Chapter 7
Conclusion

> Yesterday, you said tomorrow. So just do it.
>
> SHIA LEBOEUF

Temporal annotation is difficult for both humans and machines. The task of determining how particular events are ordered or nested is part of this temporal annotation problem and has been the goal of this book. This is known as the temporal link labelling problem. The state of the art in this problem has advanced slowly in recent years, without reaching high enough performance levels to consider it solved. This book has investigated the problem of temporal link labelling.

A principled investigation began with a data-driven exploration of temporal links in a publicly-available corpus. This led to the identification of a set of difficult links, which many modern approaches cannot automatically label correctly. Formal and subjective analyses of this difficult link set were conducted. Results suggested multiple avenues of research (in the form of types of information seemingly used to label temporal links) and the two that were selected for investigation were signal-based links and links where there is a change of tense or aspect.

For the part of the signals, these were characterised as words or phrases associated with a pair of events or timexes that provide explicit information about their temporal relation. Experimentation with a machine learning approach showed that they were very helpful in link labelling, giving about a 50 % error reduction. However, they are under-annotated in TimeBank, so attention turned to the task of automatically annotating signals. This was broken down into a two part task: discriminating signals (e.g. finding which phrases occur in text with a temporal sense and in a link labelling-supportive function) and association of signals, that is, determining which pair of events or timexes has its relation described by a given signal. Machine learning approaches and feature sets were identified for both these tasks. Finally, automatic signal annotation was attempted on a corpus initially devoid of signals and the automatically-found signals used to help classifier-based temporal link labelling on

L.R.A. Derczynski, *Automatically Ordering Events and Times in Text*,
Studies in Computational Intelligence 677, DOI 10.1007/978-3-319-47241-6_7

that corpus, yielding an overall benefit compared to automatic labelling without any signal information.

To address the cases of tense shifts, Reichenbach's framework of tense was investigated. This included multiple interpretations from the framework to TimeML, including various mappings from tense and aspect pairs into its own tense structure. The framework proposes event and time orderings in simple and complex situations, based on a point-wise temporal logic. The framework also includes capacity for expression of abstract temporal points that is not present in TimeML. Initial validation suggested that the model could be of use for constraining the types of temporal relation between a given linked pair of event verbs. The model's output was added as a feature in a machine learning approach for temporal link labelling, and found to be of some utility in most cases. However, the problem of determining which events and times may be linked through this framework is open, and difficult to solve with existing tools. Critically, no existing resources are available in which this "temporal context" is annotated. A markup acting as an extension to TimeML is proposed for supporting this functionality, as well as supporting reasoning with and annotation for other aspects of Reichenbach's framework.

Overall, an investigation began with analysis of difficult temporal relations. Potential sources of information were identified that could be used to improve automatic system's performance when determining the types of these difficult relations. Of these, two were investigated – explicit temporal signals, and tense – and both exploited in such a way as to improve temporal relation typing. In the course of this exploitation a better understanding of discourse temporal relations and of both phenomena was reached, explained within this book.

7.1 Contributions

The work presented in this book furthered the understanding of some mechanisms used to convey temporal information in language.

7.1.1 Survey of Relations and Relation Typing Systems

Chapter 4 contained a data-driven analysis of temporal relation systems, in an attempt to first identify which relations are the hardest to automatically assign types to, and then to analysis this set of "difficult" links. TempEval-2 was an evaluation exercise where many systems attempted temporal relation labelling over a common data set. The exercise comprises the first analysis of the TempEval-2 participants' performance at relation-level, and the most in-depth analysis of any TempEval exercise.

As well as developing a definition of difficult links and defining a set of those links that are the hardest to automatically label within the TempEval-2 corpus, the chapter presents quantitative and qualitative analyses of the difficult link set. In this set, there were large groups of temporal links using explicit signals and others using tense shifts. These phenomena form the basis of the remainder of the book' investigation.

7.1.2 Temporal Signals

Chapter 5 investigated the role of explicit temporal signals in discourse, with regard to temporal relations. This chapter introduced a method for using signals to achieve a large relation typing performance boost on the temporal links that they co-ordinate. Seeing that signals can be useful, a characterisation of signals is presented, as well as a corpus survey of them. Finding under-annotation in TimeBank, temporal signal annotation guidelines are clarified and an augmented version of TimeBank including extra signals (and, as a result, some extra events, timexes and temporal links) is created. Given evidence for the utility of signals and high-quality ground truth data, the chapter turns to the automatic annotation of temporal signals. This annotation task is split into two sub-parts: signal discrimination (distinguishing temporal from non-temporal uses of signal words) and signal association (finding which timexes or events a given signal co-ordinates). Successful automatic methods for independent signal discrimination and signal association are introduced. These two sub-parts are then joined, in a joint annotation approach, and this approach for signal annotation evaluated, with satisfactory results. Finally, the question of the approach's ability to contribute to the overall temporal relation typing task is addressed. The joint approach is used to label signals and connect them to temporal relations. The results indicate an improvement in temporal relation labelling after this chapter's signal annotations are applied to a document.

