

# **Studies in Computational Intelligence**

Volume 677

## **Series editor**

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland  
e-mail: [kacprzyk@ibspan.waw.pl](mailto:kacprzyk@ibspan.waw.pl)

### *About this Series*

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/7092>

Leon R.A. Derczynski

# Automatically Ordering Events and Times in Text

 Springer

Leon R.A. Derczynski  
Department of Computer Science  
The University of Sheffield  
Sheffield  
UK

ISSN 1860-949X                      ISSN 1860-9503 (electronic)  
Studies in Computational Intelligence  
ISBN 978-3-319-47240-9              ISBN 978-3-319-47241-6 (eBook)  
DOI 10.1007/978-3-319-47241-6

Library of Congress Control Number: 2016953285

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To Nanna Inie; heartfelt thanks for your  
attention, focus and passion.*

# Foreword

I am delighted to be able to write a few words of introduction to this new book on time and language. It is published at a very important time, in the midst of an explosion in artificial intelligence, where humans, hardware, data, and methods have combined at a fantastic rate to help not only us, but also our tools and computers, better understand our world.

Across the globe, in almost every language we encounter, we discover that we have evolved the ability to reason about time. Terms such as ‘now’ and ‘tomorrow’ describe regions of time; other terms reference events, such as ‘opened’ or ‘hurricane’. This ability to refer to times or to events through language is important and gives humans much great ability in planning, storytelling, and describing the world around us. However, referring to events and times is not quite enough—we also need to be able to describe how these pieces all fit together, so that we can say when an event, like the ‘hurricane’, happened. This temporal structure can be thought of being built from relations that link each event and each time like a net. These temporal relations are encoded in the way we use language around events and times. Discovering how that code works, and what temporal relations a text is communicating to us, is the key to understanding temporal structure in texts.

Traditionally, computational linguistics—the study of computational techniques for language—has given the tools used to address automatic extraction of temporal information from language. Temporal information extraction typically involves identifying events, identifying times, and trying to link them all together, following patterns and relations in the text. One of the harder parts of this extraction process is linking together of events and times, to understand temporal structure. There have been many clever approaches to the task, from scholars and researchers in industry around the world. It is so hard that there has been, and still is, a long-running set of shared exercises, just for this: the TempEval challenges. The first of this series was proposed almost a decade ago in 2006 by me and my collaborators, which we started in order to advance temporal semantic annotation and the plethora of surrounding tasks.

Later, it was actually through one of these TempEval tasks that I first met Dr. Derczynski, and thereafter over many coffees and late dinners at venues like

LREC, or ISA, the semantic annotation workshop. Since, we have collaborated on temporal information extraction, co-organizing more recent TempEval tasks. Our current forthcoming work is a full-length textbook with Marc Verhagen on temporal information processing, with plenty of examples and thorough discussion of the multitude of issues in this fascinating and open area of science.

However, despite our and the community's years of work, and the heavy focus of many researchers through shared task series such as TempEval and i2b2, the problem of extracting temporal structure remains one of the hardest to solve in extracting temporal structure, and also the most important. Clearly, some fresh knowledge is needed.

This book adopts a different tactic to many others' research and describes a data-driven approach to addressing the temporal structure extraction problem. Based on a temporal relation extraction exercise involving systems submitted by researchers across the world, the easy and difficult parts of temporal structure are separated. To tell us where the hardest parts of the problem are, there is an analysis of the temporal relations that few or even none of the systems get right. Part of this analysis then attributes to various sources of linguistic information regarding temporal structure. Each source of information is drawn from a different part of linguistics or philosophy, incorporating ideas of, for example, Vendler, Reichenbach, Allen, and Comrie. The analysis then drives into the later parts of the book, where different sources of temporal structure information are examined in turn. Each chapter discussing a source of this information goes on to present methods for using it in automatic extraction, and bringing it to bear on the core problem: getting the structure of times and events in text.

