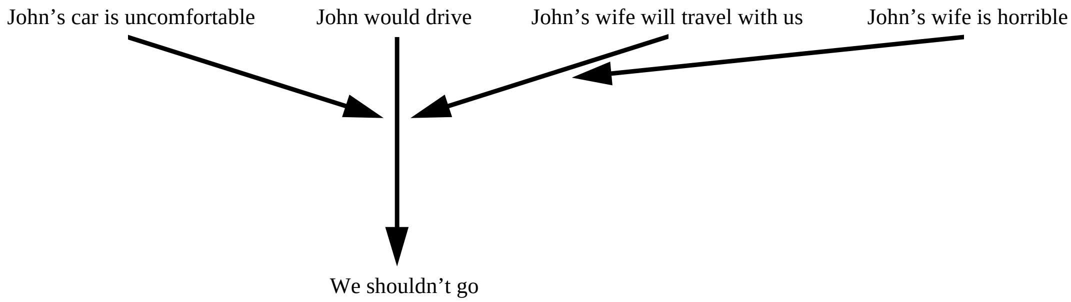

Figure 1

Importantly for the generation process, this framework also lends itself to the problem introduced at the beginning of this section: how to select premises. By operationalising schemes, a computational system has a means by which to bring new evidence into an argument: if a current goal exists of increasing a hearer's belief in a proposition *p*, all those schemes which can achieve that (almost all of them, in this case) are considered in turn[2]. For each scheme, the knowledge base is checked for suitable information - thus, for MP, the knowledge base needs to contain a statement of the form $X \supset p$. If such information exists, the scheme is selected and instantiated appropriately; if not, the next scheme is considered. For some schemes (particularly inductive and fallacious) the selection criteria are rather more rigorous, so that, for example, the supporting instances referred to in an inductive generalisation must all have a property in common with the conclusion.

Given that all available scheme instantiations applicable to a given goal are identified, the procedure attempts to include as many lines of argumentation - as many separate supports - as the knowledge base can supply. Thus there is no notion of the sufficiency of any premise, argument, set of supports or body of evidence for a given conclusion. This full structure, however, only represents a possible maximum for the realised argument, for there are then a battery of mechanisms for reducing the extent of the argument, some of which contribute to what can ultimately be seen as a notion of sufficiency[3].

One simple heuristic for pruning an argument is to restrict the number of supports for a given conclusion, keeping only the strongest evidence. Determining how much to delete is a dynamic decision, taken in conjunction with a complementary heuristic restricting the depth (i.e. serially arranged supports) of an argument. The activity of both of these heuristics is then further impacted by global features of an argument such as the medium in which it is presented (oral arguments may tend to have less complex structure than textual arguments) and overriding constraints such as limits on column inches or allotted speaking time.

Another pruning technique is to ensure that arguments refuting an opponent's counterclaim carry sufficient weight, and, though less widely applicable, that arguments supporting one's own claim are sufficiently strong. In the latter case, an accumulation of weak arguments may have little or no effect on a conclusion, but in the former case, it has been demonstrated that weak refutation arguments are usually detrimental to a conclusion (Hass and Linder, 1972).

Finally, though not strictly pruning, it is also apposite to consider the means by which enthymemes are created; i.e. how the decision is made to leave components of single argument schemes implicit. This process forms a component of the mirror image of the pragma-dialectical process of reconstruction: where the latter tries to supply missing premises, the former tries to remove explicit mention of superfluous premises. Enthymeme contraction, then, proceeds on the basis of several factors - though here again, the process is crudely simplified, but captures enough to offer interesting results. In the first place, a rough, generic measure of hearer 'competence' (that is, his ability to follow argumentation) is used to determine whether the hearer would still follow a particular argument unit if a component were left implicit. A low score for hearer competence would mean less enthymematic contraction in an argument. Secondly, the overall scepticism of a hearer is also estimated; if this estimate is high, enthymemes will again be rarer. Thirdly, and most importantly, enthymeme contraction is dependent upon the specific attitude of the hearer with respect to the proposition in hand. If the hearer is known to believe a proposition, or, more frequently known not to disbelieve that proposition, then enthymematic contraction is much likely than if the hearer is known to disbelieve the proposition. Finally, context is also a key determinant of the licensing or otherwise of a given enthymeme: in a sorites, for example, to leave too many components implicit could quickly render the argument incomprehensible.

Concluding Remarks

The previous section has offered an overview of some of the features of a computational model of the argument generation process, and has demonstrated how an operationalisation of argumentation schemes (or at least of abstractions of argumentation schemes) leads to an intuitive way of diagramming argument structure. In addition to diagrammation, however, the approach also touches on a number of other issues of direct relevance to the problems of argument analysis.

In the first place, it lends very direct and literal support to the pragma-dialectic conception of functionalization (albeit with an intermediate level of organisation between discourse and interpersonal goals on the one hand, and speech acts on the other). From a goal-directed generation point of view, speech acts, of necessity, have a functional role to play and to be co-ordinated.

Secondly, the approach lends itself very directly to the generation of 'clue words': words or phrases whose primary role is to make explicit the structure of the argument. It is demonstrated in (Reed, 1998) that the *but* commonly associated with Modus Tollens, conclusion markers such as *so* and *therefore*, and a number of other clue words and even punctuation can be generated directly by the model. The importance of such clue words in the analysis of argument has long been recognised, both in general terms (identifying conclusions by spotting markers such as *therefore*), and in specific cases (Kienpointner, for example, associates some clue words with particular schemes), and a fuller study is also under way within the pragma-dialectical framework (Snoeck Henkemanns, 1997). It is reasonable to hope that an improved understanding of the rules which license the introduction of such clues might also inform the process of uncovering argument structure by reference to clues.

