



UNIVERSITÀ DEGLI STUDI DI NAPOLI "L'Orientale"

Ph.D in East and South Asia
VII N.S.

Doctoral Dissertation

**Tense, Aspect and the Semantics of Event Description
Towards a Contrastive Analysis of Italian and Japanese**

Ph.D Candidate

Patrizia Zotti

Research Advisor

Professor Paolo Calvetti

Research Coordinator

Professor Giorgio Amitrano

2009-2010

To my family

To my son

Tense, Aspect and the Semantics of Event Descriptions Towards a Contrastive Analysis of Italian and Japanese

Patrizia Zotti

Abstract

Temporal analysis of events is an important part of many natural language processing tasks. We expect events to contain the same temporal information regardless of the language, however, the way temporal information is encoded via lexical or grammatical properties differs from language to language.

While the lexical aspect of event semantics is relatively similar across languages, the grammatical aspect varies greatly, and concepts of tense and aspect may not be shared across languages, preventing a simple correspondence of verbal forms between two languages from being established, causing trouble for both human and machine translators alike.

Divergence in temporal representations is illustrated by the following examples:

- (1) 田中さんは 学校に電車で通っている。

Tanaka san wa gakkō ni densha de kayotte iru

Il sig. Tanaka va a scuola in treno.

‘Mr Tanaka commutes to school by train.’

- (2) 田中さんは只今学校に電車で通っている。

Tanaka san wa tadaima gakkō ni densha de kayotte iru

Il sig. Tanaka ora sta andando a scuola in treno.

‘Mr Tanaka is now commuting to school by train.’

The Italian and English translations use different verbal forms and aspectual configurations, the imperfective (habitual) aspect in (1) and the progressive aspect in (2), whereas both Japanese sentences contain the same verbal form *-te iru* and make use of the adverb *ima* ‘now’ to indicate the progressive aspect. In translating these sentences, we need to be aware of these differences in temporal information representations.

This work represents a first theoretical and empirical contrastive investigation on how temporal information is conveyed in Japanese and Italian with a focus on the semantics of events and how they are reported. We include a classification of eventuality types and their expression in tense and aspect, making this work useful for researchers working on tasks that require detailed, cross-lingual temporal analysis.

We empirically support our theoretical proposal by semi-automatically annotating a parallel Japanese-Italian corpus of novels, parliamentary proceedings and newspapers translations with temporality tags and conducting an investigation into the representation of temporal phenomena across languages.

Keywords:

tense, aspect, temporality tags, Japanese, Italian, Japanese Italian parallel corpus

Acknowledgements

I would like to thank everyone helped me during this work. It is almost impossible to mention all of them and to express all my gratitude.

My first and biggest thanks goes to my advisor, Professor Paolo Calvetti for his guidance and support, for criticism and discussion, for the patience in carefully reading and commenting my drafts, and for his constant encouragement.

I would also like to thank Professor Matsumoto Yuji – Director of the Computational Linguistics Laboratory at NAIST, Japan – and assistant Professor Asahara Masayuki. I'm very grateful to Professor Matsumoto for welcoming me in the lab when I was unfamiliar with computational linguistics, and for allowing me to spend two years there. I'm very thankful to professor Asahara Masayuki too, for opening my eyes to totally new perspectives, and for his guidance and support in designing and developing the tagging procedure and the computational task.

I'm very grateful to Professor Silvio Vita – Director of ISEAS (Italian School of East Asian Studies) for motivating, helping and supporting me in more than two years in Japan. I want to thank Yamamoto Makimi (ISEAS) too, for her kindness and help in everyday life.

My deepest gratitude goes to Shoyu Club for supporting this research in Japan for two years. I'm very thankful to its Executive Director, Daigo Tadahisa, for patiently reading all my periodic reports and for his questions and comments.

Many thanks goes to Dr. Gianluca Coci from University of Turin, Dr. Laura Testaverde and Dr. Alessandro Clementi, for helping in the corpus development providing the pre-print versions of their Japanese novels' translations into Italian.

I would like to thank Professor Franco Mazzei, Professor Giorgio Amitrano and Professor Silvana De Maio from Università degli Studi di Napoli 'L'Orientale' for their help and support.

A big thanks goes to Professor Maddalena Toscano from Università degli Studi di Napoli 'L'Orientale', for her precious advice, and for encouraging me to apply for this Ph.D.

