A-dur: Action Duration Calculation System.

Marek KRAWCZYK, Yuki URABE, Rafał RZEPKA and Kenji ARAKI
Graduate School of Information Science and Technology, Hokkaido University
{marek, y_urabe, kabura, araki}@media.eng.hokudai.ac.jp

Abstract:
In this paper, we present A-dur, a tool for calculating duration of actions expressed in Japanese language. A-dur was created in order to be implemented into emotional and social consequences text mining system. However it potentially can be used as a stand-alone time distance calculation solution. It utilizes an extendable temporal expression rules database handcrafted to the system's specifications. We also present the results of our preliminary experiments estimating the database accuracy.
Keywords: Temporal Information Processing, temporal expressions, pattern matching

1. Introduction

To the best of the authors' knowledge, the only research concerning temporal expressions interpretation in Japanese language was performed in the late 1990's (Matsushita and Kato, 1999). Since then the topic was neglected, despite of a wide interest of foreign researchers dealing with other languages. The need for a Japanese language temporal expressions analysis and interpretation system is clearly visible while reading recent publications concerning such systems working on other languages: the potential use of such system is extensive and reaches many branches of Natural Languages Processing research.

The system described in this paper was prepared as an element of emotional and social consequences text mining system which is currently being developed by Rzepka and Araki (2012). Its functionality is limited to the role required by the before-mentioned system and therefore it is is far from being a complete solution for the temporal expression analysis and interpretation problem. However its development revealed a rich potential of this sphere of Natural Language Processing and can be considered as a small step towards a fully functional engine comparable to the systems already developed abroad.

2. Previous research

TempEx (Mani and Wilson, 2000), which was further extended to become GUTime (Verhagen et al., 2005), was one of the first systems dealing

with temporal expressions. The system was utilized to process temporal expressions included in news. It recognized and normalized timexes, but it did not deal with durations (e.g. “three weeks”) and fuzzy expressions (e.g. “in a few minutes”). The normalization component of the system detected explicit and implicit, content-dependent timexes. In case of implicit timexes, it differentiated between those relative to the document creation time and those relative to previously expressed timexes.

Another system was called Chronos (Negri and Merseglia, 2004) and it followed TIMEX2 (Ferro et al., 2001) annotation standard. It also differentiated between absolute and relative timexes. The latter were divided into those relative to the document creation time and to previously marked time-point.

A system that improved upon Chronos was named TERSEO (Saquete et al., 2006). It also used TIMEX2 standard, but additionally it accounted for period and fuzzy expressions. The system included a module for an automatic extension of rules to multiple languages.

DANTE (Mazur and Dale, 2007) was a rule-based system using TIMEX2 standard. Unspecified timexes differentiation and their local semantics definition was the distinguishing feature of this system.

More recent systems focus on the TIMEX3 (Pustejovsky et al., 2003) standard. Those which gave the best results were all rule-based.

One of such systems was called HeidelTime (Strotgen and Gertz, 2010). It covered explicit, durative, implicit and relative timexes. The system reached the best score (85%) of normalization performance at Temp-Eval-2 (Verhagen et al., 2010) international evaluation.

TIPSem (Llorens et al., 2010) addressed the problem of normalization with a hybrid strategy. First a classifier specifies the type of normalization (explicit, relative to document creation time, relative to previous statement, duration, set or vague). Next, depending on the selected normalization, a proper pattern matching-based rules are applied.

TERNIP (Northwood, 2010) was created on the basis of GUTime. The system utilizes system-independent rule database as well as complex syntax patterns.

3. State of the art

The most recent systems continue to use TIMEX3 format.

TIMEN (Llorens et al., 2012) was created with an assumption that the system's architecture should be modular: the algorithms conducting the normalization are to be separated from the knowledge and rules necessary for the process. This way the system is flexible enough to be adaptable to work with other languages. The system operates in the following manner: first, the target time expression is selected together with some contextual data, next the phrase is converted to a symbolic representation with the use of the provided knowledge base. A set of

rules is then matched against the representation. Finally the system normalizes the output following TIMEX3 standard.

