Salience-based Modeling of Discourse Context

SHIRAMATSU, Shun
Abstract

The aim in this thesis is to establish a computational model of discourse context on the basis of discourse salience. Dealing with discourse context is an important requirement for various discourse processing systems (e.g., discussion support ones and conversation ones) because the meaning of language expressions essentially depends on the discourse context. The discourse context gradually changes as the discourse progress because the targets of the participants’ attention change with each utterance unit. Hence, we have to take into account the time-series change of discourse salience (i.e., the perceptual degree of the participants’ attention to the discourse entities).

This thesis consists of the technical layer and the application layer. The technical layer deals with the salience-based model of discourse context. The application layer deals with information provision systems (e.g., those that support public involvement) by using the salience-based model of discourse context.

The goal for the technical layer is to establish a computational model of discourse context that takes into account the time-series change of discourse salience. It can be broken down to the three technical issues: (1) formulate discourse salience, (2) formulate referential coherence, and (3) formulate con-
Abstract

text similarity. To address these issues, we used the following approaches: (1) probabilistic formulation of salience, (2) game-theoretic formulation of referential coherence, and (3) vectorial formulation of context similarity. We formulate (2) and (3) on the basis of (1). We empirically evaluated the effectiveness of our formulations by using corpora. Furthermore, we discuss the academic contributions of our modeling. We empirically find out that our model is more effective for spontaneous spoken language than for written language. We also empirically prove the hypothesis that the referential coherence can be derived from game theory.

The goal for the application layer is to develop systems that provide information through effectively using the context of users’ discourse. For instance, in order to support consensus building in various communities, an application that provides contextual information is an effective one. In particular, we develop two prototype systems: (1) a system that provides a time-series overview of a long discourse (e.g., conference minutes and judicial records) and (2) a system that provides information related to the users’ discourse context. We discuss their effectiveness in supporting consensus building in Public Involvement processes.

In Chapter 1, we introduce our motivation, our goals, and the key issues for this study. We need to establish a computational model of discourse context, which gradually changes along with the discourse progress, in order to realize advanced processing of discourse.

Chapter 2 surveys the literature related to discourse context, discourse salience, information provision, and game-theoretic pragmatics. We describe
the contemporary condition that various salience factors proposed in different research fields have not been integrated.

Chapter 3 describes development of GDASDK (GDA Software Development Kit), a Java package for processing Global Document Annotation (GDA). It consists of APIs for processing GDA, automatic annotation system with GDA, and simple viewers of GDA.

In Chapter 4, we provide our probabilistic formulation of the discourse salience. We develop a probabilistic method to integrate salience factors and to optimize the salience calculation.

Our game-theoretic formulation of the referential coherence is provided in Chapter 5. To verify the language universality of a hypothesis that the referential coherence can be explained by game theory, we statistically formulate pronominalization and the expected utility.

In Chapter 6, we empirically evaluate our formulations by using corpora. To deal with the dynamic transition of discourse context, we empirically determine the optimal window functions. We find out that our approach to formulate salience is more effective for spontaneous conversation than for newspaper articles. Furthermore, we find out empirical evidences of the hypothesis that the referential coherence can be explained by game theory by using large Japanese and English corpora.

In Chapter 7, we describe the applications of our model to information provision and public involvement. We develop prototype systems to support sharing discourse context: one is the system visualizing dynamic transition of salience, and the other is the system providing information related to current
Abstract

discourse context.

In Chapter 8, we discuss our major contributions and the potential applications of our work. Especially, we discuss our future perspectives to support the public involvement process.

Finally, we present the conclusion of our study in Chapter 9.
Acknowledgments

This work was carried out at Okuno Laboratory, Graduate School of Informatics, Kyoto University. I would like to thank Professor Hiroshi G. Okuno for his thoughtful supervision throughout my studies at Tokyo University of Science (TUS) and Kyoto University. His insightful and visionary advices helped me to decide the direction of my study.

Dr. Kōiti Hasida (AIST) also greatly supported my research. He provided me with fundamental concepts and methodologies for my study when I was working under his supervision at AIST.

This work has been supported by other many people. Professors Tatsuya Kawahara and Sadao Kurohashi provided me with insightful comments as co-advisors. Dr. Kazunori Komatani provided me with judicious and helpful advices for my study. Dr. Tetsuya Ogata provided me with trenchant discussions about my research. Dr. Takashi Miyata provided me with substantial suggestions about basic language processing at AIST. I was also mentally supported by Mr. Keisuke Ishida, Dr. Yutaka Matsuo (Tokyo University), Dr. Junichiro Mori (Tokyo University), and Mr. Masahiro Ikemoto in AIST. Ms. Hayeong Jeong and Professor Kiyoshi Kobayashi, collaborators for the study on public debate, provided me with a great suggestion in respect of
application of our technologies. Ms. Yoko Nakahara provided me with manual annotation for a corpus and treated me to several dinners during writing this thesis. Mr. Tetsuya Watanabe provided me with helpful suggestion to apply my work to call center support systems. Mr. Takashi Sakao (JST), Ms. Tomoka Matsuo and Ms. Mayu Yagi helped me to apply a patent of the technology for contextual information provision. Ms. Miki Nishii provided me with clerical support and relaxed atmosphere at our laboratory. Mr. Yuichiro Fukubayashi, Mr. Satoshi Ikeda, Mr. Masaki Katsumaru and Mr. Naoyuki Kanda have given me helpful opinions at dialogue-team meetings every week. I also have been given encouraging support by the other members of the Okuno Laboratory: Dr. Kazuyoshi Yoshii, Mr. Ryu Takeda, Mr. Hisashi Kanda, Mr. Shun Nishide, Mr. Ryunosuke Yokoya, Mr. Katsutoshi Itoyama, and so on. Mako, a small restaurant, has supplied the members and me with dinners and dining table every night. I also encouraged by graduated members of Okuno Laboratory: Dr. Tesuro Kitahara (Kwansei Gakuin University), Mr. Masamitsu Murase, Mr. Masahiro Nishiyama, Mr. Hiromasa Fujihara (AIST), Mr. Haruhiko Niwa (Wabot-House Lab.), and so on. Ms. Yuki Yamanaka also encouraged me by giving me a good-luck amulet for scholars. Dr. Katsuko Murakami has mentally helped me through biweekly sensitive counseling. I have been especially given emotional support by Ms. Mari Takahara. The Japan Society for the Promotion and Science (JSPS) financially supported my life as a researcher. Finally, I wish to thank my family for their understanding during my lengthy studies.
Contents

Abstract i

Acknowledgments v

Contents xi

List of Figures xv

List of Tables xviii

List of Symbols xix

1 Introduction 1
  1.1 Motivation .................................. 1
  1.2 Goals and Issues ............................ 3
  1.3 Our Approach: Salience-based Modeling ...... 7
  1.4 Thesis Organization .......................... 8

2 Literature Review 11
  2.1 Reviews on Salience and Discourse Context .... 11
      2.1.1 Centering Theory ........................ 11
## CONTENTS

2.1.2 Anaphora Resolution ........................................... 14  
2.1.3 Recency and Primacy Effects ................................. 15  
2.1.4 Term-Weighting Schemes ................................. 15  
2.1.5 Direct and Indirect Priming Effects ...................... 16  
2.1.6 Spreading Activation Theory .............................. 16  
2.1.7 Aspect Model ................................................ 17  
2.1.8 Text Segmentation ......................................... 18  
2.1.9 Rhetorical Structure Theory ............................. 18  
2.1.10 Dynamic Semantics ...................................... 19  
2.2 Reviews on Information Provision and Discourse Context ... 20  
2.2.1 Information Visualization Systems ........................ 20  
2.2.2 Query-Free Information Retrieval ........................ 21  
2.2.3 Topic Detection and Tracking .......................... 21  
2.2.4 Discussion Analysis ..................................... 21  
2.2.5 Call Center Support Systems ............................. 22  
2.3 Reviews on Game-theoretic Pragmatics ....................... 22  
2.3.1 Signaling Game ............................................ 24  
2.3.2 Game-theoretic Disambiguation ......................... 24  
2.3.3 Meaning Game .......................................... 25  

3 GDA Software Development Kit ................................. 27  
3.1 Specification of GDA Tag set ................................. 28  
3.1.1 Dependency Structure .................................... 28  
3.1.2 Anaphoric Structure ..................................... 29  
3.2 Developing API for GDA .................................... 30
3.3 Developing GDA Viewers ............................................. 32
  3.3.1 GDA Tagging Systems ........................................... 33
3.4 Extending CSJ Corpus .................................................. 33

4 Statistical Formulation of Discourse Salience 35
  4.1 Introduction ............................................................ 35
    4.1.1 Requirement and Issue ........................................... 35
    4.1.2 Our Approach ...................................................... 36
  4.2 Formulating Salience: Reference Probability ....................... 38
  4.3 Calculation Method .................................................... 40
    4.3.1 Extracting Samples ............................................... 42
    4.3.2 Extracting Features .............................................. 43
    4.3.3 Windowing Features .............................................. 45
    4.3.4 Feature Integration: Leaning and Calculation .................. 46
  4.4 Evaluation Criteria ................................................... 47
  4.5 Conclusion ............................................................. 48

5 Meaning-Game-based Centering Model 49
  5.1 Introduction ........................................................... 49
  5.2 Issues on Centering Theory ........................................... 52
  5.3 Meaning Game .......................................................... 55
    5.3.1 General Framework .............................................. 55
    5.3.2 Derivation of Centering Theory .................................. 55
  5.4 Formulating Referential Coherence ................................... 56
    5.4.1 Generalized Formulation ...................................... 56
CONTENTS

5.4.2 Statistical Definition of Parameters .......................... 60
5.5 Conclusion .................................................................. 61

6 Empirical Evaluation .................................................. 63
6.1 Corpus Specification .................................................. 63
6.2 Evaluating Reference Probability as Discourse Salience ...... 66
  6.2.1 Clarifying Characteristics of Recency Effect in Discourse Context ........................................... 67
  6.2.2 Optimizing and Evaluating Calculation ................. 71
6.3 Verifying Meaning-Game-based Centering Model ............ 76
  6.3.1 Theoretic Verification of Parameter Definitions ...... 77
  6.3.2 Verification of Preference 1a ............................... 81
  6.3.3 Verification of Preference 1b ............................... 84
  6.3.4 Verification of Preference 2 ............................... 86
6.4 Discussion ............................................................... 88
  6.4.1 Characteristics of Expected Utility as a Scale of Referential Coherence ...................................... 88
  6.4.2 Quantitative Comparison of Japanese and English Corpora .................................................. 89
6.5 Conclusion ............................................................... 91

7 Application to Information Provision ............................ 93
7.1 SalienceGraph: Providing Time-Series Overview of Discourse . 93
7.2 Providing Information Related to Users’ Discourse Context .. 96
  7.2.1 Representing Discourse Context as Salience Vector .................................................. 96
# List of Figures

1.1 Discourse salience in the joint attentional state influenced by the preceding discourse context ........................................ 2
1.2 Time-series transition of the discourse salience ..................... 3
1.3 Goals and issues .................................................................. 4
1.4 Visualizing overview of an entire discourse ............................. 5
1.5 Organization of thesis .......................................................... 9
2.1 Referentially coherent and incoherent examples ....................... 13
2.2 Serial position curve ............................................................ 16
2.3 Graphical image of LDA ......................................................... 17
2.4 Example of rhetorical structure .............................................. 19
3.1 GDAViewer: Standard visualization of syntactic tree of GDA .... 32
3.2 SemGraphViewer: Visualizing dependency graph of GDA ....... 32
4.1 Serial position curve ............................................................ 37
4.2 Approach: Window function (shifting with each utterance unit) 37
4.3 Example of reference probability .......................................... 39
4.4 An example method for calculating \( \Pr(e|\text{pre}(U_i)) \) ................ 41
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>Extracting sample $\langle e, U_i \rangle$ to calculate $\Pr(e</td>
<td>\text{pre}(U_i))$</td>
</tr>
<tr>
<td>4.6</td>
<td>Complete co-reference chain of “京都” (Kyoto) annotated manually</td>
<td>44</td>
</tr>
<tr>
<td>4.7</td>
<td>Superficial occurrence chain of “京都” (Kyoto)</td>
<td>44</td>
</tr>
<tr>
<td>4.8</td>
<td>Using only automatically extractable features</td>
<td>45</td>
</tr>
<tr>
<td>4.9</td>
<td>Feature windowing according to $\text{dist}(w, U_{i+1})$</td>
<td>47</td>
</tr>
<tr>
<td>5.1</td>
<td>Two issues</td>
<td>50</td>
</tr>
<tr>
<td>5.2</td>
<td>Referentially coherent and incoherent examples</td>
<td>53</td>
</tr>
<tr>
<td>5.3</td>
<td>Derivation of Rules 1 and 2 from the meaning game</td>
<td>56</td>
</tr>
<tr>
<td>5.4</td>
<td>Preferences 1a and 1b: Generalization of Rule 1 of centering theory</td>
<td>57</td>
</tr>
<tr>
<td>6.1</td>
<td>Relationship of reference probability and recency ($CSJ$)</td>
<td>66</td>
</tr>
<tr>
<td>6.2</td>
<td>An indication of the extent of feature influence</td>
<td>70</td>
</tr>
<tr>
<td>6.3</td>
<td>The optimized calculation method</td>
<td>75</td>
</tr>
<tr>
<td>6.4</td>
<td>Comparing $\Pr(e</td>
<td>\text{pre}(U_i))$ with naive term-weighing schemes</td>
</tr>
<tr>
<td></td>
<td>($CSJ$)</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>Comparing $\Pr(e</td>
<td>\text{pre}(U_i))$ with naive term-weighing schemes</td>
</tr>
<tr>
<td></td>
<td>($Mainichi$)</td>
<td></td>
</tr>
<tr>
<td>6.6</td>
<td>Reference probability calculated from $\text{dist, chain (Mainichi)}$</td>
<td>79</td>
</tr>
<tr>
<td>6.7</td>
<td>Reference probability calculated from $\text{dist, chain (WSJ)}$</td>
<td>79</td>
</tr>
<tr>
<td>6.8</td>
<td>Determination of $U_{t_0}$ according to the consistency of Preference 2 with Rule 2</td>
<td>80</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>6.9</td>
<td>Preference 1: Ratio of positive pairs is proportional to $\text{EU}(\text{refmap}<em>A(U</em>{i+1})) - \text{EU}(\text{refmap}<em>B(U</em>{i+1}))$ (Mainichi).</td>
<td>83</td>
</tr>
<tr>
<td>6.10</td>
<td>Preference 1: Ratio of positive pairs is proportional to $\text{EU}(\text{refmap}<em>A(U</em>{i+1})) - \text{EU}(\text{refmap}<em>B(U</em>{i+1}))$ (WSJ).</td>
<td>84</td>
</tr>
<tr>
<td>6.11</td>
<td>Preference 2: Avg. $\text{EU}(\text{refmap}(U_{i+1}))$ of each transition type (Mainichi).</td>
<td>88</td>
</tr>
<tr>
<td>6.12</td>
<td>Preference Rule 2: Avg. $\text{EU}(\text{refmap}(U_{i+1}))$ of each transition type (WSJ).</td>
<td>88</td>
</tr>
<tr>
<td>6.13</td>
<td>Reference probability and ratio of pronominalized entities.</td>
<td>90</td>
</tr>
<tr>
<td>7.1</td>
<td>SalienceGraph: Visualizing time-series overview of discourse</td>
<td>94</td>
</tr>
<tr>
<td>7.2</td>
<td>Zooming into demanded part</td>
<td>94</td>
</tr>
<tr>
<td>7.3</td>
<td>Comparing SalienceGraph and TF-based visualization</td>
<td>95</td>
</tr>
<tr>
<td>7.4</td>
<td>Example of salience vectors changing with the discourse progress</td>
<td>97</td>
</tr>
<tr>
<td>7.5</td>
<td>Contextual information provision based on salience vector ($\text{Pr}(e</td>
<td>\text{pre}(U_i))$): Providing text related to the current context of users’ discourse</td>
</tr>
<tr>
<td>7.6</td>
<td>Example of a generated salience vector as a query representing users’ discourse context</td>
<td>99</td>
</tr>
<tr>
<td>7.7</td>
<td>Contextual information provision based on Bag-of-Words (TF-IDF)</td>
<td>100</td>
</tr>
<tr>
<td>8.1</td>
<td>SalienceGraph: visualization of time-series flow of discourse</td>
<td>107</td>
</tr>
<tr>
<td>8.2</td>
<td>Future work: required interface of discourse browsing system</td>
<td>113</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Table Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Extracting the influencing factors of salience from related studies</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>RAP: Initial weights assigned to the salience factors</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Difference between salience estimation and language modeling</td>
<td>17</td>
</tr>
<tr>
<td>2.4</td>
<td>Prisoner’s Dilemma</td>
<td>22</td>
</tr>
<tr>
<td>2.5</td>
<td>Typical applications of the signaling game</td>
<td>24</td>
</tr>
<tr>
<td>4.1</td>
<td>Elements and candidate features for calculating reference probability</td>
<td>46</td>
</tr>
<tr>
<td>5.1</td>
<td>Transition types of Rule 2</td>
<td>53</td>
</tr>
<tr>
<td>5.2</td>
<td>Correspondence of centering theory’s account and the meaning game’s account</td>
<td>58</td>
</tr>
<tr>
<td>6.1</td>
<td>Comparing window functions optimized for each feature (CSJ)</td>
<td>68</td>
</tr>
<tr>
<td>6.2</td>
<td>Comparing window functions optimized for each feature (Mainichi)</td>
<td>69</td>
</tr>
<tr>
<td>6.3</td>
<td>Curve gentleness of optimal inverse window (CSJ)</td>
<td>71</td>
</tr>
<tr>
<td>6.4</td>
<td>Curve gentleness of optimal inverse window (Mainichi)</td>
<td>71</td>
</tr>
<tr>
<td>6.5</td>
<td>Description of the candidate features</td>
<td>72</td>
</tr>
<tr>
<td>6.6</td>
<td>Comparing features (CSJ)</td>
<td>72</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