I also want to thank all my colleagues and friends at Matsumoto Lab for day-to-day discussion, for sharing their thoughts and for hanging out with me. Thank to Eric Nichols for his useful comments on my abstracts and papers, and for helping me to sort out various tasks; to Hiromi Ōyama for her help with the Japanese language

and for her friendship; to Jordi Polo and Francisco Dalla Rosa for their advices and support; to Manabu Kimura, Jessica C. Ramirez, Hiram Calvo, Erlyn Manguilimotan, Watanabe Yotaro, and many more.

I owe a big thanks to Robert Wener and to Prof. Natalia L. Tornesello editing this thesis and making it presentable to the international community.

Finally, my deepest thanks is to my supportive husband for always believing in me, for helping in many tasks, and for taking care of our little son.

List of Abbreviations

| | |
|----------|-------------------|
| OBJ | accusative |
| GEN | genitive |
| LOC | locative |
| NEG | negative |
| NOM | nominative |
| NON PAST | non-past tense |
| PAST | past tense |
| Q | question particle |
| TOP | topic |

General Remarks

Japanese names and sentences are Romanized following the Hepburn system, according to which vowels are pronounced as in Italian, consonants as in English, and long vowels are marked with a macron.

According to the Japanese use the family name always precedes the name, as in Kindaichi Haruhiko.

Examples taken from other authors have been transcript according to the Hepburn system if originally in a different form.

The interlinear annotation of the parts of speech has been adopted only if necessary.

List of Figures

| | |
|--|-----|
| Figure 1 – Temporal System of Natural Languages | 4 |
| Figure 2 – Classification of Aspectual Oppositions (Comrie 1976)..... | 23 |
| Figure 3 – Diagram of the Tenses of the Indicative in Italian | 93 |
| Figure 4 – The structure of Italian Actionality | 101 |
| Figure 5 – The Structure of Italian Aspectuality..... | 103 |
| Figure 6 – Automatic Estimation of Tense, Aspect and Event Class | 111 |
| Figure 7 – Corpus Based Approach..... | 112 |
| Figure 8 – Corpus Driven Approach..... | 112 |
| Figure 9 – POS Tagging of Italian and Japanese Sentences | 116 |
| Figure 10 – The Italian Tag Set | 117 |
| Figure 11 – Excerpt from the PEI Corpus | 120 |
| Figure 12 – Kudo's Verb Classification | 124 |
| Figure 13 – Verb Class Distribution – Japanese Data | 127 |
| Figure 14 – Verb Class Distribution – Italian Data | 127 |
| Figure 15 – Excerpt from the Tagged Corpus (Japanese Section)..... | 128 |
| Figure 16 – Excerpt from the Tagged Corpus (Italian Section)..... | 128 |
| Figure 17 – TAGPEI Jp – The Tool for Temporal Annotation | 131 |
| Figure 18 – Parsed Output..... | 132 |
| Figure 19 – The Annotated Parallel Corpus..... | 133 |
| Figure 20 – The Automatic Estimation Pipeline | 135 |
| Figure 21 – Tense, Aspect and Event Estimation (Italian Data)..... | 139 |
| Figure 22 – Tense, Aspect and Event Estimation (Japanese Data) | 140 |
| Figure 23 – Tense, Aspect and Event Estimation 2 (Italian Data)..... | 141 |