SUTime (Chang and Manning, 2012), a rule-based temporal tagger, was built on regular expression patterns over tokens. The system recognizes time, duration, interval and set in accordance with TIMEX3 standards. It also detects nested time expressions and duration ranges. In order to recognize temporal expressions, the system applies three types of rules: text regex rules (mappings from simple regular expressions over characters or tokens to temporal representations), compositional rules (mappings from regular expressions over chunks, consisting of both tokens and temporal objects, to temporal representations), and filtering rules (to remove ambiguous expressions that are likely to not be temporal expressions). After recognizing all temporal expressions, the system converts relative times to an absolute time, which is then presented in a TIMEX3 annotation.

4. System overview

The proposed system utilizes regex-based rules set consisting of 497 items handcrafted on the basis of Japanese temporal expressions list. The list was created with the help of JUMAN – Japanese morphological analysis system[1]. JUMAN was fed with a fragment of Ameba corpus (Maciejewski et al., 2010); the result of this operation was filtered in search of linguistic items labeled as temporal expression by the analyzer and arranged in a reusable knowledge database. The obtained list entries were enriched with temporal expression type tag and time-line value. There were four types of tags established: 'point in time', 'duration', 'shift' and 'not applicable'. 'Point in time' tag is self-explanatory- it refers to expressions describing a point markable on a time-line, for example “今朝” (“this morning”), “正午” (“noon”) etc; also some expressions commonly regarded as denoting timespan were included in this category, for example “夏” (“summer”), for reasons explained later in this paper. 'Duration' tag is attached to expressions describing a particular timespan, for example “年間” (“a year”) or “週間” (“a week”). 'Shift' tag represents expressions that do not hold time value by themselves, but modify the accompanied point in time expressions, for example “初め” (“the beginning”). 'Not applicable' tag accompanies temporal expressions that because of their ambiguous nature can not be utilized by the current version of the system. Also explicit, numerical time expressions, such as “14:45” are currently not processed by the system.

The time-line value that was manually added to each expression in the database is a time distance between the time point described by the given temporal expression and the conventional 'zero' point set up for July 2nd 12:00 AM. July 2nd is the 183rd day of the year, which is the midpoint

1 http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN

of a common year. If the temporal expression on the list refers to a time point placed after the 'zero' point on the time-line, a positive value is assigned (counted by the number of days), and analogically, expressions describing earlier time-points have negative value assigned. For example “朝” (“morning”) (considered as 6:00AM) is accompanied by [-0.25] time-line value, and “明日” (“tomorrow”) is marked with [1] time-line value.

The reason for including some of temporal expressions referring to timespan into the 'point in time' category comes from the main purpose of the system: to determine the duration of action described by an input sentence. If the input sentence includes expressions such as “夏から冬まで” (“from summer till winter”), those two timespan expressions are treated as point in time expressions and the distance between them is calculated in order to determine the duration of a given action. In this case the “夏” (“summer”) time-point is set up for July 15th 12:00 AM and the “冬” (“winter”) time-point is set up for January 15th 12:00 AM.

The authors are aware of the fact that many temporal expressions may be perceived differently depending on the region, for example the before mentioned seasons may start and finish at different dates in Okinawa and Hokkaido. Therefore we asked the third author born in Aichi prefecture to be the judge in those ambiguous cases and decide about the adjustment of the most universal time-line values to the database items.