My hope with this line of work is that it will bring some new knowledge about what is really going on with how temporal relations related to language. We can see the many types of qualitative linguistic theoretical knowledge compared with the hard reality of computational systems' outputs of temporal relations, and firm links emerge between the two. For example, we see links between iconicity—the textual order of elements in a document—and temporal ordering; or, an elegant validation of Reichenbach's philosophically based tense calculus, which, by including the progressive, ends up at Freksa's formal semi-interval logic almost by accident, while continuing to be supported by corpus evidence.

Bringing together all these threads of knowledge about time in language, while coupling them with empirically supported methods and evidence from the data that we have, has been a fruitful activity. This book advances work on some big outstanding problems, raising many interesting research questions along the way for both computer science and linguistics. Most importantly, it represents a valuable contribution to temporal information extraction, and thus to our overall goal: understanding how to process our human language.

June 2016

James Pustejovsky  
TJX/Feldberg Chair of Computer Science  
Department of Computer Science  
Volen Center for Complex Systems  
Brandeis University  
Arlington, MA

# Acknowledgements

A very special thanks to Robert Gaizauskas for his extensive help and guidance at many points; and to Yorick Wilks and Mark Steedman for their comments on an earlier version. The whole could not have been possible without the vision for the field and vast groundwork laid describing time in language, which is due to in great part to James Pustejovsky, as is gratitude for the foreword.

Finally, the book was produced during a period where I received support from, in no particular order: the EC FP7 project TrendMiner, the EC FP7 project PHEME, the EC H2020 project Comrades, an EPSRC Enhanced Doctoral Training Grant, the University of Sheffield Engineering Researcher Society, and the CHIST-ERA EPSRC project uComp.



# Contents

<b>1 Introduction</b> . . . . .	1
1.1 Setting the Scene . . . . .	1
1.2 Aims and Objectives . . . . .	4
1.3 New Material in This Book . . . . .	5
1.4 Structure of the Book . . . . .	6
References . . . . .	7
<b>2 Events and Times</b> . . . . .	9
2.1 Introduction . . . . .	9
2.2 Events . . . . .	10
2.2.1 Types of Event . . . . .	10
2.2.2 Schema for Event Annotation . . . . .	11
2.2.3 Automatic Event Annotation . . . . .	12
2.3 Temporal Expressions . . . . .	15
2.3.1 Temporal Expression Types . . . . .	15
2.3.2 Schema for Timex Annotation . . . . .	17
2.3.3 Automatic Timex Annotation . . . . .	20
2.4 Chapter Summary . . . . .	22
References . . . . .	22
<b>3 Temporal Relations</b> . . . . .	25
3.1 Introduction . . . . .	25
3.2 Temporal Relation Types . . . . .	26
3.2.1 A Simple Temporal Logic . . . . .	27
3.2.2 Temporal Interval Logic . . . . .	28
3.2.3 Reasoning with Semi-intervals . . . . .	29
3.2.4 Point-Based Reasoning . . . . .	31
3.2.5 Summary . . . . .	31
3.3 Temporal Relation Annotation . . . . .	32
3.3.1 Relation Folding . . . . .	33
3.3.2 Temporal Closure . . . . .	36
3.3.3 Open Temporal Relation Annotation Problems . . . . .	37