6.7 Comparing features (*Mainichi*) ........................................ 73
6.8 Evaluation of candidate feature sets (*CSJ*) .............................. 73
6.9 Evaluation of candidate feature sets (*Mainichi*) ........................ 74
6.10 Average reference probability for each grammatical function
    (*Mainichi*) ........................................................................ 77
6.11 Average reference probability for each grammatical function
    (*WSJ*) .............................................................................. 78
6.12 Perceptual utility for each referring expression (*Mainichi*) ... 81
6.13 Perceptual utility for each referring expression (*WSJ*) ........... 82
6.14 Preference 1b: Spearman’s rank correlation between \( \text{Pr}(e|\text{pre}(U_i)) \)
    and \( \text{Ut}(w) \) (*Mainichi*) .............................................. 85
6.15 Preference 1b: Spearman’s rank correlation between \( \text{Pr}(e|\text{pre}(U_i)) \)
    and \( \text{Ut}(w) \) (*WSJ*) ...................................................... 85
6.16 Preference 2: Avg. \( \text{EU}(\text{refmap}(U_{i+1})) \) of each transition type
    (*Mainichi*) ........................................................................ 86
6.17 Preference 2: Avg. \( \text{EU}(\text{refmap}(U_{i+1})) \) of each transition type
    (*WSJ*) .............................................................................. 86
## List of Symbols

### Symbols used in this thesis

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>([U_1, \ldots, U_n])</td>
<td>Discourse (i.e., sequence of utterance units)</td>
</tr>
<tr>
<td>(U_i)</td>
<td>The (i)-th utterance unit in discourse</td>
</tr>
<tr>
<td>(e)</td>
<td>Discourse entity</td>
</tr>
<tr>
<td>(w)</td>
<td>Word</td>
</tr>
</tbody>
</table>

### Symbols used in Chapter 2 or later

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S)</td>
<td>Sender of an utterance, i.e., writer or speaker</td>
</tr>
<tr>
<td>(R)</td>
<td>Receiver of an utterance, i.e., reader or listener</td>
</tr>
</tbody>
</table>

### Symbols used in Chapter 3 or later

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{pre}(U_i))</td>
<td>([U_1, \ldots, U_i]), the preceding discourse of (U_i)</td>
</tr>
<tr>
<td>(w \xrightarrow{\text{ref}} e)</td>
<td>(w) refers to (e)</td>
</tr>
<tr>
<td>(\langle e, U_i \rangle)</td>
<td>A sample representing a tuple consisting of a target (e) and an target moment in which (U_i) is conveyed</td>
</tr>
<tr>
<td>(\text{isRef}(e, U_{i+1}))</td>
<td>1 if (\exists w \xrightarrow{\text{ref}} e) in (U_{i+1}), otherwise 0</td>
</tr>
<tr>
<td>(\text{dist}(w, U_{i+1}))</td>
<td>Utterance distance from (w) to (U_{i+1}) (i.e., ((i + 1) - j) when (w) is in the utterance (U_j))</td>
</tr>
<tr>
<td>(W(\text{dist}))</td>
<td>Window function according to (\text{dist}(w, U_{i+1}))</td>
</tr>
<tr>
<td>(\Pr(e</td>
<td>\text{pre}(U_i)))</td>
</tr>
</tbody>
</table>
List of Symbols

\text{evalSal}(m) \quad \text{Evaluation scale of a method } m \text{ calculating salience,}
\text{defined as correlation coefficient between}
a salience value of } e \text{ at } U_i \text{ calculated by } m \text{ and}
isRef(e, U_{i+1}) \text{ in a test-set corpus}

\textbf{Symbols used in Chapter 4 or later}
\langle w, e \rangle \quad \text{Candidate reference mapping of } w \xrightarrow{\text{ref}} e

\text{Ut}_0 \quad \text{Utility when } S \text{ and } R \text{ cooperatively select}
\text{the same candidate mapping } \langle w, e \rangle

\text{I}(w) \quad \text{Perceptual cost of } w,
\text{defined as } -\log(p(w)),
to formulate reversed degree of pronominalization

\text{Ut}(w) \quad \text{Perceptual utility of } w,
\text{defined as } \text{Ut}_0 - \text{I}(w) = \text{Ut}_0 + \log(p(w)),
to formulate utility to use } w \text{ with considering pronominalization}

\text{refmap}(U_{i+1}) \quad \text{Candidate set of reference mappings } \langle e, w \rangle,
\text{defined as } \{ \langle e, w \rangle; w \xrightarrow{\text{ref}} e \text{ in } U_{i+1} \},

\text{EU(refmap}(U_{i+1})) \quad \text{Expected utility of refmap}(U_{i+1}),
\text{defined as } \sum_{w \xrightarrow{\text{ref}} e \text{ in } U_{i+1}} \Pr(e|\text{pre}(U_i))\text{Ut}(w),
to formulate referential coherence between
the preceding pre(U_i) and the succeeding U_{i+1}

\textbf{Symbols used in Chapter 5 or later}
\text{v}(U) \quad \text{Salience vector, which represents discourse context}
at the moment when the utterance } U \text{ is conveyed,
\text{defined as } [\Pr(e_1|\text{pre}(U)), \ldots, \Pr(e_N|\text{pre}(U))]^T

\text{sim}\left(\text{v}(U), \text{v}(S)\right) \quad \text{Context similarity between } \text{v}(U) \text{ and } \text{v}(S),
\text{defined as the cosine similarity } \frac{\text{v}(U) \cdot \text{v}(S)}{||\text{v}(U)|| ||\text{v}(S)||}
Chapter 1

Introduction

1.1 Motivation

The meaning of language expressions essentially depends on the discourse context. Without taking into account the preceding context of a sentence, it is impossible to accurately understand the meaning of the sentence. Hence we aim to establish a computational model of discourse context. This is an important issue for various discourse processing systems, e.g., discussion support ones, conversation ones, and text summarization ones. In particular, supporting discussion is an important application for consensus building in various communities. For instance, in order to give different perspectives, it is useful to provide information by using conference minutes. We aim to develop elemental technologies that can be applied to the discussion support system through modeling the discourse context.

We suppose that discourse salience, i.e., the perceptual degree of attention to each entity referred to in the discourse, is an important factor for modeling the discourse context. The discourse context gradually changes as the discourse progress because the targets of the discourse participants’ attention change with each utterance unit.\footnote{Discourse participants include speakers, listeners, writers, and readers.} To represent the time-series
transition of discourse context, dealing with the time-series change of the discourse salience is indispensable. It changes with the transition of joint attention between the discourse participants. By sharing the joint attention, the participants ensure that their communication is successful because communication requires cooperative use of language. In other words, although the salience depends on each individual’s perception, we assume that discourse participants who are communicating collaboratively can focus their attention on the same targets. Although joint attention can also be triggered by the non-verbal expressions (e.g., the gaze direction), we focus on the joint attention caused by the verbal expressions.

As shown in Figure 1.1, the state of the joint attention is influenced by the shared preceding discourse context. The salience of each entity is represented as the size of the word in the figure. It gradually changes according to the time-series progress of the discourse (Figure 1.2). The discourse participants
cooperatively produce utterances, which are predictable from their joint attention. Thus, the salience in the joint attention plays a significant role in the cooperative communication. Hence we must develop a computational model of the discourse salience.

1.2 Goals and Issues

Our goals for the technical layer and the application layer are shown in Figure 1.3. Our goal for the application layer is to provide information by effectively using the discourse context.

We suppose that users’ understanding of discourse context can be supported by systems that provide contextually consistent information. Here, we term such systems “context-grounding support systems”. We conjecture that such contextual information provision is effective in order to support consensus building in various communities because context is closely interlinked with common grounds[1].
For the application layer, we aim to develop the following applications:

[Applicative Goal 1] Provide time-series overview of discourse: To help users to grasp the time-series overview of a long discourse (e.g., conference minutes and judicial records), we aim to visualize a contextual flow from the start to the end of the discourse. A graph visualizing an overall flow of discourse salience is shown in Figure 1.4. It enables us to develop the discourse browsing interface, which satisfies the Visual Information-Seeking Mantra[2]: “Overview first, zoom and filter, then details on demand.” We conjecture that, because it effectively builds shared understanding about the overview of discussions, such visualization can be applied to support the facilitation of a discussion.

Figure 1.3: Goals and issues
[Applicative Goal 2] Provide information related to discourse context: To help users to know the related information of a discourse, we aim to provide information related to users’ discourse context. For instance, we conjecture that providing such information helps to avoid disproportion of discussion because considering diverse viewpoints about the agenda when the participants determine their opinions is helpful.

For instance, these contextual provisions can be applied to supporting Public Involvement (PI) process. PI, a citizen participation process in the decision making of public policy, is characterized as an interactive communication process among stakeholders[3, 4, 5]. One of the significant roles of PI is to understand diverse perceptions possessed by the civil society members and make judgment on the appropriateness of the projects[5]. PI processes such as public deliberation, town meeting, workshop, and so on have been examined. The stakeholders, however, sometimes argue on different planes because they have diverse background contexts. Hence, to encourage appropriate decision making in the PI processes, it is required to develop discussion support systems with the discourse processing technologies. We conjecture

![Figure 1.4: Visualizing overview of an entire discourse](image-url)
CHAPTER 1. INTRODUCTION

that the contextual information provision systems that we aim help the debate participants to share their diverse background contexts.

To develop the above applications, we have to establish a computational model of the discourse context. This is the technical goal described in this thesis. This goal can be broken down to the following technical issues:

[Technical Issue 1] **Formulate discourse salience**: Quantitatively formulating the salience is an important issue because we suppose that the contextual influence of an entity (in the preceding discourse) can be measured as the salience of the current one. For example, we must quantify the similarity of the context flow in order to retrieve a sentence influenced by the particular preceding context. Quantifying the discourse salience is indispensable in order to quantify the similarity of discourse context that changes along with the discourse progress. Hence we aim to quantitatively formulate the discourse salience.

[Technical Issue 2] **Formulate referential coherence**: The referential coherence is important to determine “how to say (provide)” for information provision systems. The systems are required to provide an easy-to-understand discourse. The systems have to automatically determine the optimal order of sentences on the basis of the referential coherence because the referential coherence is a critical factor for establishing the understandability of discourse.

[Technical Issue 3] **Formulate context similarity**: The context similarity is important to determine “what to say (provide)” for information provision systems. The systems are required to provide contextually consistent information. The systems have to automatically determine content whose context is similar to that of the users. For instance, the systems can select the provided content from candidate texts whose discourse context is similar to the users’ conversational context.

To solve these technical issues, we need to solve the following scientific issues.
1.3. OUR APPROACH: SALIENCE-BASED MODELING

[Scientific Issue 1] Clarify influencing factors of discourse salience:
To formulate the discourse salience, we have to clarify what factors influence it. In particular, we suppose that the *recency effect* is an important factor to deal with dynamic change of the discourse salience. The recency effect means that the more recent the occurrence is, the more likely it is to be recalled. The influence of the recency effect to the discourse salience, however, has not been clarified yet.

[Scientific Issue 2] Clarify behavioral principle of referential coherence:
To formulate the referential coherence, we need to clarify what behavioral principle is behind it. This clarification is required for cooperative interaction with language between humans and machines. The principle, however, has not been clarified yet.

1.3 Our Approach: Salience-based Modeling

To establish a computational model of the discourse context, we employed the following solutions:

[Solution 1] Probabilistic formulation of discourse salience: We can measure the discourse salience with a probabilistic approach because the more salient the discourse entity is, the more probably it becomes referred to in the succeeding utterance. In other words, the participants focus their attention on the entity that is likely to be successively referred to. This characteristic of salience is backed up by *centering theory*, which is a rule-based theory of referential coherence. Furthermore, we propose an evaluation scale for optimizing the calculation method of salience. We empirically evaluate the calculation method optimized on the basis of the calculation scale.

[Solution 2] Game-theoretic formulation of referential coherence: We can formulate the behavioral principle behind the referential coherence with a game-theoretic approach because this coherence is a
strategic preference caused by cooperative behavior between the discourse participants. In other words, centering theory, a theory about the referential coherence, can be derived from game theory. We empirically verified this hypothesis by using large Japanese and English corpora.

**Solution 3** Vectorial formulation of context similarity: Because the dynamic influences of the preceding discourse are coded into the discourse salience, we can represent the discourse context as a vector consisting of the discourse salience. Hence, we formulate the context similarity on the basis of the vectorial representation of the discourse context.

Solution 1 is the basis of our model. In other words, Solutions 2 and 3 is based on Solution 1.

### 1.4 Thesis Organization

The organization of this thesis is shown in Figure 1.5. Chapter 2 surveys the literatures related to discourse context, discourse salience, information provision, and game-theoretic pragmatics. As a preparation, Chapter 3 describes development of a Java package for processing Global Document Annotation (GDA). Chapter 4 provides our probabilistic formulation of the discourse salience. Chapter 5 provides our game-theoretic formulation of the referential coherence. Then, we empirically evaluate our formulations by using corpora in Chapter 6. In Chapter 7, we describe the applications of our model to information provision and Public Involvement. Chapter 8 discusses our major contributions and the potential applications of our work. Finally, Chapter 9 concludes our thesis.
1. Introduction

2. Literature Review

3. GDA Software Development Kit
   - Formulate salience (changing along with the progress of discourse context)

4. Statistical Formulation of Discourse Salience
   - Formulate referential coherence

5. Meaning-Game-based Centering Model

6. Empirical Evaluation
   - Formulate context similarity

7. Application to Information Provision

8. Perspectives and Future Works

9. Conclusion

Figure 1.5: Organization of thesis
Chapter 2

Literature Review

This chapter provides reviews of the literatures related to discourse context, discourse salience, information provision, and game-theoretic pragmatics.

2.1 Reviews on Salience and Discourse Context

There are several theories or models related to discourse salience in the fields of linguistics and cognitive science. Table 2.1 shows the association between the related studies and influencing factors of salience. The influencing factors surrounded by the oval box are dealt with by this thesis.