7.1.3 Framework of Tense and Aspect

Building on the earlier analysis of difficult links, Chap. 6 introduces a theoretical framework for dealing with tense and aspect – that of [1]. This chapter first introduced tense and the framework, and suggested extensions to the framework to account for positional use of the speech point. Before applying the framework to the temporal relation typing task, it was rational to validate it. This was attempted using a minimal interpretation of the framework, with negative results. Failure analysis led to a new, advanced interpretation, including several novel concepts: an account of progressives; the notion of temporal context (groups within which certain tense rules can be applied); and the discovery that event-event relation typing based on tense suggests relations in semi-interval-link groupings. This advanced interpretation led to the first empirical validation of Reichenbach's framework of tense and aspect. Continuing, techniques for integrating the framework in supervised approaches for event-event and event-timex relation typing were introduced, giving slight benefits over the same approaches without information suggested by the framework. Problems were found with accurately automatically determining temporal context; a lack of context detection limits the applicability of the framework. The chapter closed with the description of a markup language for Reichenbach's framework, integrated with a current temporal annotation schema, in order to further research in this demonstrably promising area.

7.2 Future Work

The book suggests many directions of future work throughout. This section highlights some key points.

7.2.1 Sources of Difficult Links

The failure analysis of temporal relation typing given in Chap. 4 suggests a large number of directions for further investigation. Only two of the problem areas discovered are explored in the rest of the book: signals and tense shifts. Many questions are raised about, for example, the impact that modality, iconicity, world knowledge and textual proximity have upon temporal relations. All these linguistic phenomena are worthy of further investigation, so that their rôle in temporal relation typing might be determined.

Recurrent is the theme of inference: the idea that the configuration of some temporal relations has a constraining impact on the possible configurations of other temporal relations. Temporal closure forms the basic part of this concept, but the role of temporal inference still remains largely unexplored. Approaches that attempt to use it often see only small improvements, though because global temporal constraint is difficult to perform, they have only included reduced-scope models of temporal inference. In an area full of noisy supervised learning output, it would be interesting to see a better integration of global temporal constraints. Prior work on temporal constraint networks [2] has come close to this area. Techniques that can integrate the noisy, uncertain classifier output with global temporal constraints and discourse structure may yield new levels of temporal relation typing performance.

7.2.2 Temporal Signals

While Chap. 5 introduced successful approaches for both annotating temporal signals and exploiting them for temporal relation typing, each of these approaches is a prototype and the first of its kind. There is certainly scope for improvement on each front.

Signal discrimination can be seen as a simplified word sense disambiguation (WSD) task: we are distinguishing temporal from non-temporal uses of expressions. While part-of-speech was shown not to be enough to determine whether or not a given signal was temporal, the approach taken still ignores the majority of the WSD literature [3]. For example, no context is taken into account when performing discrimination. Testing state-of-the-art WSD approaches on this binary classification task may lead to interesting results. Perhaps also the signal discrimination approach given in the chapter may contribute to some WSD tasks.

Signal association is a non-trivial task, and the approach given has some intrinsic limitations. For example, with the best-performing approach, only interval pairs within a certain number of sentences of each other are considered. This is shown by data from the corpus to already exclude some relations where the pair of intervals lie far apart. Other approaches to signal association, perhaps incorporating different discourse relations or some knowledge of pragmatics, may remove these boundaries and lead to increased performance.

Spatial and temporal signals are shown to have a lot in common. Spatial signals also seem to be critical in description of some spatial relations. It follows that the approach detailed in this book may be mapped without too much difficulty to the problem of automatically annotating spatial signals, and perhaps even to using them in automatic spatial relation typing [4].

Finally, given the success of the signal annotation approach and the lack of signal annotation capability in current temporal annotation tools (e.g. [5]), a next logical step is to package the techniques developed during the course of this book into a distributable tool for temporal signal annotation.

7.2.3 Reference Time and Temporal Context

The work presented on Reichenbach's framework, and the new evaluation of its validity, progress many existing problems in computational linguistics concerning the management and interpretation of time in discourse. The chapter presents a big problem: that of determining temporal context. Clearly this is a direction for further work, marshalling current progress in discourse segmentation, syntactic analysis and the behaviour of temporal expressions. The results in this chapter suggest that automatically understanding temporal context permits accurate event-event and event-time relation typing.