3.4	Automatic Temporal Relation Typing. . . . .	38
3.4.1	Closure for Training Data. . . . .	38
3.4.2	Global Constraints. . . . .	39
3.4.3	Task Description. . . . .	40
3.4.4	Evaluation. . . . .	40
3.5	Prior Relation Annotation Approaches . . . . .	43
3.5.1	Feature and Classifier Engineering . . . . .	44
3.5.2	Rule Engineering . . . . .	46
3.5.3	Syntactic and Semantic Information . . . . .	48
3.5.4	Linguistic Context. . . . .	49
3.5.5	Global Constraint Satisfaction. . . . .	50
3.5.6	Summary. . . . .	51
3.6	Analysis . . . . .	54
3.6.1	Data Sparsity. . . . .	54
3.6.2	Moving Beyond the State of the Art. . . . .	55
3.7	Chapter Summary. . . . .	56
	References. . . . .	56
<b>4</b>	<b>Relation Labelling Analysis</b> . . . . .	<b>63</b>
4.1	Introduction . . . . .	63
4.2	Survey of Difficult TLINKs . . . . .	64
4.2.1	The TempEval Participant Dataset . . . . .	64
4.2.2	Defining What Constitutes “difficult” Temporal Links . . . . .	66
4.2.3	Comparative Distribution Analysis . . . . .	71
4.2.4	Attribute Distribution Summary . . . . .	74
4.3	Extra-Feature Analysis . . . . .	76
4.3.1	Characterisation. . . . .	77
4.3.2	Analysis . . . . .	78
4.3.3	Signals Versus Tense Shifts . . . . .	79
4.3.4	Extra-Feature Analysis Summary . . . . .	81
4.3.5	Next Directions. . . . .	81
4.4	Analysing TLINKs Through Dataset Segmentation . . . . .	81
4.4.1	Core Approach . . . . .	82
4.4.2	Theoretical Assumptions. . . . .	82
4.5	Chapter Summary. . . . .	83
	References. . . . .	84
<b>5</b>	<b>Using Temporal Signals</b> . . . . .	<b>85</b>
5.1	Introduction . . . . .	85
5.2	The Language of Temporal Signals . . . . .	86
5.2.1	Related Work . . . . .	87
5.2.2	Signals in TimeML. . . . .	88
5.3	The Utility of Temporal Signals. . . . .	89
5.3.1	Introducing Signals to the Relation Labelling Feature Set . . . . .	92

- 5.3.2 TLINK Typing Results Using Signals. . . . . 93
- 5.3.3 Utility Assessment Summary . . . . . 95
- 5.4 Corpus Analysis . . . . . 95
  - 5.4.1 Signals in TimeBank. . . . . 96
  - 5.4.2 Relation Type Ambiguity . . . . . 97
  - 5.4.3 Temporal Versus Non-temporal Uses . . . . . 99
  - 5.4.4 Parallels to Spatial Representations in Natural Language . . . . . 100
- 5.5 Adding Missing Signal Annotations . . . . . 101
  - 5.5.1 Preliminary Signal Discrimination. . . . . 102
  - 5.5.2 Clarifying Signal Annotation Guidelines. . . . . 102
  - 5.5.3 Curation Procedure . . . . . 103
  - 5.5.4 Signal Re-Annotation Observations. . . . . 103
  - 5.5.5 TB-Sig Summary . . . . . 108
- 5.6 Signal Discrimination . . . . . 110
  - 5.6.1 Problem Definition . . . . . 110
  - 5.6.2 Method . . . . . 110
  - 5.6.3 Discrimination Feature Extraction . . . . . 110
  - 5.6.4 Discrimination Evaluation . . . . . 115
  - 5.6.5 Discrimination on Unseen Data. . . . . 119
  - 5.6.6 Summary. . . . . 121
- 5.7 Signal Association . . . . . 122
  - 5.7.1 Problem Definition . . . . . 122
  - 5.7.2 Method . . . . . 122
  - 5.7.3 Dataset . . . . . 125
  - 5.7.4 Automatic Association Evaluation. . . . . 126
  - 5.7.5 Association Summary . . . . . 129
- 5.8 Overall Signal Annotation . . . . . 130
  - 5.8.1 Joint Annotation Task . . . . . 130
  - 5.8.2 Combined Signal Annotation and Relation Typing. . . . . 133
- 5.9 Chapter Summary . . . . . 134
- References. . . . . 135
- 6 Using a Framework of Tense and Aspect . . . . . 139**
  - 6.1 Introduction . . . . . 139
  - 6.2 Timelines in Language . . . . . 141
  - 6.3 Description of the Framework . . . . . 144
    - 6.3.1 Time Points. . . . . 144
    - 6.3.2 Reichenbachian Tenses . . . . . 144
    - 6.3.3 Verb Interactions. . . . . 145
    - 6.3.4 Temporal Context . . . . . 148
    - 6.3.5 Quoted Speech . . . . . 150
    - 6.3.6 Limitations of the Framework. . . . . 150