2.1.1 Centering Theory

Centering theory is a standard theory gives a model of the local transition of the discourse participants’ attentional state[6]. It formalizes discourse salience, pronominalization, and referential coherence through rule-based approach. Referential coherence means smoothness in the transition of attention between the target utterance and the succeeding one. Centering theory consists of the two rules about referential coherence. Figure 2.1 shows examples of referential coherence and incoherence. In the coherent example, the
CHAPTER 2. LITERATURE REVIEW

Table 2.1: Extracting the influencing factors of salience from related studies

<table>
<thead>
<tr>
<th>Related study</th>
<th>Influencing factor of salience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centering theory</td>
<td>Grammatical function</td>
</tr>
<tr>
<td>Anaphora resolution</td>
<td>Anaphora</td>
</tr>
<tr>
<td>Term-weighing scheme</td>
<td>Type of named entity</td>
</tr>
<tr>
<td>Direct priming effect</td>
<td>Term frequency,</td>
</tr>
<tr>
<td>Indirect priming effect</td>
<td>Reference frequency</td>
</tr>
<tr>
<td>Spreading activation theory</td>
<td>Term co-occurrence,</td>
</tr>
<tr>
<td>Aspect model</td>
<td>Latent topic</td>
</tr>
<tr>
<td>Recency effect</td>
<td>Recency effect of occurrences</td>
</tr>
<tr>
<td>Primacy effect</td>
<td>Primacy effect of occurrences</td>
</tr>
<tr>
<td>Text Segmentation</td>
<td>Boundary of text segments</td>
</tr>
<tr>
<td>Rhetorical structure theory</td>
<td>Rhetorical relation</td>
</tr>
<tr>
<td>Dynamic semantics</td>
<td>Scope of existential quantifier</td>
</tr>
<tr>
<td>Turn-taking model</td>
<td>Speaker</td>
</tr>
<tr>
<td>Paralinguistics</td>
<td>Paralinguistic information</td>
</tr>
</tbody>
</table>

*: Target of this thesis

salient entity (i.e. center) is successively referred to by the pronoun. Thus, the coherent example satisfies the two rules about referential coherence.

Prototype ideas of the theory were proposed in [7, 8, 9, 10]. Although [6] integrated and assembled the variations, different researchers has proposed various branched versions of centering theory[11]. Although the formalization of centering theory is well-designed, it has not provided any hypothesis of a behavioral principle behind the phenomena. It handle the discourse salience by defining the heuristic ranking among grammatical functions as follows [12, 13]:

12
2.1. REVIEWS ON SALIENCE AND DISCOURSE CONTEXT

(A) Referentially coherent

$U_1$: The licensing company added Johnson & Johnson to its lawsuit over rights to Retin-A acne medicine.

$U_{1,1}^i$: It seeks Johnson & Johnson’s profits from sales of Retin-A, estimated at $50 million.

Follows Rule 1: The pronoun “it” refers to $Cb(U_{1,1}^i) = \text{“the licensing company”}$.
Follows Rule 2: $Cb(U_1^i) = Cb(U_{1,1}^i) = Cp(U_{1,1}^i) = \text{“the licensing company”}$

(B) Referentially incoherent

$U_1$: The licensing company added Johnson & Johnson to its lawsuit over rights to Retin-A acne medicine.

$U_{1,1,1}^i$: Its profits from sales of Retin-A, estimated at $50 million, was sought by the company.

Does not follow Rule 1: The pronoun “its” refers to non-$Cb$ “Johnson & Johnson”.
The non-pronoun “the company” refers to $Cb(U_{1,1}^i) = \text{“the licensing company”}$.
Does not follow Rule 2: $Cb(U_1^i) = Cb(U_{1,1}^i) = \text{“the licensing company”}$

(Retain) $\neq Cp(U_{1,1}^i) = \text{“Johnson & Johnson’s profits”}$

Figure 2.1: Referentially coherent and incoherent examples

**English Cf-ranking:**

Subject $\succ$ Direct object $\succ$ Indirect object $\succ$ Other complements $\succ$ Adjuncts

**Japanese Cf-ranking:**

Topic ($wa$) $\succ$ Subject ($ga$) $\succ$ Indirect object ($ni$) $\succ$ Direct object ($o$) $\succ$ Others

This heuristic definition of salience, Cf-ranking, has the following problems.

1. No statistical basis to justify the Cf-rankings due to its a priori definition
2. Hard to use for calculation because it is not quantitative

3. Hard to integrate with the other salience factors

4. Cover entities only in the last utterance

Salience Referent List (SRL) is an extension of Cf-ranking, which tries to cover entities not only in the last utterance but also in preceding utterances[14]. The extension, however, is also based on a heuristic approach.

2.1.2 Anaphora Resolution

Anaphora resolution, automatic determination of anaphora relation between an anaphor and its antecedent, is a complicated problem in discourse analysis. Discourse salience is one of the features for anaphora resolution because an antecedent tends to be salient in the target context. Methods for anaphora resolution are classified into rule-based approaches and machine learning approaches. Although the mainstream approaches are based on machine learning in recent years[15, 16, 17], the rule-based approaches more emphasize discourse salience than machine learning approaches.

Centering theory has been regarded as a rule-based approach for the anaphora resolution. In addition, SKK (Stock of Shared Knowledge)[18, 19] and RAP (Resolution of Anaphora Procedure)[20] are rule-based approaches intimately related to the discourse salience[21].

SKK represents the hierarchical structure of the salience of discourse entities. It defines the discourse salience as a heuristic function on the basis of a priori rule. The empirical basis of the heuristics, however, has not been clarified.

RAP presents a heuristic method to integrate various salience factors. The integration method is based on the initial weights assigned to each salience factors shown in Table 2.2. The empirical basis of the heuristic score, however, also has not been clarified.
2.1. REVIEWS ON SALIENCE AND DISCOURSE CONTEXT

Table 2.2: RAP: Initial weights assigned to the salience factors

<table>
<thead>
<tr>
<th>Influencing factor of discourse salience</th>
<th>Initial weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence recency</td>
<td>100</td>
</tr>
<tr>
<td>Subject emphasis</td>
<td>80</td>
</tr>
<tr>
<td>Existential emphasis</td>
<td>70</td>
</tr>
<tr>
<td>Accusative emphasis</td>
<td>50</td>
</tr>
<tr>
<td>Indirect object and oblique complement emphasis</td>
<td>40</td>
</tr>
<tr>
<td>Head noun emphasis</td>
<td>80</td>
</tr>
<tr>
<td>Non-adverbial emphasis</td>
<td>50</td>
</tr>
</tbody>
</table>

2.1.3 Recency and Primacy Effects

Recency effect and primacy effect are shown in Figure 2.2[22]. They are cognitive phenomena that recent occurrences and primary occurrences are more likely to be recalled than middle occurrences in a sequence of occurrences. The curve representing probability of recall is termed as the serial position curve. Notice that these effects are confirmed in the free recall task. In a free recall task, trial subjects are given a list of items, usually words, and are asked to remember and to report the list [23]. This situation, however, is different from that of the discourse context that we deal with. Little is yet known about the serial position curve in the discourse context. In particular, we have to clarify the recency effect in the discourse context because we suppose that the discourse salience greatly affected by the recency effect.

2.1.4 Term-Weighting Schemes

The term weighting schemes based on term frequency has been popularly used for various language processing (e.g., document retrieval). TF-IDF [24], proposed to apply to information retrieval, is the representative method. Although various variations have been proposed, the schemes are basically based only on the term frequency.
2.1.5 Direct and Indirect Priming Effects

Priming effect is a cognitive phenomenon wherein the preceding context affects the perceptual importance of the succeeding occurrence [25, 26]. The priming effect is classified into the following types:

- **Direct priming effect**: A succeeding stimulus is primed by the same type of the stimulus in the preceding context.

- **Indirect priming effect**: A succeeding stimulus is primed by the related stimulus in the preceding context.

The term-weighting schemes based on the term frequency correspond to the direct priming effect. The word association corresponds to the indirect priming effect.

2.1.6 Spreading Activation Theory

Spreading activation theory gives a model of the cognitive memory based on the semantic memory network consisting of the word [27, 28, 29, 30, 31]. In
the theory, the activity of a word node corresponds salience. The activity of the word node propagates to the other nodes for each step. Although it well deals with the priming effect, it is hard to be applied to formulate the discourse salience because the semantic memory network is hard to be automatically constructed from the discourse context.

### 2.1.7 Aspect Model

*Aspect model* is a statistic method for unsupervised learning of a statistical language model from corpus. It is based on the assumption that a word $w$ and a document $d$ are stochastically generated from a *latent topic* $z$. The

<table>
<thead>
<tr>
<th></th>
<th>Salience estimation</th>
<th>Language modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation target</td>
<td>entity topicalization</td>
<td>word occurrence</td>
</tr>
<tr>
<td>Transition grain</td>
<td>event (clause)</td>
<td>word</td>
</tr>
</tbody>
</table>

![Figure 2.3: Graphical image of LDA](image)

Figure 2.3: Graphical image of LDA
assumption is represented as the following equation:

\[ p(w|d) = \sum_z p(w|z)p(z|d), \]

where \( p(w|d) \) denotes the probability of \( w \) to be generated in \( d \), \( p(w|z) \) denotes the probability of \( w \) to be generated from \( z \), and \( p(z|d) \) denotes the probability of \( z \) to generate \( d \). The aspect model often has been used for language modeling. The transition grain of the salience estimation that we aim is rougher than that of the language modeling (Table 2.3).

In particular, PLSA (Probabilistic Latent Semantic Analysis)[32] is widely used. The acquired language model help the language processing systems to deal with the indirect priming effect because the related terms in the corpus are automatically allocated to the same latent topic.

LDA (Latent Dirichlet Allocation)[33, 34], an improvement of PLSA shown in Figure 2.3, presents a method to adapt the language model to the target document or the target discourse context.

2.1.8 Text Segmentation

Text segmentation means subdividing texts into multi-paragraph units that represent passages or subtopics. TextTiling is a technique for the text segmentation[35]. The TextTiling algorithm consists of tokenization, lexical score determination, and boundary identification. The boundaries between segments are determined by comparing adjacent two blocks. The blocks are represented as the rectangular windows. The aim of TextTiling algorithm is not to estimate the discourse salience. The boundaries between segments, however, are related to discourse salience. Although this thesis does not deal with the boundaries, we have to consider it in the future.

2.1.9 Rhetorical Structure Theory

Rhetorical Structure Theory (RST) is a theory of discourse structure based on the rhetorical relation (e.g., cause, contrast, elaboration, and question)[36,
2.1. REVIEWS ON SALIENCE AND DISCOURSE CONTEXT

37, 38, 39]. The theory handle the discourse as the RST-tree structure that consists of the text segments and the rhetorical relations between the segments, which is shown in Figure 2.4. Cross-document Structure Theory (CST) is a extended version of RST, which deals with the sentences distributed in cross-documents[40]. The rhetorical structure represents a semantically high-level structure of the discourse context. Although it is an influencing factor of discourse salience, this thesis does not deal with it because automatic parsing of the RST-tree or CST-tree is prohibitively difficult.

2.1.10 Dynamic Semantics

Dynamic semantics (dynamic interpretation) is a logic-based approach that deals with the processes changing the information held by the discourse participants. It deals with the scope of existential quantifier in logical formulas representing the semantics of discourse. The studies of dynamic semantics

![Figure 2.4: Example of rhetorical structure](image-url)
are listed as follows: Discourse Representation Theory (DRT)[41], Context-Change Semantics[42], Dynamic Predicate Logic[43], and Typed Dynamic Logic[44]. Although the scope of existential quantifier is an influencing factor of discourse salience, this thesis does not deal with it because the automatic conversion from discourse to logical formulas is a pretty hard problem.

2.2 Reviews on Information Provision and Discourse Context

Information provision is an important application to help users to successfully communicate with each other because it can help them to share their understanding about their target.

2.2.1 Information Visualization Systems

To support understanding the theme transition, visualization is an effective approach.

There have been some visualizers of chronological change of topics. ThemeRiver presents chronological transition of themes in a collection of documents[45, 46]. Topic Matrix presents cross relationship on the basis of latent contexts in a document set[47, 48]. Although these visualizers deal with chronological transition in a set of documents, they do not deal with the salience transition in a particular document.

There also have been numerous visualizers of the network structures of discourse. LyberWorld presents networks consisting of documents and terms[49]. KeyGraph also presents term network on the basis of term co-occurrence[50]. Network Tsukuru-kun (Mr. Network Extraction) presents a term network extracted from the Web on the basis of co-occurrences of the terms[51]. Conversation Analysis Tool (ChAT) provides a discourse browsing interface based on the network structure of the discourse participants[52].

Although the network structure of discourse is an effective information to
support understanding the discourse, this thesis focuses not on the network structure but on the dynamic transition of discourse salience.

2.2.2 Query-Free Information Retrieval

Query-free information retrieval is a task to find documents that are relevant to a user’s current activity[53, 54]. It requires automatic creation of queries from the user’s activities.

In particular, to support debate participants to know the diverse viewpoints related to the current agenda, we need to automatically create queries from their current discussion context. Although the conventional studies of query-free retrieval are informative guide, the query creation from the users’ temporal discourse context has not been studied yet.

2.2.3 Topic Detection and Tracking

Topic Detection and Tracking (TDT) is an event-based information organization task[55, 56]. It is to develop automatic identification of topically related stories within a stream of news media.

To support debate participants to know the diverse viewpoints related to the current agenda, we need to automatically identify information topically related to their current discussion context. Although the conventional studies of TDT are informative guide, the topic detection according to the users’ temporal discourse context has not been studied yet.

2.2.4 Discussion Analysis

There are many methodologies to analyze discussion. For instance, visualizing discussion structure as a network is an effective approach to analyze discourse. Discussion Structure Visualizer provides the network of discussion structure consisting of text segments on the basis of KeyGraph and Spreading Activation[57]. Semantic Authoring, which is an authoring platform based on the network structure consisting of ontologies and rhetorical
relations, can be used as a discussion platform[58]. The discussion analysis on the Semantic Authoring has been researched[59].

In particular, for the Public Involvement (PI) processes, corpus-based discussion analysis is needed[5, 60]. Public debates need technological support for discussion because the participants sometimes argue on different planes due to their diverse backgrounds. Recently, discussion analysis for face-to-face offline meetings has been carried out[61]. Such technologies are desired to become a social infrastructure to support the PI process.

### 2.2.5 Call Center Support Systems

Call center provides a telephone-based human-service operation[62]. An appropriate operation only by speech is sometimes difficult because the customers and the service agents are remote from each other. We consider the call center an applicable field of the discourse support technologies. Although management of staffing, scheduling[63], and automatic call routing[64, 65] have been focused on in the studies for call center support systems, supporting to share the conversational context of service agents and customers has not been focused on.

### 2.3 Reviews on Game-theoretic Pragmatics

*Game theory* is a mathematical theory about interaction among multiple players (i.e., autonomous agents) [66, 67]. It formulates the player’s rational behavior in various strategic interactions.

<table>
<thead>
<tr>
<th>Table 2.4: Prisoner’s Dilemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>prisoner A</td>
</tr>
<tr>
<td>cooperater</td>
</tr>
<tr>
<td>betray</td>
</tr>
</tbody>
</table>

Nash equilibrium
The *Prisoner’s Dilemma* is a well-known strategic situation in which two players may cooperate with or betray each other. Table 2.4 is its payoff matrix. The values in this matrix represent players’ benefits. In this game, betrayal is the individually optimal strategy of each player in response to any strategy of the other player: Each player gains more by betraying whether the other player cooperates or betrays. Generally speaking, such a combination of individually optimal strategies is called the *Nash equilibrium*. That is, the Nash equilibrium is a combination of players’ strategies in which no player is motivated to change her strategy as far as the other players keeps their strategies. The equilibrium in Prisoner’s Dilemma, however, does not maximize the players’ benefits in total. Hence, individual optimization does not always conform to social optimization.

On the other hand, communication is an inherently collaborative situation. For instance, even enemies must cooperate in order to convey their animosities to each other. To formulate this, let \( i \) be the proposition that the sender \( S \) intends to communicate some semantic content \( c \) to the receiver \( R \). Then \( i \) entails that \( S \) intends that \( R \) should both recognize \( c \) and believe \( i \). This is the core of Grice’s *nonnatural meaning*\(^{[68, 69]}\).\(^{1}\) This restricted sense of nonnatural meaning implies that communication is inherently collaborative, because both \( S \) and \( R \) want \( R \) to recognize \( c \) and \( i \). \( S \) of course wants it, and so does \( R \) because it is generally beneficial to know what \( S \) intends to make \( R \) believe or obey.

Discourse is a kind of strategic interaction. In recent years, there have been numerous game-theoretic studies on communication, pragmatics, and discourse analysis\(^{[70, 71, 72]}\).