List of Tables

| | |
|---|-----|
| Table 1 – Hornstein’s Basic Tense Representation | 14 |
| Table 2 – Reichenbach’s Diagrams for Complex Tenses | 14 |
| Table 3 – Soga’s Diagrams for the Japanese Tense System | 15 |
| Table 4 – Bertinetto’s Diagrams for the Italian Tense System (Bertinetto 1986a) | 17 |
| Table 5 – Interaction between Lexical and Grammatical Aspect | 27 |
| Table 6 – Vendler Classification | 29 |
| Table 7 – Rothstein Classification | 31 |
| Table 8 – Van Lambalgen and Hamm Classification | 32 |
| Table 9 – Possible Interpretations of Tense Forms | 50 |
| Table 10 – Possible Time Relationships in Complex Sentences | 54 |
| Table 11 – Kindaichi Classification | 56 |
| Table 12 – Kindaichi and Fujii Classification (merged) | 61 |
| Table 13 – Nakamura’s Features | 65 |
| Table 14 – Nakamura’s Verb Classification and Related Aspect | 66 |
| Table 15 – Matsumoto and Oishi’s Feature-Based Framework | 73 |
| Table 16 – Class of Adverbs in the Japanese Language | 75 |
| Table 17 – Effects of the Temporal Adverbs on the Aspectual Values | 76 |
| Table 18 – Aspectual Morphemes and their Semantic Value | 91 |
| Table 19 – Aspect Values and Semantics in Italian | 107 |
| Table 20 – The PEI Corpus | 119 |
| Table 21 – Event Class – English, Italian and Japanese Verbs | 125 |
| Table 22 – Tense, Aspect and Event Class Attributes | 134 |
| Table 23 – ISST ConLL Extract | 137 |
| Table 24 – POS Tagging with the two Italian POS Tag Sets | 137 |
| Table 25 – <i>V-te iru</i> Distribution in the PEI Corpus | 146 |
| Table 26 – The Process to Determine the Meaning of the Form <i>V-te iru</i> | 160 |
| Table 27 – The Process to Determine the Meaning of the Aspectual Periphrasis ... | 173 |

Contents

| | |
|---|-----------|
| CONTENTS..... | IX |
| 1 INTRODUCTION..... | 1 |
| 1 Motivation..... | 1 |
| 2 Methodological Note | 2 |
| 3 Literature | 2 |
| 4 Thesis Organization | 3 |
| 2 THEORETICAL BACKGROUND..... | 4 |
| 2.1 Overview | 4 |
| 2.2 Language and Cognition in the Temporal Domain | 6 |
| 2.3 Conceptualising Time | 6 |
| 2.3.1 Time as Duration | 7 |
| 2.3.2 Temporal Perspective: Past, Present and Future | 8 |
| 2.3.3 Time as a Succession | 9 |
| 2.4 Time and Language | 10 |
| 2.4.1 Events and States | 10 |
| 2.4.2 Theories of Tense and Aspect | 11 |
| 2.4.2.1 Tense | 12 |
| 2.4.2.2 Aspect..... | 17 |
| 2.4.2.3 Cross Linguistic Differences | 18 |
| 2.5 Concluding Remarks | 19 |
| 3. A THEORY OF ASPECT | 20 |
| 3.1 Overview | 20 |
| 3.2 Lexical Aspect | 20 |
| 3.3 Grammatical Aspect | 23 |
| 3.4 Interaction between Lexical and Grammatical Aspect..... | 25 |
| 3.5 Event Classification..... | 28 |
| 3.5.1 Vendler (1957) | 28 |
| 3.5.2 Dowty (1979)..... | 30 |
| 3.5.3 Rothstein (2004)..... | 30 |

| | | |
|-----------|---|-----------|
| 3.5.4 | Van Lambalgen and Hamm (2005) | 31 |
| 3.6 | Aspect as a Choice (Smith 1991) | 33 |
| 3.7 | Compositional Aspect | 34 |
| 3.8 | Concluding Remarks..... | 36 |
| 4. | TENSE, ASPECT AND EVENTS: THE JAPANESE CASE..... | 38 |
| 4.1 | Overview | 38 |
| 4.2 | Tense | 40 |
| 4.2.1 | Relative Tense Theory | 41 |
| 4.2.1.1 | Tense Forms..... | 43 |
| 4.2.1.2 | Order of Events | 51 |
| 4.3 | Event Classification..... | 54 |
| 4.3.1 | Kindaichi Haruhiko | 55 |
| 4.3.1.1 | Problems Arising from Kindaichi's Classification | 59 |
| 4.3.1.2 | Kindaichi (1950) and Vendler (1957): a Comparison | 62 |
| 4.3.2 | Okuda (1978a; 1978b) and Jacobsen (1992) | 63 |
| 4.3.3 | Nakamura (2001)..... | 65 |
| 4.3.4 | Kudo M. (1995) | 69 |
| 4.3.5 | Matsumoto and Oishi (1997)..... | 73 |
| 4.3.6 | Temporal Adverbs | 74 |
| 4.4 | Aspect Morphemes | 77 |
| 4.4.1 | <i>V-te iru</i> | 77 |
| 4.4.2 | <i>V-te aru</i> | 80 |
| 4.4.3 | <i>V-te shimau</i> | 81 |
| 4.4.4 | <i>V-te iku</i> and <i>V-te kuru</i> | 82 |
| 4.4.5 | <i>VB2-owaru</i> and <i>VB2-oeru</i> | 87 |
| 4.4.6 | <i>VB2-tsuzukeru/tsuzuku</i> | 88 |
| 4.4.7 | <i>VB2-hajimeru/dasu</i> | 88 |
| 4.4.8 | <i>VB2-kakeru</i> | 89 |
| 4.5 | Concluding Remarks..... | 90 |
| 5. | TENSE, ASPECT AND EVENTS: THE ITALIAN CASE..... | 92 |
| 5.1 | Overview | 92 |