6.4	Validating the Framework Against TimeBank . . . . .	153
6.4.1	Minimal Interpretation of Reichenbach’s Framework . . . . .	155
6.4.2	Advanced Interpretation of Reichenbach’s Framework . . . . .	158
6.5	Applying Reichenbach’s Framework to Temporal Relation Typing . . . . .	164
6.5.1	Same Context Event-Event Links . . . . .	165
6.5.2	Same Context Event-Timex Links. . . . .	167
6.5.3	Summary. . . . .	169
6.6	Annotating Reichenbach’s Framework . . . . .	169
6.6.1	Motivation for Annotating the Framework’s Points. . . . .	170
6.6.2	Proposed Solution. . . . .	170
6.6.3	Special RTMLINKs . . . . .	172
6.6.4	Example RTMML. . . . .	173
6.7	Chapter Summary . . . . .	175
	References. . . . .	175
<b>7</b>	<b>Conclusion</b> . . . . .	<b>179</b>
7.1	Contributions . . . . .	180
7.1.1	Survey of Relations and Relation Typing Systems . . . . .	180
7.1.2	Temporal Signals . . . . .	181
7.1.3	Framework of Tense and Aspect. . . . .	181
7.2	Future Work. . . . .	182
7.2.1	Sources of Difficult Links . . . . .	182
7.2.2	Temporal Signals . . . . .	182
7.2.3	Reference Time and Temporal Context. . . . .	183
	References. . . . .	185
	<b>Appendix A: Annotated Corpora and Annotation Tools</b> . . . . .	<b>187</b>
	<b>Appendix B: RTMML Reference.</b> . . . . .	<b>199</b>
	<b>Appendix C: CAVaT Reference.</b> . . . . .	<b>203</b>
	<b>Index</b> . . . . .	<b>205</b>

# List of Figures

Figure 3.1	Temporal graph of a simple story . . . . .	30
Figure 3.2	TLINK relation type assignment difficulty increases with the distance between link arguments . . . . .	45
Figure 4.1	Frequencies of event attribute values in the TempEval-2 English test data . . . . .	65
Figure 4.2	Proportions missing events attribute values in the TempEval-2 English test data . . . . .	66
Figure 4.3	Frequencies of timex attribute values in the TempEval-2 English test data . . . . .	67
Figure 4.4	TempEval-2 relation labelling tasks, showing the proportion of relations organised by number of systems that failed to label them correctly . . . . .	68
Figure 4.5	Composition of the set of difficult links . . . . .	70
Figure 4.6	Proportion of each TempEval-2 task’s links that are difficult . . . . .	70
Figure 4.7	Comparative analysis of features for TempEval-2 task E. . . . .	72
Figure 4.8	Comparative analysis of features for task F, relating events with their subordinate events . . . . .	73
Figure 4.9	Comparative analysis of features for TempEval-2 task C. . . . .	75
Figure 4.10	Comparative analysis of features for task D, relating events to DCT. . . . .	76
Figure 5.1	Signalled TLINKs by argument type (event-event or event-tlink) in TimeBank 1.2 and the AQUAINT TimeML corpus. . . . .	94
Figure 5.2	An example SBAR-TMP construction around a temporal signal. . . . .	99
Figure 5.3	An example of the common syntactic surroundings of a <i>before</i> signal . . . . .	106
Figure 5.4	Typical mis-interpretation of a spatial (e.g. non-temporal) usage of <i>before</i> . . . . .	106

Figure 5.5 Example of a non-annotated signal (*former*)  
from TimeBank’s `wsj_0778.tml` . . . . . 107