\(^{1}\)Grice’s original notion of nonnatural meaning further entails (\( S \)’s intention of) \( R \)’s believing (when \( c \) is a proposition or a reference) or obeying (when it is an order or a request) \( c \), but we disregard this aspect here.
Table 2.5: Typical applications of the signaling game

<table>
<thead>
<tr>
<th></th>
<th>Job market</th>
<th>Mate selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sender S</td>
<td>Job seeker</td>
<td>Male deer</td>
</tr>
<tr>
<td>Receiver R</td>
<td>Employer</td>
<td>Female deer</td>
</tr>
<tr>
<td>Type T</td>
<td>Competence</td>
<td>Vitality</td>
</tr>
<tr>
<td>Message M</td>
<td>Education</td>
<td>Antler size</td>
</tr>
<tr>
<td>Action A</td>
<td>Hiring</td>
<td>Mating</td>
</tr>
</tbody>
</table>

2.3.1 Signaling Game

A signaling game consists of S’s sending a message (a signal) to R and R’s doing some action in response to that message. In this game, S should send a costly message to get a large payoff. Table 2.5 shows some typical cases of signaling game. In job market signaling[73], a job seeker S signals her competence (type) to a potential employer R with the level of her education as the message, and R decides the salary to offer her. A competent job seeker tends to be highly educated, and the potential employer will offer her a high salary. In mate selection[74], a male deer S indicates its strength to potential mate R by the size of its antlers. A strong deer will grow extra-large antlers to demonstrate its survival competence despite the handicap of keeping those practically useless antlers — the handicap principle[75].

2.3.2 Game-theoretic Disambiguation

One linguistic expression may have multiple meanings depending on multiple contexts[76]. This efficiency of language presupposes disambiguation of meanings conveyed by linguistic messages. Here, let $m_1$ and $m_2$ be linguistic messages and $c_1$ and $c_2$ be their content candidates. Suppose that both $c_1$ and $c_2$ can be referenced by $m_1$, and that only $c_2$ can be referenced by $m_2$, then $S$ and $R$ can determine the optimal correspondence such that $m_1$ refers to $c_1$ and $m_2$ refers to $c_2$. Parikh explains this disambiguation in game-theoretic terms as a collaboration between $S$ and $R$[77, 70]. This sort of collaborative
2.3. REVIEWS ON GAME-THEORETIC PRAGMATICS

disambiguation may extend to referential coherence in general, but Parikh does not consider that direction.

2.3.3 Meaning Game

The meaning game is a game-theoretic framework to formulate intentional communication (e.g., linguistic communication) \[78, 79\]. The meaning game is a model to capture the core of non-natural meaning on the basis of game-theoretic principles. Let $C$ be the set of semantic contents and $M$ be the set of the linguistic messages. The sender $S$ sends a message $m \in M$ to convey content $c_S \in C$. The receiver $R$ interprets $m$ as meaning $c_R \in C$. Therefore, a turn of communication is represented as $\langle c_S, m, c_R \rangle \in C \times M \times C$. $c_S = c_R$ is a necessary condition for this turn of communication to be successful.

The utility function $U_{tX}$ of player $X$ would thus be a real-valued function from $C \times M \times C$ (the set of turns). Only if the communication is successful are $U_{tS}$ and $U_{tR}$ mutually positive. $Pr(c_S, m, c_R)$ is the probability that a turn $\langle c_S, m, c_R \rangle$ occurs (i.e., $S$ selects $\langle c_S, m \rangle$ and $R$ selects $\langle m, c_R \rangle$).

The meaning game assumes the following:

1. $c_S = c_R$, i.e., $S$ and $R$’s communication is always successful, because $U_{tS}$ and $U_{tR}$ are mutually positive only if the communication is successful.

2. $S$ and $R$’s common belief includes the distribution of $Pr(c_S, m, c_R)$.

The expected utility of player $X$ in the succeeding turn is defined as follows:

$$\sum_{\langle c_S, m, c_R \rangle \in C \times M \times C} Pr(c_S, m, c_R) U_{tX}(c_S, m, c_R)$$

$S$ and $R$ respectively decide the strategy $x = \langle c_S, m, c_R \rangle$ according to their expected utilities. Under the above assumptions, $S$ and $R$ cooperatively decide the strategy $x = \langle c, m, c \rangle$ for success of communication.
CHAPTER 2. LITERATURE REVIEW

Furthermore, the assumptions lead the speculation that $S$ and $R$ select the effective solution $x$ that should be a *Pareto-optimal evolutionarily stable strategy (ESS)* [80] satisfying the following conditions:

$$\text{Ut}_X(x, x) \geq \text{Ut}_X(y, x) \quad \text{for all } y$$
$$\text{Ut}_X(y, x) = \text{Ut}_X(x, x) \Rightarrow \text{Ut}_X(y, y) < \text{Ut}_X(x, y)$$

Here, $\text{Ut}_X(x, y)$ represents the utility of $X$ when $X$’s strategy is $x$ and the other player’s strategy is $y$. The first condition indicates that an ESS is a Nash equilibrium. In this way, the social optimization of intentional communication can be reduced to the game theoretic account.

---

2When most players’ strategy is an ESS $x$, any other strategy $y$ is disadvantageous.
Chapter 3

GDA Software Development Kit

We need tools for corpus management because modeling discourse context requires corpora in order to evaluate and to verify the model. To formulate discourse context, we use corpora annotated with Global Document Annotation (GDA) tags. Hence, we develop tools for handling GDA corpora.

GDA is an XML vocabulary designed for representing linguistic structure[81, 82]. It can represent morphemic information, syntactic one, dependency one, anaphoric one, and rhetorical one. We need develop a library for processing GDA because implementing a system processing GDA is not easy. To prepare and manage required corpora for modeling discourse context, this chapter describes development of a Java package for processing GDA. We term the Java package the GDA Software Development Kit (GDASDK), which includes an API (Application Programming Interface) for handling GDA, lightweight viewers for GDA, and automatic annotation tool for GDA. It is available on our web site\(^1\).

\(^1\)http://winnie.kuis.kyoto-u.ac.jp/~siramatu/gda/gdasdk.html
CHAPTER 3. GDA SOFTWARE DEVELOPMENT KIT

3.1 Specification of GDA Tag set

GDA nodes are organized as a DOM (Document Object Model) tree[83] represented by GDA tags inserted between morphemes. This section describes how to interpret the GDA tree structure.

3.1.1 Dependency Structure

Dependency structure is represented as relationship between sibling nodes. To explain the specification to represent dependency, we have to describe two types of GDA elements: head elements and phrase elements.

- **Head element**: A type of element that can be depended by the other element. `<n>`, `<v>`, `<aj>`, `<ad>`, `<seg>`, and so on.

- **Phrase element**: Another type of element that cannot be depended by the other element. `<np>`, `<vp>`, `<ajp>`, `<adp>`, `<segp>`, and so on.

Basically, a phrase element depends on a sibling head element. For the following example, `<np>現行制度での準備<np>`, a phrase element, depends on the sibling head element `<ad>も</ad>.

```
<adp>
  <np>現行制度での準備<np><ad>も</ad></adp>
  <v>怠らない</v>
```

Additionally, `<adp>現行制度での準備も<adp>`, also a phrase element, depends on the sibling head element `<v>怠らない</v>`. When a parent element has multiple head elements, the dependency structure in the child nodes has ambiguity (unless the direction of the dependency is set by attribute syn="f" or syn="b"). Hence, an element is desired to have only one head element in its children.
3.1.2 Anaphoric Structure

Anaphoric structure is represented by *id* attribute and relation attributes. An element annotated with the *id* attribute represents an antecedent. An element annotated with the relation attribute represents an anaphor. For the following example, `<persname id="KonoYohei">河野洋平</persname>` represents an antecedent and `<v agt="KonoYohei">示し</v>` represents a zero anaphora as an agent of action.

```
<su syn="f">
<adp opr="topic.fit.agt">
  <persname id="KonoYohei">河野洋平</persname> ⋯①
  氏は</adp>
<adp opr="cnt"><q>「新制度の下で戦う」</q>と</adp>
<v紧张></v>。
</su>
```

("We will fight under the new system", Mr. KONO Yohei emphasized.)

```
<su syn="f">
<adp>しかし</adp>,
<adp opr="obj">現行制度での準備も怠らない構えを</adp>
<v agt="KonoYohei">示し</v> ⋯②
た。</su>
```

("But φ(he) also indicated an intention to keep ready for the current system.

In this example, *agt* is a *relation identifier* that represents an agent of the verb “示し”. A relation identifier takes two arguments. For the above instance, the first argument of *agt* is the verb “示し” and the second one is the antecedent “河野洋平”. GDA tag set contains many relation identifiers that represent deep cases or various linguistic relationships between concepts[81]. Let us partially list up the capital ones.

*agt*  Agent of action.
CHAPTER 3. GDA SOFTWARE DEVELOPMENT KIT

**obj** Object of action or event.

**res** Result, which is another special case of **obj**. A resulting event or object.

**src** Source. The second argument of **src** is the initial position or state of the entity denoted by the subject or object of the verb denoting the first argument of **src**.

**gol** Goal. The second argument of **gol** is the final position or state of the entity denoted by the subject or object of the verb denoting the first argument of **gol**.

**exp** Experiencer.

**pos** Possessor.

**cnt** Content of thought, belief, speech, promise, rumor, plan, request, and so forth.

**eq** Equivalence.

**arg** Primary or unique argument, such as the arguments of auxiliary verbs, prepositions, and relational nouns.

The relation identifiers of GDA are used to represent not only direct anaphora but also indirect anaphora and rhetorical relation.

### 3.2 Developing API for GDA

To facilitate implementation of systems processing GDA, we should define API (Application Programming Interface) for processing GDA. The API can be defined on the basis of DOM (Document Object Model)[83]. As a foundation for the GDASDK, we used Xerces2-J, a DOM parser implemented for Java[84].
3.2. DEVELOPING API FOR GDA

DOM defines the interface `org.w3c.dom.Node` including methods to process XML nodes. It, however, does not include methods for processing GDA, e.g., processing dependency and anaphoric structure. Hence, we define an interface `jp.go.aist.gda.dom.GDANode` that extends `org.w3c.dom.Node`. In the interface `GDANode`, the principal methods for dependency structure is as follows:

- **getGovernor()**: Return a GDANode depended by this node, i.e., a node governing this node.
- **getDependants()**: Return a list of GDANode that this node depends on.
- **canGovern()**: Return true if this node can be depended on by other nodes. For example, return true if this node is a head element. Return false if this node is a phrase element.
- **canDepend()**: Return true if this node can depend on another node.
- **dependsOn(GDANode anode)**: Return true if this node depends on the anode.
- **getHeadLeaf()**: Return a leaf node that represents a head morpheme of this node.
- **getSentence()**: Return a sentence element that includes this node.
- **isHeadOfParent()**: Return true if this node is a head of its parent node.

We implemented these methods in the classes `jp.go.aist.gda.dom.GDAElement` and `jp.go.aist.gda.dom.GDAText`. `GDAElement` and `GDAText` further include implemented methods in interfaces `org.w3c.dom.Element` and `org.w3c.dom.Text`, respectively. Furthermore, this API also includes utility classes and methods for anaphoric structure, part-of-speech, XPath, and so on. This API is included in the GDASDK that we developed.
3.3 Developing GDA Viewers

To manage GDA corpora, we developed lightweight software to visualize structure of GDA on the basis of the API mentioned above. Figures 3.1 and 3.2 show the software. When we carry out experiments by using GDA corpora, we use these viewers to investigate details of data. Figure 3.1, GDAViewer, shows a syntactic tree structure represented as a hierarchy of GDA tags. It is a standard way for visualization. Figure 3.2, SemGraphViewer, represents a graph structure of dependency and anaphora. The graph structure is converted from the tree structure by using the API for GDA. The SemGraphViewer can more effectively represent the dependency structure than the former GDAViewer. These lightweight viewers are included in the GDASDK that we developed.
3.3.1 GDA Tagging Systems

We developed an automatic annotation system for GDA on the basis of dependency structure analyzed by CaboCha [85, 86]. CaboCha [85, 86] is a dependency parser for Japanese. Our annotation system just converts the output format of CaboCha to the GDA format. The problem on the conversion is the difference of the formats between GDA and Cabocha. GDA handles the dependency between morphemes, chunks, and phrases (see also Figure 3.2). On the other hands, the result by CaboCha represents dependency between chunks. More concretely, although GDA requires a semantic head morpheme in each chunk, the output format of CaboCha does not represent it. Hence, our system restores the semantic head in chunks. The automatic annotation system is included in the GDASDK that we developed.

Furthermore, we also need an authoring tool for manual annotation. GDA Tagging Editor, a manual authoring tool, had been developed at AIST through RWC and CREST projects. It will soon become available on the web of Gengo Shigen Kyokai (GSK) [87].

3.4 Extending CSJ Corpus

CSJ (Corpus of Spontaneous Japanese) consists of spoken discourses [88]. It includes manual dictations, phonological information, and morphological information. Although we need dependency structure and anaphoric one to formulate discourse context, CSJ does not include them. Hence, we extend CSJ by using the GDASDK and GDA Tagging Editor.

We selected four spontaneous dialogues (D03F0001, D03M0013, D03F0040, D03M0048) from CSJ with considering the length of dialogue and the balance of gender. Firstly, we automatically annotated the dialogue data with dependency tags by using the GDASDK. Secondly, we let a human operator (who has an experience in GDA annotation) manually annotate the dialogue data with anaphoric tags. The details of the manual annotation are described in Chapter 6.
Chapter 4

Statistical Formulation of Discourse Salience

In this chapter, we provide a corpus-based formulation of discourse salience, which incorporates the dynamic transition along with the discourse progress.

4.1 Introduction

This section describes what requirements we have to fulfill in order to formulate the discourse salience and what approaches we take to satisfy the requirements.

4.1.1 Requirement and Issue

To quantitatively formulate the discourse salience, we have to satisfy the following three requirements:

[Requirement 1] Deal with the dynamic change of salience: Discourse salience dynamically changes along with the progress of discourse. Hence, it is indispensable to estimate the salience value for each utterance unit. To satisfy this requirement, it is essentially im-
important to take into account the recency effect (mentioned in Chapter 2) of the expressions referring to the entity.

[Requirement 2] **Integrate the various influencing factors:** Discourse salience is influenced by not only the recency effect, but also various factors, i.e., term frequency, type of grammatical function, type of named entity, anaphora, primacy effect, and latent topic. To accurately estimate salience, we have to formulate salience through integrating these factors. Furthermore, to optimally integrate the influencing factors, we need a quantitative criterion to evaluate the integration method.

[Requirement 3] **Calculate automatically:** To apply the salience formulation to information provision or discourse analysis, it is desirable to automatically calculate the salience value by using present technologies of language processing. Notice that although the salience calculation must be automatic, creation of training data for optimizing salience formulation can be manual. We consider the manual annotation of corpora an allowable labor for supervised learning.

### 4.1.2 Our Approach

To satisfy the three requirements, we take the following approaches:

[Approach 1] **Window function:** To take into account the recency effect, we adopt window function that corresponds to the *serial position curve* [22, 89, 90] in the discourse structure. The serial position curve represents the recency and primacy effects in the short-term memory (Figure 4.1). We assume that the “recency part” of the serial position curve shift according to the discourse progress. We can implement this assumption by using a *window function* borrowed from the field of signal processing. The window function corresponding to the recency effect is shown in Figure 4.2. We empirically find an optimal window function that corresponds to the recency part in the serial position curve.
4.1. INTRODUCTION

[Approach 2] **Probabilistic formulation**: To integrate various influencing factors, we formulate discourse salience by a probabilistic approach. We assume that the more salient an entity is, the more frequently it becomes referred to in the succeeding utterance. This idea is backed up by the rules about the referential coherence of centering theory. Con-
cretely, we define a scale of salience as the probability of an entity to be referred to in the succeeding utterance to integrate the influencing factors of salience. Furthermore, we design a evaluation criterion of salience scales to optimize our salience formulation.

[Approach 3] **Automatically extractable features:** To enable us to automatically calculate the salience value, we employ automatically extractable features for calculation. Although the optimization of salience formulation needs the manually annotated information of anaphors’ referent, the calculation of the salience value should not need such manual information.

Concisely, the following sections in this chapter describe the salience formulation, the calculation method, and the evaluation scale of salience calculation according to the just mentioned approaches.