| | | |
|-----------|---|------------|
| 5.2 | Tense | 92 |
| 5.2.1 | Simple Present, Simple Future and Compound Future | 93 |
| 5.2.2 | Simple Past, Compound Past and Imperfect..... | 95 |
| 5.2.2.1 | The ‘Narrative Imperfect’ | 98 |
| 5.3 | Situation and Viewpoint Aspect | 100 |
| 5.3.1 | Situation Aspect | 100 |
| 5.3.2 | Viewpoint Aspect. Perfectivity vs. Imperfectivity..... | 102 |
| 5.3.3 | Lexical Morphemes and Aspectual Periphrasis..... | 104 |
| 5.4 | Temporal Adverbs in Italian | 108 |
| 5.5 | Concluding Remarks | 109 |
| 6. | AUTOMATIC ESTIMATION OF TENSE, ASPECT AND EVENTS | 111 |
| 6.1 | Overview | 111 |
| 6.2 | Corpus Driven Approach and Definitional Issues | 111 |
| 6.2.1 | Corpus Annotation as a Methodology | 115 |
| 6.3 | The PEI Corpus..... | 118 |
| 6.4 | An Annotation Scheme for Temporal Relations | 120 |
| 6.4.1 | The Temporal Labels..... | 121 |
| 6.5 | Corpus Annotation..... | 129 |
| 6.5.1 | Methodology..... | 129 |
| 6.5.2 | TAGPEI – the Annotation Tool | 129 |
| 6.6 | Tense/Aspect/Event Class Estimation – The Experimental Setup..... | 134 |
| 6.6.1 | Results and Discussion | 138 |
| 6.7 | Concluding Remarks | 142 |
| 7 | ITALIAN AND JAPANESE: A COMPARISON | 143 |
| 7.1 | Overview | 143 |
| 7.2 | <i>V-te iru</i> | 144 |
| 7.2.1 | Permanence of Results and Pure State | 146 |
| 7.2.2 | Pure State Meaning..... | 151 |
| 7.2.3 | Progressive Meaning | 152 |
| 7.2.4 | Habitual and Iterative-Durative Meaning | 157 |
| 7.2.5 | Experiential Meaning..... | 158 |

| | | |
|-----------|--|------------|
| 7.2.6 | Discussion | 159 |
| 7.3 | The Italian Periphrasis and their Japanese Counterparts | 161 |
| 7.3.1 | The Progressive Periphrasis..... | 161 |
| 7.3.2 | The Imminent Periphrasis | 163 |
| 7.3.3 | Phasal Periphrasis: inchoative, continuative, and conclusive | 165 |
| 7.3.3.1 | The inchoative periphrasis (VB2- <i>dasu/hajimeru</i> , V-te <i>ikulkuru</i>) | 165 |
| 7.3.3.2 | The continuative periphrasis..... | 167 |
| 7.3.3.3 | The Terminative or Conclusive Periphrasis | 170 |
| 7.3.3.4 | Discussion | 171 |
| 7.4 | Concluding Remarks..... | 174 |
| 8. | CONCLUSIONS AND FUTURE WORK | 176 |
| 8.1 | Conclusions | 176 |
| 8.2 | Future Work | 177 |
| | REFERENCES..... | 179 |
| | APPENDICES..... | 193 |
| | A – TAGPEI..... | 194 |
| | B – TAGGED EXAMPLES | 210 |
| | C – THE PEI CORPUS | 213 |

*[...] Me l'astipo
e dimane m' 'o bevo?'
Dimane nun esiste.
E 'o juorno primma,
siccome se n 'è gghiuto,
manco esiste.
Esiste solamente
stu mumento
'e chistu rito 'e vino int' 'a butteglia. [...]
E allora bevo...
E chistu surz' 'e vino
vence 'a partita cu l'eternità!*