Figure 5.6 Example of an SBAR-TMP where the first child is a  
signal qualifier (*several months*) and the second child  
the signal word itself (*before*) . . . . . 112

Figure 6.1 An example of permanence of the reference point. . . . . 147

Figure 6.2 Error reduction in SCEE links with and without  
features representing permanence of the reference point,  
modelling temporal context as same-sentence . . . . . 166

Figure 6.3 Comparative performance on labelling event-time links  
where the time positions the reference point . . . . . 168

Figure A.1 Automatically annotating text with TTK . . . . . 193

Figure A.2 Manually annotating text with Callisto . . . . . 194

Figure A.3 Overseeing a BAT annotation project . . . . . 195

Figure B.1 RTMML for a passage from David Copperfield . . . . . 200

# List of Tables

Table 3.1	Allen’s temporal interval relations . . . . .	28
Table 3.2	TimeML temporal relations . . . . .	32
Table 3.3	The relation set used in TempEval and TempEval-2 . . . . .	33
Table 3.4	Relation folding mappings used in this book . . . . .	34
Table 3.5	Distribution of TLINK relation types in TimeBank 1.2 . . . . .	35
Table 3.6	Distribution of relation types over TimeBank 1.2, as per Table 3.5 and folded using the mappings in Table 3.4 . . . . .	35
Table 3.7	Prior work on automatic temporal relation classification . . . . .	53
Table 4.1	Proportion of difficult links in each TempEval-2 task . . . . .	68
Table 4.2	Error rates in TempEval-2 Task C, event-timex linking . . . . .	69
Table 4.3	Error rates in TempEval-2 Task D, event-DCT linking . . . . .	69
Table 4.4	Error rates in TempEval-2 Task E, linking main events of subsequent sentences . . . . .	69
Table 4.5	Error rates in TempEval-2 Task F, linking events to events that they subordinate . . . . .	70
Table 4.6	Temporal ordering phenomena and their occurrence in difficult links . . . . .	78
Table 4.7	Co-occurrence frequencies for temporal signals and tense shifts in event-event difficult links . . . . .	80
Table 4.8	Co-occurrence frequencies for temporal signals and tense shifts in all TimeBank v1.2’s event-event links . . . . .	80
Table 5.1	A sample of phrases most likely to be annotated as a signal when they occur in TimeBank . . . . .	89
Table 5.2	Signal expressions and the TimeML relations that they can denote . . . . .	90
Table 5.3	TLINKs and signals in the largest TimeML-annotated corpora . . . . .	91
Table 5.4	Results from replicating a prior experiment on automatic relation typing of event-event relations . . . . .	92
Table 5.5	TLINK classification with and without signal features . . . . .	93

Table 5.6	Predictive accuracy from Carafe’s maximum entropy classifier, using features that do or do not include signal information, over signalled and non-signalled event-event TLINKs in ATC . . . . .	94
Table 5.7	How <SIGNAL> elements are used in TimeBank. . . . .	96
Table 5.8	The number of TLINKs associated with each temporal signal word/phrase, in TimeBank. . . . .	96
Table 5.9	A closed class of temporal signal expressions . . . . .	98
Table 5.10	Distribution of part-of-speech in signals and the first word of signal phrases . . . . .	99
Table 5.11	The set of signal words and phrases suggested by the SBAR-TMP model . . . . .	99
Table 5.12	Frequency of signal usage for different types of spatial link in the ACE 2005 English SpatialML Annotations Version 2. . . . .	101
Table 5.13	Signal texts that are hard to discriminate; error reduction performance compared to the most common class (“change”) is based on a maximum entropy classifier, trained on TimeBank. . . . .	104
Table 5.14	Frequency of candidate signal expressions in TimeBank and TB-sig. . . . .	109
Table 5.15	Comparison of the effect that decomposing values of the preceding_subtree and following_subtree features has, using our extended feature set and TimeBank data . . . . .	114
Table 5.16	Performance of four constituent-tag based baselines over TimeBank . . . . .	115
Table 5.17	Signal discrimination performance on the plain TimeBank corpus. . . . .	117
Table 5.18	Signal discrimination performance on the curated corpus . . . . .	118
Table 5.19	Signal discrimination performance on the TimeBank corpus, with an extended feature set . . . . .	118
Table 5.20	Sample features useful for signal discrimination, based on our curated TimeBank data, TB-sig . . . . .	120
Table 5.21	Characteristics of the N45 section of the AQUAINT TimeML corpus, before and after signal curation . . . . .	121
Table 5.22	Performance of a TB-sig trained signal discriminator on unseen data. . . . .	121
Table 5.23	Distribution of sentence distance between intervals linked by a signal, for TB-sig. . . . .	125
Table 5.24	Performance at the signal:interval association task, with 5-fold cross validation . . . . .	127
Table 5.25	Performance at the signal:interval-pair association task . . . . .	128