### 4.2 Formulating Salience: Reference Probability

We assume that discourse participants focus their attention on the discourse entity $e$ that is likely to be referred to in the succeeding utterance unit $U_{i+1}$. This assumption is consistent with the tendency of referential coherence, which is formalized by centering theory\[6\]; i.e., the saliently referenced entity tends to be successively referred to in the succeeding utterance. Under this assumption, the discourse salience of $e$ at the moment when $U_i$ is conveyed can be formulated as follows:

\[
(Salience \ of \ e \ at \ U_i) = \Pr(\exists w \xrightarrow{ref} e \ in \ U_{i+1}\mid\text{pre}(U_i)) \\
= \Pr(e|\text{pre}(U_i))
\]

Here and hereafter, we let $w \xrightarrow{ref} e$ denote that the referring expression $w$ refers to the discourse entity $e$, and $\text{pre}(U_i)$ denote $[U_1, \ldots, U_i]$, a preceding
4.2. FORMULATING SALIENCE: REFERENCE PROBABILITY

We term the probability $\Pr(\exists w \xrightarrow{\text{ref}} e \text{ in } U_{i+1}\mid \text{pre}(U_i))$ the reference probability of $e$ at the moment when $U_i$ is conveyed. Let $\Pr(e\mid \text{pre}(U_i))$ be the simplified notation of the reference probability. This probabilistic formulation corresponds to the Approach 2 described above. It means the discourse salience of $e$ can be measured as the probability of $e$ being referred to in the succeeding utterance $U_{i+1}$. For example, Figure 4.3 shows $\Pr(\text{Kyoto}\mid \text{pre}(U_i))$, the salience of “Kyoto” given the context of pre($U_i$).

Although centering theory defines the discourse salience only by grammatical function, we have to take into account various influencing factors of the discourse salience (as described in Chapter 2).

The salience formulation $\Pr(e\mid \text{pre}(U_i))$ has a theoretical ground as above. Then, how to calculate $\Pr(e\mid \text{pre}(U_i))$? The following section describes the statistical calculation based on a corpus.
### 4.3 Calculation Method

The reference probability $\text{Pr}(e|\text{pre}(U_i))$ is calculated by using a regression model obtained from a training corpus, which is annotated with anaphora information. In other words, the calculation of $\text{Pr}(e|\text{pre}(U_i))$ requires the training phased as a preparation before the calculation. Notice that although the training phase requires a corpus annotated with anaphora information due to the definition of $\text{Pr}(e|\text{pre}(U_i))$, the practical calculation phase does not require the anaphora annotation because we use only automatically extractable features without anaphora annotation in the practical calculation phase.

**Basic Idea**

Here, let $\langle e, U_i \rangle$ be a target example to calculate $\text{Pr}(e|\text{pre}(U_i))$, where $e$ denotes a target discourse entity and $U_i$ denotes the utterance conveyed at a target moment. Basically, $\text{Pr}(e|\text{pre}(U_i))$ is calculated by using samples $\langle e', U'_j \rangle$ extracted from a training corpus as follows:

$$
\text{Pr}(e|\text{pre}(U_i)) = \frac{\#\{(e', U'_j); \text{eqfeat}(\langle e', U'_j \rangle, \langle e, U_i \rangle) \land \exists w \xrightarrow{\text{ref}} e' \text{ in } U'_{j+1}\}}{\#\{(e', U'_j); \text{eqfeat}(\langle e', U'_j \rangle, \langle e, U_i \rangle)\}}
$$

where

$$
\text{eqfeat}(x, y) = \begin{cases} 
\text{True} & (\text{features of } x \text{ equal those of } y) \\
\text{False} & (\text{otherwise}) 
\end{cases}
$$

This basic idea, however, cannot be used for practical calculation due to data sparseness. We have to cope with data sparseness in calculation because the samples in a training corpus are not always densely distributed in the feature space. To cope with the data sparseness, we employ regression analysis as mentioned in the following subsections.

Additionally, let us notice that the characteristic of training corpus should not be so dissimilar from that of the target discourse. For instance, if the
4.3. CALCULATION METHOD

Figur e 4.4: An example method for calculating \( \text{Pr}(e | \text{pre}(U_i)) \)

target discourse is conversation, the training corpus is desired to consist of conversations also.

Calculation Procedure

Figure 4.4 shows an example method for calculating \( \text{Pr}(e | \text{pre}(U_i)) \) through integrating features by using logistic regression\[91\]. The procedures of the training phase and the practical calculation phase are as follows:

[Training Phase]

1. **Sample set extraction**: Extract a sample set \( \{ \langle e, U_i \rangle; \exists w \xrightarrow{\text{ref}} e \text{ in } \text{pre}(U_i) \} \) from the training corpus annotated with anaphora information.

2. **Feature extraction**: Extract features of referring expressions \( w \xrightarrow{\text{ref}} e \) in \( \text{pre}(U_i) \) for each sample \( \langle e, U_i \rangle \).
3. **Feature windowing**: Apply a window function to each feature value in order to incorporate the curve of the recency effect.

4. **Feature integration (Learning)**: Train a regression model to calculate $\Pr(e|\text{pre}(U_i))$ by integrating the features extracted from the training corpus.

**[Practical Calculation Phase]**

1. **Sample extraction**: Extract a target sample $\langle e, U_i \rangle$ to calculate $\Pr(e|\text{pre}(U_i))$ from the target discourse context, which does not need anaphora annotation.

2. **Feature extraction**: Extract the features from the target sample $\langle e, U_i \rangle$.

3. **Feature windowing**: Apply a window function to each feature value.

4. **Feature integration (Calculation)**: Calculate $\Pr(e|\text{pre}(U_i))$ by using the obtained regression model. In other words, integrate the features into the value of $\Pr(e|\text{pre}(U_i))$ with the regression.

Let us detail the sample extraction, the feature extraction, the feature windowing, and the feature integration.

**4.3.1 Extracting Samples**

For example of Figure 4.5, when the current utterance $U_i$ is assumed to be $U_{273}$, the samples listed in the figure are extracted. In the training phase, extract all samples $\langle e, U_i \rangle$, which comprise every combination of $e$ and $\text{pre}(U_i)$ through shifting $U_i$, from the training corpus. In the practical calculation phase, extract a target sample $\langle e, U_i \rangle$ for calculating $\Pr(e|\text{pre}(U_i))$. 
4.3.2 Extracting Features

We use only automatically extract for automatic calculation. Although Figure 4.6 shows a complete co-reference chain of “京都 (Kyoto)”, we cannot use its complete features because it contains zero pronouns, which are hard to be automatically extracted. We extract the features only from the superficial occurrence chain instantiated in Figure 4.7. Figure 4.5 shows their differences in respect of the examples of feature values.

Candidate features

Table 4.1 shows the candidate features we deal with. These are automatically extractable features according to Approach 3 above mentioned. The feature set for calculation should be empirically optimized on the basis of corpus in Chapter 6.

![Figure 4.5: Extracting sample \( \langle e, U_i \rangle \) to calculate \( \Pr(e|\text{pre}(U_i)) \)
Assigning real number to discrete feature

In case of discrete features (e.g. grammatical role and part-of-speech), we need to assign a real value to each discrete feature in order to integrate the features by using the logistic regression.

Here, let $w$ be a referring expression that refers to the entity $e$ in $\text{pre}(U_i)$, i.e., $w \xrightarrow{\text{ref}} e$ in $\text{pre}(U_i)$. Let $\text{feature}(w)$ be a discrete feature of $w$. When $\text{feature}(w) = x$, we assign the real number $\text{avgPr}(x)$ to a particular discrete feature $x$ as follows:

$$
\text{avgPr}(x) = \frac{\#\{\langle w, U_i \rangle; w \xrightarrow{\text{ref}} e \text{ in } \text{pre}(U_i) \land \text{feature}(w) = x \land \exists w' \xrightarrow{\text{ref}} e \text{ in } U_{i+1}\}}{\#\{\langle w, U_i \rangle; w \xrightarrow{\text{ref}} e \text{ in } \text{pre}(U_i) \land \text{feature}(w) = x\}}
$$
4.3. CALCULATION METHOD

4.3.3 Windowing Features

In case of a feature depending on the recency effect, the recent referring expressions \( w \xrightarrow{\text{ref}} e \) are more important than the other ones. We can deal with this by incorporating window function \( W(\text{dist}) \) shown in Figure 4.9 (where \( \text{dist} = \text{dist}(w, U_{i+1}) \) denotes the utterance distance from \( w \xrightarrow{\text{ref}} e \) to \( U_{i+1} \), i.e., \( (i + 1) - j \) when \( w \) is in the utterance \( U_j \)). This is according to Approach 1 above mentioned.
### Table 4.1: Elements and candidate features for calculating reference probability

<table>
<thead>
<tr>
<th>Element for calculation</th>
<th>Influencing factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition of Pr(e</td>
<td>pre(U_i))</td>
</tr>
<tr>
<td>Window function W(dist(w, U_{i+1}))</td>
<td>Recency effect of w</td>
</tr>
<tr>
<td>[ \sum_{w \xrightarrow{\text{ref}} e \text{ in } \text{pre}(U_i)} \left( W(\text{dist}(w, U_{i+1})) \times \text{(feature of } w) \right) ]</td>
<td>Frequency of recent w</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Candidate feature for calculation</th>
<th>Influencing factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>dist: Utterance distance from (w \xrightarrow{\text{ref}} e) to U_{i+1}</td>
<td>Recency effect of w</td>
</tr>
<tr>
<td>freq: (\frac{1}{i}) (Frequency of (w \xrightarrow{\text{ref}} e) in \text{pre}(U_i))</td>
<td>Frequency of w</td>
</tr>
<tr>
<td>gram: Function word depended by (w \xrightarrow{\text{ref}} e)</td>
<td>Grammatical function of w</td>
</tr>
<tr>
<td>pos: Part-of-speech of (w \xrightarrow{\text{ref}} e)</td>
<td>Type of named entity of w</td>
</tr>
<tr>
<td>title: Whether e is referred to in a title</td>
<td>Primacy effect of e</td>
</tr>
</tbody>
</table>

### 4.3.4 Feature Integration: Learning and Calculation

We adopt logistic regression to integrate features and to calculate Pr(e|pre(U_i)). Logistic regression requires a training phase, i.e., an estimation of the regression weights. Given that \{feature_1((e, U_i)), \ldots, feature_n((e, U_i))\} is the feature set for calculation, the regression weights satisfy the following logit formula:

\[
\log \frac{\text{Pr}(e|\text{pre}(U_i))}{1 - \text{Pr}(e|\text{pre}(U_i))} = b_0 + \sum_{k=1}^{n} b_k \cdot \text{feature}_k((e, U_i)).
\]

The regression weights are calculated with the maximum-likelihood method by using training corpus. The dummy variable to train the regression model is defined as follows:
4.4 Evaluation Criteria

To optimize a calculation method of the salience, we need evaluation criteria. Hence, we define the corpus-based evaluation measure of a calculation method $m$.

**Figure 4.9: Feature windowing according to $\text{dist}(w, U_{i+1})$**

$$\text{isRef}(e, U_{i+1}) = \begin{cases} 1 & (\exists w \xrightarrow{\text{ref}} e \text{ in } U_{i+1}) \\ 0 & \text{(otherwise)} \end{cases}$$

That is to say, we can calculate $\Pr(e|\text{pre}(U_i))$ by using the logistic regression model that explains $\text{isRef}(e, U_{i+1})$. By using the trained regression model, $\Pr(e|\text{pre}(U_i))$ can be calculated by

$$\Pr(e|\text{pre}(U_i)) = \left(1 + \exp\left(-\left(\sum_{k=1}^{n} b_k \cdot \text{feature}_k(\langle e, U_i \rangle)\right)\right)^{-1}.$$
Given that sal\(_m(e|\text{pre}(U_i))\) is a calculated value by \(m\), the evaluation measure \(\text{evalSal}(m)\) is defined on the basis of the test-set corpus as follows:

\[
\text{evalSal}(m) = \text{cor}\left(\left[\text{sal}_m(e|\text{pre}(U_i))\right]_{(e,U_i)}, \left[\text{isRef}(e, U_{i+1})\right]_{(e,U_i)}\right),
\]

where \(\text{isRef}(e, U_{i+1})\) is the dummy variable defined above (i.e., 1 if \(\exists w \xrightarrow{\text{ref}} e\) in \(U_{i+1}\), otherwise 0), and \(\text{cor}(x, y)\) denotes Pearson’s correlation coefficient between \(x\) and \(y\). Notice again that \(\text{evalSal}(m)\) is calculated not by using a training corpus, but by using the test-set corpus.

The evaluation scale \(\text{evalSal}(m)\) can be interpreted as “how accurate \(m\) can predict whether \(e\) is referred to in \(U_{i+1}\)”. When \(\text{sal}_m(e|\text{pre}(U_i)) = \text{Pr}_m(e|\text{pre}(U_i))\), i.e., the reference probability calculated by \(m\), \(\text{evalSal}(m)\) corresponds to \(m\)’s approximation accuracy for the definition of the reference probability.

### 4.5 Conclusion

In this chapter, we formulated the reference probability \(\text{Pr}(e|\text{pre}(U_i))\) as the discourse salience of the entity \(e\) at the moment when \(U_i\) is conveyed. We proposed a regression-based method to calculate \(\text{Pr}(e|\text{pre}(U_i))\). This method enabled us to integrate features. In particular, to deal with dynamic transition, we incorporated the recency effect as a feature. To incorporate the recency effect, we adopted the window function shifting with each utterance unit. Furthermore, we proposed \(\text{evalSal}(m)\) as an evaluation scale of a target method to calculate salience. This enabled us to empirically evaluate our formulation in Chapter 6.
Chapter 5

Meaning-Game-based Centering Model

This chapter describes a game-theoretic formulation of referential coherence. We term the formulation “Meaning-Game-based Centering Model”.

5.1 Introduction

Discourse structure essentially consists of cooperative use of language among communicative participants to extend their common beliefs (e.g., shared discourse context). A rational individual prefers to cooperatively use language so as to communicate information to her interlocutor. At the same time, she prefers to simplify her utterances within the inferable range for her interlocutor. For instance, they cooperatively prefer referential coherence (i.e., smooth transition of attention between adjacent utterances) resulting from topic continuity and pronominalization. They tend to continue the same topic of discourse because the continuing topic can be easily predicted. They also tend to pronominalize (i.e., to simplify or to ellipt) the continuing topics within the inferable range because such contextually correct pronominalization reduces their perceptual loads. The goal of this study is theoretical formulation of the behavioral principle behind referential coherence.
CHAPTER 5. MEANING-GAME-BASED CENTERING MODEL

Centering theory [6] is a standard theory of referential coherence. It explains referential coherence with heuristic rules about salience, topic continuity, and pronominalization. The rules are helpful in automatic generation or compilation of referentially coherent discourse. In this paper, we deal with the following two theoretical drawbacks. (1) Centering theory gives no principle behind the cooperative behavior of rational individuals. We aim to formulate the behavioral mechanism behind referential coherence in various languages. (2) Heuristic rules of centering theory are not designed on statistical grounds. In order to absorb language-specific differences, (e.g., grammatical functions, pronoun types, etc.), a corpus-based design is more desirable than a heuristic design. We aim to statistically formulate the perceptual factors related to referential coherence (i.e., degrees of salience and pronominalization).

Figure 5.1 shows our approaches to these two issues. (1) We suppose that the behavioral principle in the cognitive process independent from language can be formulated in terms of game theory. (2) We also suppose that the low-level perceptions depending on language-specific expressions can be formulated by using corpus statistics. In resolving these issues, we aim to construct a bottom-up formulation of the cognitive principle behind referential coherence from low-level perceptions.

Figure 5.1: Two issues
5.1. INTRODUCTION

We believe that the behavioral principle behind referential coherence in various languages can be formulated in game-theoretic terms. Game theory is a mathematical theory about interaction among multiple players (i.e., autonomous agents) [66, 67]. It formulates the player's rational behavior in various strategic interactions. Discourse is a kind of strategic interaction. In recent years, there have been numerous game-theoretical studies on pragmatics or discourse analysis [70, 71, 72].