However, temporal context is not the sole avenue for further research based on Reichenbach's framework. Multiple problems have called for a means of determining and reasoning with reference time. Aside from the temporal relation typing task, timex normalisation (interpreting an expression of a time) and story generation both require nuanced temporal reasoning, including awareness of the reference point.

Some existing temporal expression normalisation systems heuristically approximate reference time. GUTime [6] interprets the reference point as “the time currently being talked about”, defaulting to document creation date. Over 10 % of errors in this system were directly attributed to having an incorrect reference time, and correctly tracking reference time is the only way to resolve them. TEA [7] approximates reference time with the most recent timex temporally (as opposed to textually) before the expression being evaluated, excluding noun-modifying temporal expressions; this heuristic yields improved performance in TEA when enabled, showing that modelling reference time helps normalisation. HeidelTime [8] uses a similar approach to TEA but does not exclude noun-modifying expressions.

The model is of use when generating language, for determining which tense to use. In fact, it is necessary to consider abstract temporal entities such as the reference point in order to know when to shift tense and how to properly describe events in other temporal frames of reference. A formal application of the model as it extends TimeML may prove useful to accurate language generation. Elson [9] describes how to relate events based on a “perspective” which is calculated from the reference and event times of an event pair. The authors construct a natural language generation system that requires accurate reference times in order to correctly write stories.

Portet [10] found reference point management critical to medical summary generation, in a situation where many small reports were generated with shifting speech and reference points, in order to helpfully unravel the meanings of tense shifts in minute-by-minute patient reports.

The WikiWars corpus of TIMEX2 annotated text prompted the comment that there is a “need to develop sophisticated methods for temporal focus tracking if we are to extend current time-stamping technologies” [11]. Resources that explicitly annotate reference time will be direct contributions to the completion of this task.

A computational model of the sequence of tenses may offer improvements in automatic machine translations. This is because accurately capturing temporal context permits more precise “analytical interlingual translation” [12].

There is also demand in journalism for changing a stock wire articles between present, past and anterior past, in order to suit a particular outlet’s style guidelines. This mood switching can be accomplished using Reichenbach’s framework.

Finally, the problem of datestamping documents automatically is not trivial. Reichenbach’s framework provides the notion of speech time and means of bounding using permanence of the reference point between same-context events and attachment of events to fixed times via positional use of the reference point with a document’s timexes. The model may therefore provide insights into this problem.

In summary, automatic determination of reference time for verbal expressions is an open problem, the solution of which is useful for a number of computational language processing tasks.

References

1. Reichenbach, H.: The tenses of verbs. In: Elements of Symbolic Logic. Dover Publications (1947)
2. Dechter, R., Meiri, I., Pearl, J.: Temporal constraint networks. Artif. Intell. **49**(1), 61–95 (1991)
3. Navigli, R.: Word sense disambiguation: a survey. ACM Comput. Surv. **41**(2), 1–69 (2009)
4. Kordjamshidi, P., Van Otterlo, M., Moens, M.: Spatial role labeling: towards extraction of spatial relations from natural language. ACM Trans. Speech Lang. Process. (TSLP) **8**(3), 4 (2011)
5. Verhagen, M., Mani, I., Sauri, R., Knippen, R., Jang, S., Littman, J., Rumshisky, A., Phillips, J., Pustejovsky, J.: Automating temporal annotation with TARSQI. In: Proceedings of the ACL 2005 on Interactive Poster and Demonstration Sessions, p. 84. Association for Computational Linguistics (2005)
6. Mani, I., Wilson, G.: Robust temporal processing of news. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, pp. 69–76. Association for Computational Linguistics (2000)
7. Han, B., Gates, D., Levin, L.: From language to time: a temporal expression anchorer. In: Proceedings of the 13th International Symposium on Temporal Representation and Reasoning (TIME) (2006)
8. Strötgen, J., Gertz, M.: HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 321–324. Association for Computational Linguistics (2010)
9. Elson, D., McKeown, K.: Tense and aspect assignment in narrative discourse. In: Proceedings of the Sixth International Conference in Natural Language Generation (2010)
10. Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., Sykes, C.: Automatic generation of textual summaries from neonatal intensive care data. Artif. Intell. **173**(7–8), 789–816 (2009)
11. Mazur, P., Dale, R.: WikiWars: a new corpus for research on temporal expressions. In: Proceedings of the EMNLP, pp. 913–922 (2010)
12. Horie, A., Tanaka-Ishii, K., Ishizuka, M.: Verb temporality analysis using Reichenbach's tense system: Towards interlingual MT. In: Proceedings of the International Conference on Computational Linguistics, pp. 471–482. Association for Computational Linguistics (2012)