Table 5.26	Confusion matrix for signal association performance with a MaxEnt classifier on TimeBank with a window including the signal sentence and two preceding ones . . . . .	128
Table 5.27	Distribution of sentence distance between intervals and signal that links them . . . . .	129
Table 5.28	Performing of a TB-sig trained signal associator on unseen data. . . . .	129
Table 5.29	Details of the joint approach to signal annotation . . . . .	131
Table 5.30	Sample signals and arguments found in N45. . . . .	131
Table 5.31	Details of the constrained joint approach to signal annotation . . . . .	132
Table 5.32	TLINK stats over corpora used for extrinsic evaluation. . . . .	133
Table 5.33	Training dataset sizes from TB-sig used for signal annotation models . . . . .	133
Table 5.34	TLINK labelling accuracy over corpora used for extrinsic evaluation . . . . .	134
Table 6.1	Frequency of TimeML tense and aspect on verb events in TimeBank . . . . .	140
Table 6.2	Reichenbach's tenses; from [16] . . . . .	145
Table 6.3	Minimal schema for mapping TimeML event tense and aspects to Reichenbach's framework. . . . .	156
Table 6.4	Event orderings based on the Reichenbachian tenses that are available in TimeML . . . . .	156
Table 6.5	Accuracy of Reichenbach's framework with a subset of links manually annotated for being tensed verbs in the same temporal context . . . . .	157
Table 6.6	TimeML tense/aspect combinations, in terms of the Reichenbach framework. . . . .	159
Table 6.7	Example showing disjunctions of TimeML intervals applicable to describe the type of relation between A and B given their tense and aspect . . . . .	160
Table 6.8	Freksa semi-interval relations; adapted from Freksa (1992). . . . .	161
Table 6.9	TimeML tense/aspect pairs with the disjunction of TimeML relations they suggest. . . . .	162
Table 6.10	Consistency of temporal relation types suggested by Reichenbach's framework with ground-truth data . . . . .	163
Table 6.11	Using Reichenbach-suggested event ordering features representing permanence of the reference point. . . . .	165
Table 6.12	Reichenbach-suggested event ordering feature representing permanence of the reference point. . . . .	166
Table 6.13	Performance when using dependency parse and Reichenbach-derived feature . . . . .	168
Table 6.14	RTMML relation types . . . . .	172

# Abstract

The ability to describe the order of events is crucial for effective communication. It is used to describe causality, to plan, and to relay stories. This temporal ordering can be expressed linguistically in a variety of ways. For example, one may use tense to describe the relation between the time of speaking and other events, or use a temporal conjunction to temporally situate an event relative to time. This ordering remains the hardest task in processing time in text. Very sophisticated approaches have yielded only small improvements over initial attempts. This book covers relevant background and discusses the problem, and goes on to conduct an analysis of temporal ordering information. This breaks the types of information used into different groups. Two major sources of information are identified that provide typing information for two segments: relations explicitly described by a signal word, and relations involving a shift of tense and aspect. Following this, the book investigates automatic temporal relation typing in both these segments, presenting results, introducing new methods, and generating a set of new language resources.