The meaning game [78, 79] is a game-theoretical framework to formulate communication in which the sender sends a message and the receiver attempts to infer its intended meaning. Hasida demonstrated that centering theory can be derived from the principle of expected utility in the meaning game framework. In other words, his hypothesis has the potential to attribute the behavioral principle behind referential coherence to a game-theoretic principle. We suppose that the hypothesis is commonly valid in various languages. To empirically verify the universality of the hypothesis, we statistically formulate Hasida's derivation. The formulation is expected to enable quantitative and game-theoretical analyses of referential coherence in various languages.

Furthermore, from a viewpoint of engineering, various discourse systems (e.g., dialogue systems, summarization systems, etc.) are required to output a referentially coherent discourse. Such systems need to select a referentially coherent expression from possible candidates of the succeeding sentence. Thus, we aim to quantify how restricted the system's selection of the expression is by the preference of referential coherence. We suppose that such systems can use the expected utility of the meaning game as a strength of preference in order to determine the succeeding output utterance. Hence we investigate the characteristics of expected utility as a quantitative scale of referential coherence between the preceding discourse \([U_1, \ldots, U_i]\) and the succeeding utterance \(U_{i+1}\). Notice that our aims are different from those of machine-learning-based studies about anaphora resolution [15, 16, 17]. We aim to develop a quantitative criterion to cooperatively select the expression
and the interpretation of the succeeding utterance.

### 5.2 Issues on Centering Theory


$U_i$ represents the $i$-th utterance unit (clause) in the target discourse. Entities referred to in $U_i$ are ordered by Cf-ranking, i.e., the salience ranking of grammatical functions, dependent on the language of the target discourse. Cf-rankings in English and Japanese are as follows [12, 13]:

**English Cf-ranking:**
- Subject $\succ$ Direct object $\succ$ Indirect object $\succ$ Other complements $\succ$ Adjuncts

**Japanese Cf-ranking:**
- Topic (wa) $\succ$ Subject (ga) $\succ$ Indirect object (ni) $\succ$ Direct object (o) $\succ$ Others

The entities referred to in $U_i$, ordered by salience, are defined as forward-looking centers, Cf($U_i$). The highest ranked entity in Cf($U_{i-1}$) that is referred to in $U_i$ is defined as the backward-looking center, Cb($U_i$). The highest ranked entity in Cf($U_i$) is defined as the preferred center, Cp($U_i$).

The two rules about referential coherence are:

**Rule 1** (pronominalization): If an entity $e$ in Cf($U_i$) is referred to by a pronoun in $U_{i+1}$, then Cb($U_{i+1}$) must also be referred to by a pronoun in $U_{i+1}$ ($e$ may be Cb($U_{i+1}$)).

**Rule 2** (topic continuity): Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift (See Table 5.1).
5.2. ISSUES ON CENTERING THEORY

Table 5.1: Transition types of Rule 2

<table>
<thead>
<tr>
<th>( \text{Cb}(U_{i+1}) = \text{Cb}(U_i) )</th>
<th>( \text{Cb}(U_{i+1}) \neq \text{Cb}(U_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue</td>
<td>Smooth-Shift</td>
</tr>
<tr>
<td>Retain</td>
<td>Rough-Shift</td>
</tr>
</tbody>
</table>

For instance, the examples in Figure 5.2 can be heuristically formalized by Rules 1 and 2. Centering theory, however, has the following drawbacks.

1. **No general principles.** Without a cooperative cognitive principle, the two rules explain only superficial phenomena about referential co-

Figure 5.2: Referentially coherent and incoherent examples

53
CHAPTER 5. MEANING-GAME-BASED CENTERING MODEL

herence. A standard theory of intentional communication requires a principle of action selection for the sender and receiver. Centering theory does not explain what kind of mechanism is working when $S$ and $R$ cooperatively prefer strong referential coherence. Furthermore, because of the lack of general principles, different researchers have proposed different versions of centering theory [11]. Their variations do not clearly specify the principle behind referential coherence. Without a principle of cooperative preference, we fear that such variations will grow disorderly.

2. **No statistical foundation.** Without corpus-based statistics, the theory’s formalizations (i.e., salience ranking and the two rules) are based on a priori heuristics. For instance, the above Cf-rankings are defined as heuristic rankings among grammatical functions. While [92] proposed an extended Cf-ranking integrated with information status and [14] proposed an extended ranking integrated with contextual information, these rankings are based on surficial observations without sufficient theoretical grounds. Although [11] discussed the parameters settings in centering theory, and [93] proposed centering-based metrics of coherence, their discussions were also based on heuristic ranking. To absorb language-specific differences in superficial expressions, a statistical design is more desirable than a heuristic one. Moreover, such heuristic rules do not give quantitative predictions. To output referentially coherent utterances, automatic discourse compilation systems (e.g., dialogue systems, summarization systems, etc.) are required to measure preference of referential coherence; i.e., how $S$ and $R$ prefer particular candidate mappings between expressions and their referents in the succeeding utterance.
5.3 Meaning Game

5.3.1 General Framework

The meaning game [78, 79] formulates intentional communication in order to capture the core of non-natural meaning on the basis of game-theoretical principles. Let \( C \) be the set of semantic contents and \( M \) be the set of the linguistic messages. The sender \( S \) sends a message \( m \in M \) to convey content \( c_S \in C \). The receiver \( R \) interprets \( m \) as meaning \( c_R \in C \). Therefore, a turn of communication is represented as \( \langle c_S, m, c_R \rangle \in C \times M \times C \). \( c_S = c_R \) is a necessary condition for this turn of communication to be successful.

The utility function \( U_{t_X} \) of player \( X \) would thus be a real-valued function from \( C \times M \times C \) (the set of turns). \( U_{t_S} \) and \( U_{t_R} \) are mutually positive only if the communication is successful. \( \Pr(c_S, m, c_R) \) is the probability that a turn \( \langle c_S, m, c_R \rangle \) occurs (i.e., \( S \) selects \( \langle c_S, m \rangle \) and \( R \) selects \( \langle m, c_R \rangle \)).

The meaning game assumes the following:

1. \( c_S = c_R \), i.e., \( S \) and \( R \)'s communication is always successful, because \( U_{t_S} \) and \( U_{t_R} \) are mutually positive only if the communication is successful.

2. \( S \) and \( R \)'s common belief includes the distribution of \( \Pr(c_S, m, c_R) \).

The expected utility of player \( X \) in the succeeding turn is defined as follows:

\[
\sum_{(c_S, m, c_R) \in C \times M \times C} \Pr(c_S, m, c_R) U_{t_X}(c_S, m, c_R).
\]

\( S \) and \( R \) respectively decide the strategy \( \langle c_S, m, c_R \rangle \) according to their expected utilities. Under the above assumptions, they cooperatively decide the same strategy \( \langle c, m, c \rangle \) for the communication to succeed.

5.3.2 Derivation of Centering Theory

Hasida demonstrated that centering theory can be derived from the meaning game framework[79]. He explained his hypothesis by using the following
CHAPTER 5. MEANING-GAME-BASED CENTERING MODEL

refmap\textsubscript{A}(U\textsubscript{2})  
(Entities in U\textsubscript{1}) (Anaphors in U\textsubscript{2})  
Pr\textsubscript{1} Fred ‘he’ Ut\textsubscript{1}  
\lor  
Pr\textsubscript{2} Max ‘the man’ Ut\textsubscript{2} 

refmap\textsubscript{B}(U\textsubscript{2})  
(Entities in U\textsubscript{1}) (Anaphors in U\textsubscript{2})  
Pr\textsubscript{1} Fred ‘he’ Ut\textsubscript{1}  
\lor  
Pr\textsubscript{2} Max ‘the man’ Ut\textsubscript{2} 

EU(refmap\textsubscript{A}(U\textsubscript{2})) = Pr\textsubscript{1}Ut\textsubscript{1} + Pr\textsubscript{2}Ut\textsubscript{2} > Pr\textsubscript{1}Ut\textsubscript{2} + Pr\textsubscript{2}Ut\textsubscript{1} = EU(refmap\textsubscript{B}(U\textsubscript{2})) 

because (Pr\textsubscript{1} − Pr\textsubscript{2})(Ut\textsubscript{1} − Ut\textsubscript{2}) > 0

Figure 5.3: Derivation of Rules 1 and 2 from the meaning game example:

U\textsubscript{1}: Fred scolded Max.

U\textsubscript{2}: He was angry with the man.

Figure 5.3 represents the meaning game on this example. The preferred interpretation of ‘he’ and ‘the man’ in U\textsubscript{2} are Fred and Max, respectively, rather than the contrary. This preference, which complies with Rules 1 and 2 of centering theory, is accounted for by the meaning game shown as Figure 5.3. Let Pr\textsubscript{1} and Pr\textsubscript{2} be the probabilities of Fred and Max being referred to in U\textsubscript{2}, respectively. We assume that Pr\textsubscript{1} > Pr\textsubscript{2} because Fred is more salient than Max according to the grammatical functions in U\textsubscript{1}. Let Ut\textsubscript{1} and Ut\textsubscript{2} be the respective utilities of ‘he’ and ‘the man’ being used in U\textsubscript{2}. We assume that Ut\textsubscript{1} > Ut\textsubscript{2} because ‘he’ costs less than ‘the man’ to perceptually process. Now we have (Pr\textsubscript{1} − Pr\textsubscript{2})(Ut\textsubscript{1} − Ut\textsubscript{2}) > 0 because Pr\textsubscript{1} > Pr\textsubscript{2} and Ut\textsubscript{1} > Ut\textsubscript{2}. Namely, the combination of mappings on the left-hand side entails a greater joint expected utility.

5.4 Formulating Referential Coherence

5.4.1 Generalized Formulation

Hasida’s derivation of centering theory can be generalized to actual examples in a real discourse structure (Figure 5.4). Hereafter, let w be a target anaphor
5.4. FORMULATING REFERENTIAL COHERENCE

(i.e., a message) in a succeeding utterance unit and \( e \) be an entity (i.e., a semantic content) as a candidate of a referent of the anaphor. Let \( \text{pre}(U_i) \) be \([U_1, U_2, \ldots, U_i]\), the preceding discourse of \( U_i \), when \( S \) and \( R \) predict the succeeding utterance \( U_{i+1} \). To generally formulate Hasida’s derivation, we define two parameters:

**The reference probability** of \( e \) at \( U_i \), represented as \( \text{Pr}(e|\text{pre}(U_i)) \), is defined as the conditional probability of \( e \) being referred to in the succeeding utterance unit \( U_{i+1} \), given the referential features of \( e \) in \( \text{pre}(U_i) \) (see Table 4.1). It represents discourse salience at the moment of \( U_i \) (i.e., degree of joint attention to \( e \) between \( S \) and \( R \)) because a salient entity tends to be continuously referred to.

**The perceptual utility** of \( w \), represented as \( \text{Ut}(w) \), is defined as the inverted measure (i.e., the reduction) of the perceptual cost for use of \( w \) when \( S \) speaks/writes it and \( R \) listens/reads it. The perceptual cost can be defined on the basis of the occurrence probability of \( w \). Frequently used expressions (e.g., pronouns) have higher perceptual utilities than rare nouns because \( S \) and \( R \) are perceptually familiarized with them.

Figure 5.4: Preferences 1a and 1b: Generalization of Rule 1 of centering theory
CHAPTER 5. MEANING-GAME-BASED CENTERING MODEL

Notice that we exclude the cost of reference resolution from the perceptual cost definition because our goal is to make a bottom-up formulation of the cognitive principle behind referential coherence from low-level perceptions (See also Figure 5.1).

We assume the communication between $S$ and $R$ is successful. Where let $\text{refmap}(U_{i+1})$ be the set of reference mappings $\{ (e, w); w \xrightarrow{\text{ref}} e \text{ in } U_{i+1} \}$, the assumption means that $S$ and $R$ can cooperatively select the same candidate of $\text{refmap}(U_{i+1})$. Therefore $\text{EU}(\text{refmap}(U_{i+1}))$, $S$ and $R$’s joint expected utility of $\text{refmap}(U_{i+1})$ can be formulated as follows:

\[
\text{EU}(\text{refmap}(U_{i+1})) = \sum_{w \xrightarrow{\text{ref}} e \text{ in } U_{i+1}} \text{Pr}(e|\text{pre}(U_{i})) \text{Ut}(w)
\]

Table 5.2: Correspondence of centering theory’s account and the meaning game’s account

<table>
<thead>
<tr>
<th>Discourse salience</th>
<th>Centering theory (non-quantified/rule-based)</th>
<th>Meaning game (quantified/corpus-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cf-ranking</td>
<td>Reference probability</td>
<td>Perceptual utility</td>
</tr>
<tr>
<td>(Subject&gt;Object&gt; · · ·)</td>
<td>Pr(e</td>
<td>pre(U_{i}))</td>
</tr>
<tr>
<td>Load reduction</td>
<td>Pronominalization</td>
<td>Expected utility</td>
</tr>
<tr>
<td>(Pronoun / Non-pronoun)</td>
<td></td>
<td>$\text{EU}(\text{refmap}(U_{i+1}))$</td>
</tr>
<tr>
<td>Referential coherence</td>
<td>Transition ranking</td>
<td></td>
</tr>
<tr>
<td>(CONTINUE &gt;RETAIN &gt;SMOOTH-SHIFT &gt;ROUGH-SHIFT)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\text{pre}(U_{i})$: Preceding discourse, $U_{i+1}$: Succeeding utterance unit,

$e$: Entity referred to in $\text{pre}(U_{i})$, $w$: Anaphor in $U_{i+1}$
5.4. FORMULATING REFERENTIAL COHERENCE

$S$ and $R$ cooperatively prefer solutions $\langle w, e \rangle$ that have higher expected utility. This is the principle of expected utility. The above meaning-game-based formulation corresponds to the original centering theory, as Table 5.2 shows. Since the reference probability is a quantification of discourse salience, it is also a quantification of Cf-ranking. Since the perceptual utility is a quantification of the reduction of perceptual load for using referring expressions, it is also a quantification of pronominalization. Since the expected utility is a quantification of the referential coherence, it is also a quantification of transition ranking of Rule 2.

Rules 1 and 2 of centering theory, which represent preference of referential coherence, can be generalized as follows:

**Preference 1a:** In Figure 5.4, (A) is preferred over (B). That is, $w_1$ refers to $e_1$ and $w_2$ refers to $e_2$ when $\Pr(e_1|\text{pre}(U_i)) > \Pr(e_2|\text{pre}(U_i))$ and $\Ut(w_1) > \Ut(w_2)$, given that both $w_1$ and $w_2$ are in $U_{i+1}$.

**Preference 1b:** There is a positive correlation between $\Ut(w)$ and $\Pr(e)$, when $w$ refers to $e$.

**Preference 2:** The higher $\EU(\text{refmap}(U_{i+1}))$ is preferred.

These preferences are derived from the principle of expected utility.

Preference 1a is our general formulation of Hasida’s derivation. Since (A) has greater expected utility than (B) in Figure 5.4, (A) is preferred over (B). It is also a generalization of Rule 1 of centering theory. When only $e_1$ and $e_2$ are referred to from $U_{i+1}$, the condition $\Pr(e_1|\text{pre}(U_i)) > \Pr(e_2|\text{pre}(U_i))$ means $\Cb(U_{i+1}) = e_1$. The condition $\Ut(w_1) > \Ut(w_2)$ means that $w_1$ is a lower-cost expression (e.g., pronominalized) than $w_2$. Therefore, $\Cb(U_{i+1})$, i.e. $e_1$, is pronominalized in (A). This means the coverage of Preference 1a includes that of Rule 1.

Preference 1a can be further generalized to Preference 1b. Preference 1a means that a high-Pr entity tends to be referred to by a high-Ut expression, and a low-Pr entity tends to be referred to by a low-Ut expression. It is generalized to the positive correlation between $\Pr(e|\text{pre}(U_i))$ and $\Ut(w)$, i.e., Preference 1b.
Preference 2 is just the principle of expected utility. We consider that Rule 2 of centering theory is attributed to expected utility. When the condition $\text{Cb}(U_i) = \text{Cb}(U_{i+1})$ in Rule 2 holds, the reference probability of Cb is higher than when it does not hold. In this case, the utility of the expression referring to the Cb also tends to become high because of Preference 1b, so that the expected utility also tends to be high. Similarly, when $\text{Cb}(U_{i+1}) = \text{Cp}(U_{i+1})$ holds, the reference probability of Cb and the utility of the expression referring to Cb are high; Thus, the expected utility also tends to be high. Furthermore, the first condition has a stronger influence than the second, because the first one represents the observed referential coherence between $U_i$ and $U_{i+1}$, whereas the second merely predicts the referential coherence between $U_{i+1}$ and the succeeding $U_{i+2}$. Consequently, Retain has a larger expected utility than Smooth-Shift. Thus, Preference 2 can be regarded as a generalization of Rule 2.