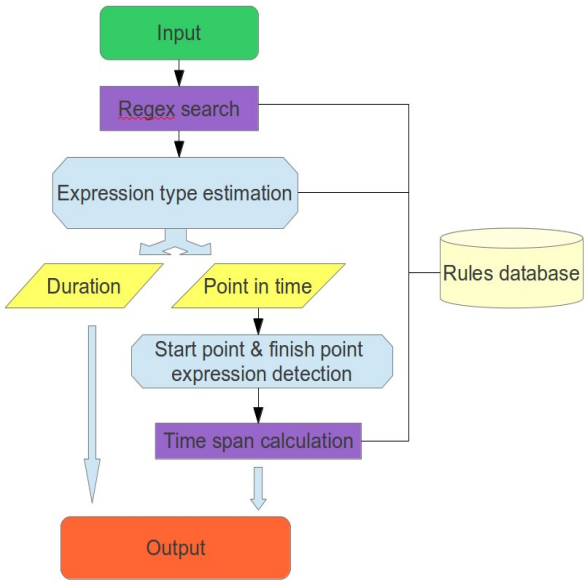

Figure 1: A-dur algorithm outline

The system works in the following manner: first it searches for a match between the input sentence (or part of the input) and temporal expressions rules database. If a match is found, the system takes into account numerals appearing before those expressions to determine the proper time-point or timespan multiplier. Next a type of temporal expression is determined. If the type is 'point in time', the program checks whether the expression is accompanied by the 'beginning' and 'end' markers, such as “から” (“from”) and “まで” (“until”). If a pair if these markers is found, the system assigns the time-line values corresponding to these expressions to two

variables and outputs the calculated difference between those two variables. In case of input sentence containing only one of the 'beginning' and 'end' markers, the timespan is not calculated. In case of 'duration' type temporal expression being found, the system outputs the time-value associated with the particular temporal expression.

5. Experiment

Currently the system's algorithm and functionality is still under development in order to create a comprehensive temporal expressions normalization solution. However the used rules database should be complete and correct enough so as to not introduce unnecessary errors while further developing the code. That is why we decided to make an experiment aiming at evaluating the content of the database and use the results to increase the database quality. The experiment would create a basis for a routine evaluation, correction and expansion of the database to be regularly implemented in the future.

For the experiment we took 20 randomly selected items from the rules database. The items consisted of Japanese temporal expression, type tag and time-line value. We asked eleven participants to evaluate whether the proposed tag and time-line value is correct or not. If not, we asked the participants to provide a correct alternative.

6. Results

The overall score of items evaluated as correct reached 77.2%. Only two items were unanimously evaluated as correct, seven items received a score of 90% and one of 80%. The rest of the items received scores between 72% and 45%.

Table 1: Evaluation results

Score	100%	90%	81%	72%	63%	54%	45%
No.of items	2	7	1	5	1	2	2

A large majority of answers pointing to incorrect items were referring to the temporal expression tags. Many respondents indicated an ambiguous nature of these items and suggested that assignment of some temporal expressions to one particular category can be difficult or even impossible without context. The rest of the suggested corrections referred to erroneous time-line values.

The conclusion from the above results is following: the proposed system will work perfectly fine while dealing with great majority of standard, explicit temporal expressions. In order to reduce the 22.8% error rate, future correction and expansion of the database has to be performed in parallel with further development of the system's functionality, mainly with implementing detection and normalization of implicit, content-dependent

timexes.

7. Conclusion and Future Work

In this paper, we presented A-dur, a tool for calculating duration of actions expressed in Japanese language. By describing functionality of previously developed methods we showed the goals we would like to achieve with further development of the system. We also presented the results of our preliminary experiments estimating the database accuracy.

Our predecessors' work clearly showed the necessity for timex analysis systems and indicated their potential use. Automated analysis of temporal expression is considered critical for accurate processing of discourse semantics (Mani and Wilson, 2000). It is also vital for other language processing applications, such as question answering (Saquete et al. 2009), information retrieval (Alonso et al., 2007), text summarization (Daniel et al., 2003) and knowledge base population (Ji et al., 2011).

While analyzing the functionality of the proposed system in comparison with English language temporal expression normalization systems it is clearly visible that A-dur requires further development. In the nearest future we would like to add the following features to the system: detection and normalization of implicit, content-dependent timexes and fuzzy expressions, as well as utilization of TIMEX3 standard. After implementing these initial modifications, we will use the distinguishing features of the systems created during the last 13 years as guidelines we would like to follow.

