DETAILED NOTES (there were what we used internally for generating the CAGs)

REPRESENTATION ISSUES

1. Nodes in the graph are “factors”.
2. We interpret an edge between two factors (source S, target T) as a stand-in for a function that relates changes in S to changes in T.
3. The polarity of the edge captures the direction of change to T. We also need a way to represent the delta of S that is associated with the delta of T. Right now, this is represented as an attached bubble with polarity.

NORMALIZATION AND REPRESENTATION

1. We have performed “linguistically licensed” normalization – e.g., if factor S increases food insecurity, we will normalize food insecurity to food security (based on knowing the linguistic relation between “security” and “insecurity”). By inference, we assume that (X increases anti-P) is equivalent to (X decreases P). That means we will represent:

*S increases food insecurity* as:

*S decreases food security*

1. We will NOT infer that inverses hold, e.g., if the sentence says:

Increase in conflict decreases food production

we will NOT infer that “decrease in conflict increases food production”

Thus, we would normalize

*Low cereal production increases food insecurity* to

*Low(cereal production) decreases food security*

But we don’t mark the “default” case where the delta of the source is positive,   
 e.g. we don’t need to write:

High(cereal production) increases food security

1. This leaves us with the problem of how to represent specific state descriptions of factors (e.g., low, collapsing, crisis, etc.). This is particularly an issue for Source factors -- the edge polarity gives direction of change for the Target. For now, we represent this as a descriptor (attribute) associated with the Source node, and the directed edge goes from the descriptor to the Target. Note that the default direction for the Source node is “increase” and is not explicitly marked – although for symmetry, we could also capture that information if desired.

WHAT TO READ

1. To gather a “background” model, we will look at some tutorial/introductory material and develop (parts of) generic CAGs from those. This mirrors the use of starting networks/ knowledge bases in Big Mech to provide a scaffold of background knowledge which is often not made explicit in reports on specific regions/periods. It is almost certainly much more efficient for reading teams to start with some basic background knowledge/models developed by human readers, rather than trying to deduce these models entirely by automated reading.
2. The initial paragraphs supplied by Joshua provide specific information for South Sudan in a particular time period. For now, we are focusing on extracting the basic relationships mentioned in the paragraphs, and are ignoring temporal and geographic information (although these are often there). That is, we are generalizing from the South Sudan specific materials by ignoring the temporal and geo-spatial information.
3. For now, we are also ignoring any qualitative or quantitative “parameterization” information beyond assigning basic polarity to the edge (and to the Source node, as needed). However, this information needs to be retained (eventually) and is noted in some of the representations.

NORMALIZATION/GROUNDING

1. There are some useful resources that give lists of terms suggesting causal linkage (from annotation guidelines used for causal relation mark-up, in Paramita Mirza and Sara Tonelli Proceedings of COLING 2014, pages 2097–2106):

C-SIGNALs are used to mark-up textual elements signaling the presence of causal relations, which include all causal uses of prepositions (e.g. because of, as a result of, due to), conjunctions (e.g. because, since, so that), adverbial connectors (e.g. so, therefore, thus) and clause-integrated expressions (e.g. the reason why, the result is, that is why).

There is more complete information in Appendix A.

1. We will provide a spreadsheet that provides a tabular view of the predicate-argument relations, plus normalizations and capture of some of the metadata related to factuality, certainty and severity. This is a first pass (for the first four paragraphs).
2. We don’t know how to represent complex concepts, such as “access to food” or “access to clean water” – or even phrases like “food prices”. For now, these are treated as primitive concepts, but these may need to be further decomposed.

Appendix A: Annotation Guidelines from the NLP Community

The attached papers describe an annotation framework used to apply causal relation annotation to a corpus already containing temporal annotations.

1. Paramita Mirza and Sara Tonelli. An Analysis of Causality between Events and its Relation to Temporal Information. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 2097–2106, Dublin, Ireland, August 23-29 2014.

Abstract

In this work we present an annotation framework to capture causality between events, inspired by TimeML, and a language resource covering both temporal and causal relations. This data set is then used to build an automatic extraction system for causal signals and causal links between given event pairs. The evaluation and analysis of the system’s performance provides an insight into explicit causality in text and the connection between temporal and causal relations.

Potentially of interest to us as we consider what to extract and represent in CAGs:

“Similar to the <TLINK> tag in TimeML for temporal relations, we introduce the <CLINK> tag to mark a causal relation between two events. Both TLINKs and CLINKs mark directional relations, i.e. they involve a source and a target event….

C-SIGNALs are used to mark-up textual elements signaling the presence of causal relations, which include all causal uses of prepositions (e.g. because of, as a result of, due to), conjunctions (e.g. because, since, so that), adverbial connectors (e.g. so, therefore, thus) and clause-integrated expressions (e.g. the reason why, the result is, that is why).

We assign a CLINK if, given two annotated events, there is an explicit causal construction linking them. Such construction can be expressed in one of the following ways:

1. Expressions containing affect verbs (affect, influence, determine, change, etc.), e.g. Ogun CAN crisis S influences the launch T of the All Progressive Congress.

2. Expressions containing link verbs (link, lead, depend on, etc.), e.g. An earthquake T in North America was linked to a tsunami S in Japan.

3. Basic constructions involving causative verbs of CAUSE, ENABLE and PREVENT type, e.g. The purchase S caused the creation T of the current building.

4. Periphrastic constructions involving causative verbs of CAUSE, ENABLE and PREVENT type, e.g. The blast S caused the boat to heel T violently. With “periphrastic” we mean constructions where a causative verb (caused) takes an embedded clause or predicate as a complement expressing a particular result (heel).

5. Expressions containing CSIGNALs, e.g. Its shipments declined T as a result of a reduction S in inventories by service centers.

We annotate both intra- and inter-sentential causal relations between events, provided that one of the above constructions is present. We do not annotate causal relations that are implicit and must be inferred by annotators, because they may be highly ambiguous and would probably affect inter-annotator agreement.”

1. Paramita Mirza, Rachele Sprugnoli, Sara Tonelli, and Manuela Speranza. 2014. Annotating Causality in the TempEval-3 Corpus. In Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL), pages 10–19, Gothenburg, Sweden, April. Association for Computational Linguistics.

Abstract

While there is a wide consensus in the NLP community over the modeling of temporal relations between events, mainly based on Allen’s temporal logic, the question on how to annotate other types of event relations, in particular causal ones, is still open. In this work, we present some annotation guidelines to capture causality between event pairs, partly inspired by TimeML. We then implement a rule-based algorithm to automatically identify explicit causal relations in the TempEval-3 corpus. Based on this annotation, we report some statistics on the behavior of causal cues in text and perform a preliminary investigation on the interaction between causal and temporal relations.

Additional details provided in this paper include discussion of event attributes:

1.       Polarity (used to distinguish affirmative and negative events)

2.       Factuality (used to distinguish facts from counterfactual and speculative/future situations)

3.       Certainty (used to distinguish certain from uncertain events)