We conjecture that the strengths of the restriction of these preferences can be measured by the expected utility $EU(\text{refmap}(U_{i+1}))$.

### 5.4.2 Statistical Definition of Parameters

Here, we describe a statistical calculation of the parameters, $\Pr$ and $\Ut$, on the basis of a corpus.

**Calculation of $\Pr(e|\text{pre}(U_i))$**

As described in Chapter 4, the discourse salience of an entity $e$ at the target utterance $U_i$ depends on the linguistic pattern of the expressions referring to $e$ from the preceding context $\text{pre}(U_i)$.

**Calculation of $\Ut(w)$**

The perceptual utility $\Ut(w)$ is defined as a reduction of the perceptual cost when $S$ speaks/writes it and $R$ hears/reads it. Let $p(w)$ be the occurrence probability of the anaphor $w$. Thus, the perceptual cost of $w$ can be defined
as \( I(w) = -\log p(w) \) because \( S \) and \( R \) are perceptually familiar with the frequent anaphors. Moreover, once \( S \) and \( R \) perceptually get used to \( w \), it becomes more frequently used because its use decreases the perceptual load. The definition conforms to Weber-Fechner’s law; i.e., perceptual intensity is proportional to the logarithm of the fundamental stimulus value [94]. Let \( U_{t_0} \) be the constant utility of successful communication; i.e., \( S \) and \( R \) cooperatively regard \( e \) as the referent of \( w \). Hereafter, we call \( U_{t_0} \) the basic value of utility. The perceptual utility of \( w \) can be calculated as follows:

\[
U_{t}(w) = U_{t_0} - I(w) = U_{t_0} + \log p(w)
\]

We should determine \( U_{t_0} \), the basic value of utility, that satisfies \( U_{t_0} \geq \max I(w) \) in order to ensure \( U_{t}(w) \geq 0 \) when the communication is successful. The value of \( U_{t_0} \) can be determined by maximizing the consistency of Preference 2 of our meaning-game-based account with Rule 2 of centering theory for the following reason: Preferences 1a and 1b do not depend on \( U_{t_0} \). The restrictions on Preferences 1a and 1b depend on only \( U_{t}(w_1) - U_{t}(w_2) = I(w_2) - I(w_1) \) in Figure 5.4. On the other hand, the consistency of Preference 2 with Rule 2 depends on \( U_{t_0} \). The consistency is kept only if the value of \( U_{t_0} \) satisfies \( U_{t_0} \geq \max I(w) \). Thus, we determine \( U_{t_0} \) according to the consistency of Preference 2 with Rule 2 in Chapter 6.

5.5 Conclusion

We proposed the meaning-game-based centering model, which gives a game-theoretical and statistical formulation of referential coherence in order to clarify the behavioral principle behind referential coherence in various languages. To formulate referential coherence, we employed Hasida’s hypothesis that centering theory can be derived from the meaning game framework. To absorb differences due to language-specific expressions, we statistically for-
mulated two parameters (i.e., reference probability and perceptual utility) by using a corpus of the target language. These formulations enabled us to empirically verify the language universality of the meaning-game-based centering model in Chapter 6.
Chapter 6

Empirical Evaluation

This chapter empirically evaluates our formulations of (1) discourse salience, (2) referential coherence. Furthermore, the following unexplained points are clarified: (1) the characteristics of recency effect in discourse and (2) the behavioral principle behind referential coherence.

6.1 Corpus Specification

Evaluating and verifying our model requires corpora, which is annotated with syntactic and anaphoric information. In this thesis, we used the following corpora:

- **CSJ**: 4 spontaneous dialogues from the Corpus of Spontaneous Japanese (CSJ)[88], which contain
  - 1,780 utterance units (IPUs; inter-pause units),
  - 6.92 morphemes per utterance unit, and
  - 1,180 anaphora relations annotated manually.

- **Mainichi**: 3,000 newspaper articles from Mainichi-Shinbun in 1994 (GSK2004-A[87]), which contain
CHAPTER 6. EMPIRICAL EVALUATION

- 63,221 utterance units (predicate clauses), 37,340 sentences,
- 10.79 morphemes per an utterance unit, and
- 86,541 anaphora relations annotated manually.

- WSJ: 2,412 newspaper articles from Wall Street Journal (GSK2004-H[87]), which contain
  - 135,278 utterance units (predicate clauses), 46,816 sentences,
  - 7.64 words per an utterance unit, and
  - 95,677 anaphora relations annotated manually.

These corpora are annotated with syntactic and anaphoric information according to GDA (Global Document Annotation) tags[81, 82]. Although the corpora includes both direct and indirect anaphora, we deals with only direct anaphora by omitting indirect anaphora. The syntactic GDA tags in CSJ are annotated with automatic dependency analysis by CaboCha[85, 86] as mentioned in Chapter 3. We selected the four dialogues (D03F0001, D03M0013, D03F0040, and D03M0048) from 58 spontaneous dialogues in the CSJ corpus due to the length of conversation and the balance of gender. Those in Mainichi are converted from RWC-DB-TEXT-95-2, RWC Text Database[95], which is annotated manually. Those in WSJ are converted from Penn TreeBank[96]. The anaphoric GDA tags in all corpora are annotated manually, which is instantiated as follows:

```
<su syn="f">
  <adp opr="topic.fit.agt">
    <persnamep id="KonoYohei">河野洋平</persnamep> 氏は</adp>
  <adp opr="cnt"><q>新制度の下で戦う</q>と</adp>
  <v>強調</v>。 </su>
```

(“We will fight under the new system”, Mr. KONO Yohei emphasized.)
6.1. CORPUS SPECIFICATION

The subject of the verb “示し (indicate)” is omitted, i.e. zero-pronominalized, at element ②. The attribute agt="KonoYohei" at ② represents the zero anaphora. The antecedent element ③ and the verb element ④ are respectively annotated with id and agt attributes. agt is an attribute type to represent an agent of action.

“We have no useful information on whether users are at risk,”</q>

said</v>

James A. Talcott</persname>

of Boston’s Dana-Farber Cancer Institute</adp>

Dr. Talcott</persname>

led</v>

a team of researchers</n>

from the National Cancer Institute</adp>
The attribute \texttt{eq="DrTalcott"} at ③ represents an anaphor referring to ③.

### 6.2 Evaluating Reference Probability as Discourse Salience

We statistically formulated the scale of discourse salience (i.e., \( \Pr(e|\text{pre}(U_i)) \)) and the evaluation measure of the salience calculation (i.e., \( \text{evalSal}(m) \)) in Chapter 4. In this section, we empirically evaluate calculation methods of the discourse salience by using corpora. To compare multiple kinds of discourse, we used \textit{CSJ}, the corpus of Japanese spontaneous dialogue and \textit{Mainichi}, the corpus of Japanese newspaper.

![Figure 6.1: Relationship of reference probability and recency (CSJ)](image)

Figure 6.1: Relationship of reference probability and recency (CSJ)
6.2. EVALUATING REFERENCE PROBABILITY AS DISCOURSE SALIENCE

6.2.1 Clarifying Characteristics of Recency Effect in Discourse Context

As preparation, we clarify the characteristics of recency effect in order to investigate the optimum method of calculating $\Pr(e|\text{pre}(U_i))$. To incorporate the recency effect, we employ window function which represents the optimum decay curve of the recency effect. This is based on the assumption that the optimum window function represents the decay curve of the recency effect. In other words, though optimizing window function, we can clarify what shape the decay curve of the recency effect in discourse context has.

Firstly, to presume the decay curve of the recency effect, we investigated the relationship between $\Pr(e|\text{pre}(U_i))$, the reference probability, and the recency of the referring expression $w_{\text{ref}} \rightarrow e$. Figure 6.1 shows the relationship between $\Pr(e|\text{pre}(U_i))$ and $\text{dist}(\text{latest}(e, U_i), U_{i+1})$, the recency of the latest $w$ referring to $e$ in $\text{pre}(U_i)$, where $\text{latest}(e, U_i)$ denotes the latest expression $w_{\text{ref}} \rightarrow e$ in $\text{pre}(U_i)$. The shape of the graph can be perceived as being similar to the inverse function. This leads the presumption that the decay curve of the recency effect in discourse is also similar to the inverse function.

Now let us verify the presumption through investigating the optimum window function. The candidate window functions are as follows:

- **Rectangular window** ($k$ is a variable parameter)

  $$W(\text{dist}) = \begin{cases} 
  1 & \text{dist} \leq k \\
  0 & \text{otherwise}
  \end{cases}$$

- **Gaussian window** ($\sigma$ is a variable parameter)

  $$W(\text{dist}) = \exp\left(-\frac{\text{dist}^2}{\sigma^2}\right)$$

- **Exponential window** ($T$ is a variable parameter)

  $$W(\text{dist}) = \exp\left(-\frac{\text{dist}}{T}\right)$$

67
CHAPTER 6. EMPIRICAL EVALUATION

- **Inverse window** \((d)\) is a variable parameter

\[
W(\text{dist}) = \frac{1}{\text{dist}^d}
\]

Notice that an optimized window function is dedicated for each feature (e.g., grammatical role and part-of-speech). That is to say, the variable parameters of the window functions have to been optimized for each feature.

The window functions can be optimized by maximizing \(\text{evalSal}(m)\), the evaluation scale defined in Chapter 4. To optimize the window functions, we regard a sum of windowed features in \(\text{pre}(U_i)\) as an exponential salience value \(\text{sal}_m(e|\text{pre}(U_i))\). More concretely, where \(W(\text{dist})\) is a candidate window function and \(\text{feature}(w)\) is a target feature of \(w\), we regard

\[
\sum_{w \text{ ref.: in } \text{pre}(U_i)} W(\text{dist})\text{feature}(w)
\]

<table>
<thead>
<tr>
<th>feature</th>
<th>window function</th>
<th>optimal parameter</th>
<th>(\text{evalSal}(m))</th>
</tr>
</thead>
<tbody>
<tr>
<td>occurrence</td>
<td>Rectangular</td>
<td>(k = 9)</td>
<td>0.1048</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>(\sigma = 3.66)</td>
<td>0.1667</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>(T = 2.51)</td>
<td>0.1727</td>
</tr>
<tr>
<td></td>
<td>Inverse</td>
<td>(d = 1.35)</td>
<td><strong>0.1731</strong></td>
</tr>
<tr>
<td>gram. func.</td>
<td>Rectangular</td>
<td>(k = 9)</td>
<td>0.1195</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>(\sigma = 5.01)</td>
<td>0.1928</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>(T = 3.39)</td>
<td>0.2013</td>
</tr>
<tr>
<td></td>
<td>Inverse</td>
<td>(d = 1.14)</td>
<td><strong>0.2063</strong></td>
</tr>
<tr>
<td>part-of-speech</td>
<td>Rectangular</td>
<td>(k = 9)</td>
<td>0.1200</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>(\sigma = 2.52)</td>
<td>0.2226</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>(T = 2.19)</td>
<td>0.2308</td>
</tr>
<tr>
<td></td>
<td>Inverse</td>
<td>(d = 1.24)</td>
<td><strong>0.2390</strong></td>
</tr>
</tbody>
</table>
6.2. EVALUATING REFERENCE PROBABILITY AS DISCOURSE SALIENCE

Table 6.2: Comparing window functions optimized for each feature (Mainichi)

<table>
<thead>
<tr>
<th>feature</th>
<th>window function</th>
<th>optimal parameter</th>
<th>evalSal(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>occurrence</td>
<td>Rectangular</td>
<td>( k = 1 )</td>
<td>0.3013</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>( \sigma = 0.99 )</td>
<td>0.3467</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>( T = 0.33 )</td>
<td>0.3467</td>
</tr>
<tr>
<td></td>
<td>Inverse</td>
<td>( d = 4.27 )</td>
<td>\textbf{0.3468}</td>
</tr>
<tr>
<td>gram. func.</td>
<td>Rectangular</td>
<td>( k = 1 )</td>
<td>0.2991</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>( \sigma = 1.14 )</td>
<td>0.3682</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>( T = 0.45 )</td>
<td>0.3685</td>
</tr>
<tr>
<td></td>
<td>Inverse</td>
<td>( d = 3.01 )</td>
<td>\textbf{0.3696}</td>
</tr>
<tr>
<td>part-of-speech</td>
<td>Rectangular</td>
<td>( k = 1 )</td>
<td>0.2985</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>( \sigma = 1.04 )</td>
<td>0.3537</td>
</tr>
<tr>
<td></td>
<td>Exponential</td>
<td>( T = 0.37 )</td>
<td>0.3537</td>
</tr>
<tr>
<td></td>
<td>Inverse</td>
<td>( d = 3.80 )</td>
<td>\textbf{0.3540}</td>
</tr>
</tbody>
</table>

as \( \text{sal}_m(e|\text{pre}(U_i)) \). As a result, the evaluation scale for optimizing a window function is deformed as

\[
\text{evalSal}(m) = \text{cor}\left( \sum_{w \text{ref} \in \text{pre}(U_i)} W(\text{dist}\text{feature}(w), \text{isRef}(e, U_{i+1})) \right).
\]

Thus, we optimized the variable parameters of each window function by maximizing \( \text{evalSal}(m) \) in order to compare the window functions. The experimental results by using \textit{CSJ} and \textit{Mainichi} are respectively shown in Tables 6.1 and 6.2. By these results, we can find that the inverse windows are most optimal for each corpus and each feature.

Furthermore, we focused on the extent to which each feature influences in the succeeding discourse. We can investigate it by observing the curve gentleness of the window function optimized for each feature. The curve
gentleness corresponds to the distance to decay to \( \frac{1}{20} \) shown in Figure 6.2.\(^1\) We observed them as shown in Tables 6.3 and 6.4. The tables show that the rank of the extent of feature influences between features is as follows:

grammatical function \( > \) part-of-speech \( > \) occurrence.

Moreover, we can interpret the results from a viewpoint of the difference between spontaneous conversation and newspaper. The extent of feature influence in \( CSJ \) is wider than that in \( Mainichi \). We consider this difference being caused by the restriction of writing newspaper, that is, the newspaper writer has to shorten the article due to restriction of space. On the other hand, conversation participants in \( CSJ \) are not under such restriction. Therefore, the contextual extent of feature influence in \( CSJ \) became wider than that in \( Mainichi \).

\(^1\)We consider \( \frac{1}{20} \) the enough weak weight to represent decaying influence.
6.2. EVALUATING REFERENCE PROBABILITY AS DISCOURSE SALIENCE

Table 6.3: Curve gentleness of optimal inverse window (<i>CSJ</i>)

<table>
<thead>
<tr>
<th>feature</th>
<th>optimal $k$</th>
<th>utterance distance (IPUs) to decay to $\frac{1}{20}$</th>
<th>morpheme distance to decay to $\frac{1}{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>occurrence</td>
<td>1.35</td>
<td>9.20</td>
<td>63.7</td>
</tr>
<tr>
<td>gram. func.</td>
<td>1.14</td>
<td>13.8</td>
<td>95.8</td>
</tr>
<tr>
<td>part-of-speech</td>
<td>1.24</td>
<td>11.2</td>
<td>77.5</td>
</tr>
</tbody>
</table>

Table 6.4: Curve gentleness of optimal inverse window (<i>Mainichi</i>)

<table>
<thead>
<tr>
<th>feature</th>
<th>optimal $k$</th>
<th>utterance distance (clauses) to decay to $\frac{1}{20}$</th>
<th>morpheme distance to decay to $\frac{1}{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>occurrence</td>
<td>4.27</td>
<td>2.01</td>
<td>21.7</td>
</tr>
<tr>
<td>gram. func.</td>
<td>3.01</td>
<td>2.71</td>
<td>29.2</td>
</tr>
<tr>
<td>part-of-speech</td>
<td>3.80</td>
<td>2.20</td>
<td>23.7</td>
</tr>
</tbody>
</table>

6.2.2 Optimizing and Evaluating Calculation

We optimize the calculation method of the reference probability through maximizing the evaluation scale proposed in Chapter 4. Furthermore, we compare the optimized calculation method of $\text{Pr}(e | \text{pre}(U_i))$ with a naive term-weighting (i.e., TF windowed by optimized rectangular window).