References

Omar Alonso, Michael Gertz, and Ricardo Baeza-Yates. 2007. On the Value of Temporal Information in Information Retrieval. *SIGIR Forum*, 41(2):35–41.

Naomi Daniel, Dragomir Radev, and Timothy Allison. 2003. Sub-event based multi-document summarization. In *HLT-NAACL Text summarization workshop*, pages 9–16. ACL.

Angel Chang and Chris Manning. 2012. SUTIME: A library for recognizing and normalizing time expressions. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 3735–3740, European Language Resources Association (ELRA).

Lisa Ferro, Inderjeet Mani, Beth Sundheim and George Wilson. 2001. TIDES Temporal Annotation Guidelines Version 1.0.2. Mitre Technical Report.

Heng Ji, Ralph Grishman, Hoa Trang Dang, Xuansong Li, Kira Griffit, and Joe Ellis. 2011. Overview of the TAC2011 Knowledge Base Population Track. In *Proceedings of the Text Analytics Conference*, Gaithersburg, MD, USA.

Hector Llorens, Estela Saquete, and Borja Navarro- Colorado. 2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic

Roles in TempEval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 284–291. ACL.

Hector Llorens, Leon Derczynski, Robert Gaizauskas, and Estela Saquete. 2012. TIMEN: An Open Temporal Expression Normalisation Resource. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA).

Jacek Maciejewski, Michal Ptaszynski, Pawel Dybala. 2010. Developing a Large-Scale Corpus for Natural Language Processing and Emotion Processing Research in Japanese. In *Proceedings of the International Workshop on Modern Science and Technology(IWMST)*, pages 192–195.

Inderjeet Mani and George Wilson. 2000. Robust temporal processing of news. In *ACL Annual Meeting*, pages 69–76, NJ, USA. ACL.

Mitsumori Matsushita and Tsuneaki Kato. 1999. An Analysis Regarding the Use of Time Expressions to Understand Intention. *NLPRS-99*, pages 233–238, Beijing, China.

Pawel Mazur and Robert Dale. 2007. The DANTE temporal expression tagger. In *Proceedings of the 3rd Language and Technology Conference (LTC)*.

Matteo Negri and Luca Marseglia. 2004. Recognition and Normalization of Time Expressions: ITC-irst at TERN 2004. Technical report, Information Society Technologies.

Chris Northwood. 2010. TERNIP: Temporal Expression Recognition and Normalisation in Python. Master's thesis, University of Sheffield.

James Pustejovsky, Jos Castao, Robert Ingria, Roser Saur, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. TimeML: Robust specification of event and temporal expressions in text. In *Fifth International Workshop on Computational Semantics (IWCS-5)*.

Rafal Rzepka and Kenji Araki. 2012. Polarization of consequence expressions for an automatic ethical judgment based on moral stages theory. *IPSJ SIG Notes 2012-NL-207(14)*, pages 1–4.

Estela Saquete, Rafael Munoz, and Patricio Martinez- Barco. 2006. Event ordering using TERSEO system. *Data Knowledge Engineering*, 58(1):70–89.

Estela Saquete, Jose Luis Vicedo Gonzalez, Patricio Martinez-Barco, Rafael Munoz, and Hector Llorens. 2009. Enhancing QA Systems with Complex Temporal Question Processing Capabilities. *Journal of Artificial Intelligence Reresarch (JAIR)*, 35:775–811.

Jannik Strotgen and Michael Gertz. 2010. HeidelTime: High Quality Rule-Based Extraction and Normalization of Temporal Expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 321–324, Uppsala, Sweden. Association for Computational Linguistics.

Marc Verhagen, Inderjeet Mani, Roser Sauri, Robert Knippen, Seok Bae Jang, Jessica Littman, Anna Rumshisky, John Phillips, and James Pustejovsky. 2005. Automating temporal annotation with TARSQI. In *ACL*, pages 81–84, NJ, USA. ACL.

Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval- 2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 57–62, Uppsala, Sweden. ACL.