Thirdly, the generation framework also touches upon the issue of the distinction between logic and rhetoric. Both Rescher (1997) and Freeman (1998) have intimated that the two are active at different levels - the former that rhetoric is used, at the very least, for basic premise establishment; the latter that it is rhetorical, rather than logical, concerns which give rise to the pragma-dialectic definition of compound argumentation. Interestingly, in a generative model, it is useful to have both logic and rhetoric contributing at *every* level of abstraction. By 'logic' is here meant the structural relationships between premises and conclusions (and is thus close to Freeman's use of the word), and by 'rhetoric' are meant hearer-specific, arational, contributions to the content and arrangement of an argument (thus lying close to the Perelman and Olbcrechts-Tyteca conception). Thus at every level of abstraction (that is, every horizontal slice through a diagram of an argument) there will be logical concerns (which argument schemes to employ, how to instantiate them, whether to include support at all, etc.) and rhetorical concerns (whether to use climax ordering, whether to place a conclusion first, etc.).

Finally, there is the broader issue of whether consideration of the pressures exerted on a speaker whilst constructing an argument can be of use in reconstructing the original, deep structure. One example of this issue was considered above: enthymeme contraction. If an analyst faced with reconstructing an argument comes to a potential enthymeme, she might consider whether the speaker's intended audience would have been expected to believe (or not to disbelieve) the omitted premise. If so, the analyst might feel more justified in proposing the omitted premise than if it were one the hearer was sure to have disbelieved.

The model presented above makes a large number of important simplifications which represent potential departures for future work. The first is the - by the standards of argumentation theory - archaic reliance on a hearer's mental attitudes, rather than his commitments within the current discourse encounter. Although the implemented use of beliefs can be shown to be very close to the usual notion of commitment[4], it nevertheless relies on several simplifying assumptions. Most importantly, if an epistemic underpinning is eschewed in favour of one based on commitment, it makes the extension of the work to dialogic argument that much easier.

The restriction of the current work to monologue forms the second key area of development required: adducing the further constraints imposed on a speaker by the dialogic situation not only represents an interesting extension of the approach to generation, but also a further test of its utility within argument analysis.

Lastly, the current model focuses specifically upon persuasive discourse, and, for the most part, on such discourse with very few (or at least, easily captured) interpersonal goals. There is, however, nothing in principle barring the development of argumentation schemes which support other forms of argumentation (negotiation and deliberation, for example), bringing the work in line not only with broader concerns in NLG, but also with recent advances in informal logic. In addition to applied aims outwith argumentation theory, a key goal of such work is to contribute to the understanding of the small scale, inter-utterance structure of these varied forms of argumentation, and to represent that structure in a common format. The current work has demonstrated that such a common underlying framework is not only feasible, but that it is also of practical use within NLG. Furthermore, it is also of potential interest within argumentation theory where it emphasises the argument-as-process view, forming something of a bridge between the traditionally logical (e.g. Freeman) and the traditionally dialectical (van Eemeren and Grootendorst). There is also metatheoretical value to the project: by submitting any analytic theory of argument structure to the demands of functioning as a *generative* framework, that theory's adequacy - expressiveness, complexity, consistency, and to some extent flexibility and generality - can all be investigated. The current work thus performs two tasks: first, it presents elements of a theory of argument structuring which follows in the spirit of Freeman's work, but draws it towards the pragma-dialectical school; and second, it represents a support for the metatheoretical claim, through practical demonstration, that the demands of the generation process can yield constructive insights into the structure of natural argument.

References

Blair, H. (1838) *Lectures on Rhetoric and Belles Lettres*, Charles Daly

Freeman, J.B. (1991) *Dialectics and the Macrostructure of Arguments*, Foris Publications

Freeman, J.B. (1998) Argument Structure and Disciplinary Perspective, in *ISSA'98*, Amsterdam

Hass and Linder (1972) Counterargument availability and the effects of message structure on persuasion. *Journal of Personality and Social Psychology* 23 (2) pp219-233

Kienpointner (1992) How to Classify Arguments, in van Eemeren, F.H.,Grootendorst, R., Blair, J.A. & Willard, C.A. (eds) *Argumentation Illuminated*, SicSat

Reed, C.A. (1998) *Generating Arguments in Natural Language*, PhD Dissertation, University College London

Rescher, N. (1997) Rhetoric and Rational Argumentation, in *OSSA'97*

Snoeck Henkemanns, A.F. (1997) Verbal indicators of argumentation and explanation, in *OSSA'97*

Toulmin (1958) *The Uses of Argument*, Cambridge University Press

---

[1] Throughout this paper, the term 'speaker' is used to refer to authors of written or spoken monologues, 'hearer' to refer to the audience or reader, and 'utterance' to a written or spoken speech act corresponding to a single proposition.

[2] A crude preference order exists across argument schemes, such that pseudo-deductive operators (MP, MT, DS) are preferred over inductive operators, and those over fallacious operators; that is not to claim that such a scale is legitimate either for evaluation or for rhetorical purposes, rather, it attempts to capture the intuition that an (enlightened) speaker might generally prefer a 'deductive' argument to one further removed from formal logic.

[3] In fact, this maximal set of argument structure is never actually created, since the process of pruning the argument is intertwined with the generation process itself. Apart from the intuitive correspondence between such underspecification and the process by which human speakers invent argument, there are important computational benefits which render the process tractable.

[4] Since the beliefs of the hearer are not represented directly, but rather, are relativised to the speaker (that is, expressions involving hearer belief are always of the form BEL(S, BEL(H, x))), and because the current work focuses solely upon monologue, the speaker's reasoning about the state of the hearer can be assumed to be accurate given the accuracy of her initial model of him.