Eduardo De Filippo

Chapter 1

1 Introduction

1 Motivation

This thesis stems from two basic observations: 1. ‘events’, considered as ‘happenings’ in the real world, contain the same temporal information regardless of the language in which they are conveyed, but the way in which such information is encoded via lexical or grammatical properties differs from language to language. While the lexical aspect of event semantics is relatively similar across languages, the grammatical aspect varies greatly. Concepts of tense and aspect may not be shared across languages, thus preventing a simple correspondence of verbal forms between languages from being established; 2. Temporal analysis of ‘events’ is an important part of many natural language processing tasks such as text summarization (temporally ordering information), information extraction (*e.g.*, normalizing information for database entry) question-answering (*e.g.*, answering ‘when’ questions by temporally anchoring events) and machine-translation. Studies and experiments in the field of temporal annotation may contribute to the development of increasingly sophisticated applications.

The aim of the present work is two-fold:

1. To conduct a theoretical and empirical contrastive investigation on how temporal information is conveyed in Japanese and Italian with a focus on the semantics of events and how they are reported, which includes the notion of eventuality types and their expression in tense and aspect, empirically supported by semi-automatically annotating a parallel Japanese-Italian corpus with temporality tags;
2. To define an automatic procedure for temporal tagging to be used in natural language processing applications and by researchers working on tasks requiring detailed, cross-lingual temporal analysis.

The intended users of the annotated data are the overall linguistic community, the community involved in processing Natural Languages, system developers involved in building tagging programs to extract temporal information and those who are de-

veloping Machine Translation Systems and Second Language teachers interested in data which may be easily compared in the two languages under discussion.

2 Methodological Note

We have conducted the analysis by designing, creating and semi-automatically annotating a parallel Japanese-Italian corpus of novels, parliamentary proceedings and newspapers translations with temporality tags, and conducting an investigation into the representation of temporal phenomena across the two languages.

We have adopted the corpus-driven approach for the benefits normally associated with it: a better understanding of the phenomena of concern and resources for the training and evaluation of algorithms to automatically identify features and relations of interest. Both the Japanese and the Italian sections of the parallel corpus have been first pre-processed through stable morphological analysers and then tagged with a tool we have specifically built for the purposes of this research.

Given the lack of parallel resources in Japanese and Italian, the parallel (translation corpus) in itself may be considered one of the results of this work. The data collected may represent a relevant base of data for comparable linguistic analysis not only on event semantics, but also on other linguistic domains.

3 Literature

This thesis lies at the intersection of linguistics, logic, cognitive and computational science. To meet the increasing demand on the processing of temporal information a deep interaction between artificial intelligence, logic and formal linguistic theories is needed, namely an interaction between three domains: i) theories of tense and aspect and event structure; ii) reasoning about time and events; iii) annotation schemes and tools.

We have reviewed previous works both in the field of theoretical and computational linguistics. A full mastery of the relevant literature of all these areas is however far beyond the scope of the present work, focussing on the devices employed by Italian and Japanese speakers to express temporality.

To this extent, we have taken into account previous studies on the semantics of events and their expression in tense and aspect, as well as those on temporal annotation. We have tried to make use of the extensive literature on these two subjects in order to pinpoint the linguistic devices used in the two languages to convey events.

4 Thesis Organization

This thesis is organized as follows.

Chapter 2 is devoted to the theoretical background of which two elements are presented: 1) the main theories concerned with time; 2) the way natural languages convey it.

Chapter 3 is concerned with the theories on Tense and Aspect. The main theoretical positions about Tense and the difference between Viewpoint/Grammatical and Situation/Lexical Aspect are outlined. Traditional and innovative generalist approaches are reviewed.

Chapter 4 and 5 are respectively devoted to the way in which Japanese and Italian convey the temporal perspective with a focus on tense and aspect and on their interaction.

Chapter 6 and 7 represent the core of the research. Chapter 6 presents the parallel corpus, the temporal annotation methodology and the experimental setting for the automatic estimation of tense, aspect and event class. Chapter 7 is devoted to the contrastive analysis between Japanese and Italian as far as aspect and event classes are concerned, carried on by querying the tagged parallel corpus illustrated in chapter 6.

Conclusions are drawn in chapter 8.