As mentioned above, the inverse window functions were optimal for representing the recency effect in discourse. The parameter $k$ of the inverse window functions were optimized for each feature (Tables 6.3 and 6.4).

Let us select the optimal feature set by maximizing evalSal($m$). The candidate features are shown in Tables 6.5. Firstly, we investigated single influence of each feature. Concretely, we calculated evalSal($m$) for each feature, given that $\text{sal}_m(e | \text{pre}(U_i))$ is respectively defined by a target feature. Tables 6.6 and 6.7 show the results for each feature, where “latest” denotes
the latest expression referring to $e$ in $\text{pre}(U_i)$. The results show that $\text{dist}$ is the most effective feature. That is to say, the recency effect is the most

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{dist}(w, U_{i+1})$</td>
<td>$(i + 1) - j$ when $w \xrightarrow{\text{ref}} e$ is in the utterance $U_j$</td>
</tr>
<tr>
<td>$\text{freq}((e, U_i))$</td>
<td>$\frac{1}{i}(# w \xrightarrow{\text{ref}} e$ in $\text{pre}(U_i))$</td>
</tr>
<tr>
<td>$\text{gram}(w)$</td>
<td>Function word depended by $w \xrightarrow{\text{ref}} e$</td>
</tr>
<tr>
<td>$\text{pos}(w)$</td>
<td>Part-of-speech of $w \xrightarrow{\text{ref}} e$</td>
</tr>
<tr>
<td>$\text{title}(e)$</td>
<td>Whether $e$ is referred to in the title</td>
</tr>
</tbody>
</table>

Table 6.5: Description of the candidate features

<table>
<thead>
<tr>
<th>feature</th>
<th>gram</th>
<th>pos</th>
<th>$\text{W}(\text{dist})$</th>
<th>Regression</th>
<th>evalSal($m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alone $\text{gram}$ (avgPr(gram(latest)))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0475</td>
</tr>
<tr>
<td>Alone $\text{pos}$ (avgPr(pos(latest)))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0526</td>
</tr>
<tr>
<td>Alone $\text{dist}$ $\frac{1}{i} \text{dist(latest, } U_{i+1})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1519</td>
</tr>
<tr>
<td>Alone $\text{freq}$ $\frac{1}{i} \sum_{w \xrightarrow{\text{ref}} e \text{ in } \text{pre}(U_i)} 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0437</td>
</tr>
</tbody>
</table>

Table 6.6: Comparing features ($CSJ$)
6.2. EVALUATING REFERENCE PROBABILITY AS DISCOURSE SALIENCE

Table 6.7: Comparing features (Mainichi)

<table>
<thead>
<tr>
<th>feature</th>
<th>gram</th>
<th>pos</th>
<th>W(dist)</th>
<th>∑ [ w \rightarrow e \in \text{pre}(U_i) ]</th>
<th>Ref de (Primacy)</th>
<th>Regression</th>
<th>evalSal(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alone gram</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1171</td>
</tr>
<tr>
<td>Alone pos</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0817</td>
</tr>
<tr>
<td>Alone dist</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2831</td>
</tr>
<tr>
<td>Alone freq</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2519</td>
</tr>
<tr>
<td>Alone title</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td>0.0644</td>
</tr>
<tr>
<td>(whether ∃ w \rightarrow e \in \text{title})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8: Evaluation of candidate feature sets (CSJ)

<table>
<thead>
<tr>
<th>feature set</th>
<th>gram</th>
<th>pos</th>
<th>W(dist)</th>
<th>∑ [ w \rightarrow e \in \text{pre}(U_i) ]</th>
<th>Regression</th>
<th>evalSal(m)</th>
<th>log likelihood (avg. of 10-folds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gramFeat</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3458</td>
<td>-6982.2</td>
</tr>
<tr>
<td>posFeat</td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td>0.3591</td>
<td>-6898.5</td>
</tr>
<tr>
<td>gramFeat posFeat</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td>0.3652</td>
<td>-6857.1</td>
</tr>
</tbody>
</table>

important factor influencing to discourse salience.

To select the optimal feature set, we investigated evalSal(m) for each candidate feature set. To measure evalSal(m), we employed 10-fold cross
Table 6.9: Evaluation of candidate feature sets (*Mainichi*)

<table>
<thead>
<tr>
<th>feature set</th>
<th>gram (Gram. func.)</th>
<th>pos (Named entity)</th>
<th>title (Primacy)</th>
<th>W(dset) (Recency)</th>
<th>W(θ) (Frequency)</th>
<th>Regression (Anomaly)</th>
<th>evalSal(m) (avg. of 10-folds)</th>
<th>Log Likelihood (avg. of 10-folds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gramFeat</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3687</td>
<td>-49520.9</td>
</tr>
<tr>
<td>posFeat</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3437</td>
<td>-50000.8</td>
</tr>
<tr>
<td><strong>gramFeat posFeat</strong></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>0.3680</strong></td>
<td>-49481.0</td>
</tr>
<tr>
<td>title</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0860</td>
<td>-57523.4</td>
</tr>
<tr>
<td>gramFeat title</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3638</td>
<td>-48917.8</td>
</tr>
<tr>
<td>posFeat title</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3395</td>
<td>-49453.0</td>
</tr>
<tr>
<td>gramFeat posFeat title</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3631</td>
<td>-48899.6</td>
</tr>
</tbody>
</table>

Validation. That is to say, we measure evalSal(m) by using a test-set corpus which is a fold in 10 folds. This observation resulted in Tables 6.8 and 6.9. The bold faced rows are the optimal feature set on each corpus. In *CSJ*, evalSal(m) = 0.3652 in case of the optimal feature set. In *Mainichi*, evalSal(m) = 0.3680 in case of the optimal feature set. In both corpora, the optimized calculation method was as shown in Figure 6.3.

Here, let us compare these evaluation measures with a naive term-weighting, evalSal(m) of “simple occurrences” in Tables 6.1 and 6.2 can be regarded as the evaluation of TF (term frequency), which is windowed by the optimal rectangular window function. The results of comparison are shown in Figures 6.4 and 6.5. The evaluation measures evalSal(m) for the naive term-weighting
6.2. EVALUATING REFERENCE PROBABILITY AS DISCOURSE SALIENCE

![Figure 6.3: The optimized calculation method](image)

The evaluation measures for our proposed method (0.3652 in CSJ and 0.3680 in Mainichi) were greater than those for the naive term-weighting. These results mean that our method can more effectively predict that a target entity \( e \) become referred to in the

![Figure 6.4: Comparing \( \Pr(e|\text{pre}(U_i)) \) with naive term-weighing schemes (CSJ)](image)
succeeding $U_{i+1}$ than the naive term-weighting schemes. The effectiveness of our method in \textit{CSJ} was more significant than that in \textit{Mainichi}. This indicates that handling spoken language needs the integrating features (especially the recency effect) more than handling written language does.

### 6.3 Verifying Meaning-Game-based Centering Model

We empirically verified the hypothesis that centering theory can be reduced to the principle of the expected utility in the meaning game. To verify the language universality, we used \textit{Mainichi}, the corpus of Japanese newspaper and \textit{WSJ}, the corpus of English newspaper.

Notice that we regard one predicative or tensed clause as one utterance in the case of complex sentences [97].
6.3.1 Theoretic Verification of Parameter Definitions

Here, we verify the definitions of the reference probability $\text{Pr}(e|\text{pre}(U_i))$ and perceptual utility $U_t(w)$, from a theoretic viewpoint based on centering theory. The reference probability $\text{Pr}(e|\text{pre}(U_i))$ represents how strong $S$ and $R$’s joint attention is to the target entity. The perceptual utility $U_t(w)$ is a parameter to measure the reduction of perceptual load (i.e., how familiar $S$ and $R$ are with the expression referring to the entity). We empirically verified these parameters.

Verification of Reference Probability

The calculation of the reference probability (see also Figure 6.3) required the following preparation.

Firstly, we needed to assign a real value $\text{avgPr}(\text{gram})$ to each grammatical function. We assigned an average $\text{Pr}$ to each grammatical function, which was

<table>
<thead>
<tr>
<th>grammatical function</th>
<th># samples</th>
<th># successive references</th>
<th>avgPr(gram)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic ($wa$)</td>
<td>35,329</td>
<td>1,908</td>
<td>$5.40 \times 10^{-2}$</td>
</tr>
<tr>
<td>Subject ($ga$)</td>
<td>38,450</td>
<td>1,107</td>
<td>$2.88 \times 10^{-2}$</td>
</tr>
<tr>
<td>(no)</td>
<td>88,695</td>
<td>1,755</td>
<td>$1.98 \times 10^{-2}$</td>
</tr>
<tr>
<td>Object ($o$)</td>
<td>50,217</td>
<td>898</td>
<td>$1.79 \times 10^{-2}$</td>
</tr>
<tr>
<td>Indirect object ($ni$)</td>
<td>46,058</td>
<td>569</td>
<td>$1.24 \times 10^{-2}$</td>
</tr>
<tr>
<td>(mo)</td>
<td>8,710</td>
<td>105</td>
<td>$1.21 \times 10^{-2}$</td>
</tr>
<tr>
<td>(de)</td>
<td>24,142</td>
<td>267</td>
<td>$1.11 \times 10^{-2}$</td>
</tr>
<tr>
<td>(kara)</td>
<td>7,963</td>
<td>76</td>
<td>$9.54 \times 10^{-3}$</td>
</tr>
<tr>
<td>(to)</td>
<td>19,383</td>
<td>129</td>
<td>$6.66 \times 10^{-3}$</td>
</tr>
<tr>
<td>other postpositions</td>
<td>512,006</td>
<td>8,027</td>
<td>$1.57 \times 10^{-2}$</td>
</tr>
<tr>
<td>no gram. func.</td>
<td>153,197</td>
<td>1,315</td>
<td>$8.58 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
Table 6.11: Average reference probability for each grammatical function (WSJ)

<table>
<thead>
<tr>
<th>grammatical function</th>
<th># samples</th>
<th># successive references</th>
<th>avgPr (gram)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject (by)</td>
<td>76,147</td>
<td>16,441</td>
<td>2.16×10⁻¹</td>
</tr>
<tr>
<td>Indirect object (with)</td>
<td>1,569</td>
<td>184</td>
<td>1.17×10⁻¹</td>
</tr>
<tr>
<td>(of)</td>
<td>23,798</td>
<td>2,145</td>
<td>9.01×10⁻²</td>
</tr>
<tr>
<td>(from)</td>
<td>4,005</td>
<td>350</td>
<td>8.74×10⁻²</td>
</tr>
<tr>
<td>Object (to)</td>
<td>42,578</td>
<td>3,703</td>
<td>8.70×10⁻²</td>
</tr>
<tr>
<td>(for)</td>
<td>8,449</td>
<td>661</td>
<td>7.82×10⁻²</td>
</tr>
<tr>
<td>(on)</td>
<td>7,759</td>
<td>601</td>
<td>7.75×10⁻²</td>
</tr>
<tr>
<td>(at)</td>
<td>5,140</td>
<td>229</td>
<td>5.82×10⁻²</td>
</tr>
<tr>
<td>Complement (in)</td>
<td>4,043</td>
<td>233</td>
<td>5.76×10⁻²</td>
</tr>
<tr>
<td>other prepositions</td>
<td>183,710</td>
<td>6,848</td>
<td>3.73×10⁻²</td>
</tr>
<tr>
<td>no gram func.</td>
<td>34,105</td>
<td>3,286</td>
<td>9.10×10⁻²</td>
</tr>
</tbody>
</table>

calculated by counting samples in the Mainichi and WSJ corpora. Tables 6.10 and 6.11 show the calculated gram values from both corpora. The results for both language corpora show the consistency between the reference probability and salience ranking (i.e., Cf-ranking) of centering theory in spite of language-specific differences in grammatical functions.

Figures 6.6 and 6.7 show measured reference probabilities by logistic regression with these regression weights. Notice that the gram value is fixed in order to draw the three-dimensional graphs. The figures also show the consistency between the statistically measured value of reference probability and the heuristic knowledge.

That is, these results show the empirical validity of our statistical formu-
6.3. VERIFYING MEANING-GAME-BASED CENTERING MODEL

When gram = Pr(subject) = 0.0288

Figure 6.6: Reference probability calculated from dist, chain (Mainichi)

When gram = Pr(subject) = 0.0171

Figure 6.7: Reference probability calculated from dist, chain (WSJ)

ulation of the reference probability as a measure of discourse salience in both language corpora. This also indicates that our statistical formulation can absorb the difference between language-specific expressions.

79
Verification of Perceptual Utility

The calculation of the perceptual utility requires to determine the value of $Ut_0$ as a preparation. According to the consistency between Preference 2 of the meaning-game-based account and Rule 2, we empirically determined $Ut_0$ to be 15.1 in the Mainichi corpus and $Ut_0$ to be 12.6 in the WSJ corpus. Figure 6.8 shows the empirical basis of the determination of $Ut_0$. The consistencies of both corpora represented as Spearman’s rank correlation coefficients are maximized at those points. The figure also shows that the consistencies are kept within the range of $Ut_0 \geq \max I(w)$; i.e., $Ut_0 \geq 11.06$ in Mainichi and $Ut_0 \geq 11.82$ in WSJ.

We calculated the perceptual utility of each referring expression in both corpora. Tables 6.12 and 6.13 show the calculated perceptual utilities. Both the zero pronoun in Japanese and empty category in English are kinds of ellipsis. Both corpora had the following ranking of perceptual utilities:

ellipsis > pronoun > other noun
6.3. VERIFYING MEANING-GAME-BASED CENTERING MODEL

Table 6.12: Perceptual utility for each referring expression (Mainichi)

<table>
<thead>
<tr>
<th>referring expression</th>
<th>occurrence probability ( p(w) )</th>
<th>perceptual cost ( I(w) )</th>
<th>perceptual utility ( Ut(w) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(zero pronoun)</td>
<td>( 2.940 \times 10^{-1} )</td>
<td>1.224</td>
<td>13.88</td>
</tr>
<tr>
<td>watashi (I)</td>
<td>( 5.129 \times 10^{-3} )</td>
<td>5.273</td>
<td>9.827</td>
</tr>
</tbody>
</table>
sono (that)           | \( 3.965 \times 10^{-3} \)         | 5.530                     | 9.570                         |
kore (this)           | \( 2.973 \times 10^{-3} \)         | 5.818                     | 9.282                         |
kono (this)           | \( 1.888 \times 10^{-3} \)         | 6.272                     | 8.828                         |
Nihon (Japan)         | \( 1.809 \times 10^{-3} \)         | 6.315                     | 8.785                         |
mono (thing)          | \( 1.809 \times 10^{-3} \)         | 6.315                     | 8.785                         |
sore (it)             | \( 1.699 \times 10^{-3} \)         | 6.378                     | 8.722                         |
daitoryo (president)  | \( 1.479 \times 10^{-3} \)         | 6.516                     | 8.584                         |
|...                  |                                   | ::                      | ::                           |
|type of \( w \)      | avg. \( p(w) \)                  | avg. \( I(w) \)          | avg. \( Ut(w) \)             |
|zero pronoun         | \( 2.940 \times 10^{-1} \)         | 1.224                     | 13.88                         |
|pronoun              | \( 2.403 \times 10^{-3} \)         | 6.031                     | 9.069                         |
|other noun           | \( 2.271 \times 10^{-4} \)         | 8.390                     | 6.710                         |

Provided that \( Ut_0 = 15.1 \)

This is consistent with the heuristic knowledge about the perceptual load for using referring expressions. Therefore, the results show the empirical validity of the statistical formulation of the perceptual utility in both language corpora. This also indicates that our statistical formulation can absorb the difference of language-specific expressions.

6.3.2 Verification of Preference 1a

We empirically verified the language universality of Preference 1a by using the Mainichi and WSJ corpora. Moreover, we also verified our conjecture that the strength of the restriction of Preference 1a can be measured by the expected utility \( EU(\text{refmap}(U_{i+1})) \).
Table 6.13: Perceptual utility for each referring expression (WSJ)

<table>
<thead>
<tr>
<th>referring expression</th>
<th>occurrence probability</th>
<th>perceptual cost</th>
<th>perceptual utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty category</td>
<td>2.547 × 10⁻¹</td>
<td>1.368</td>
<td>11.23</td>
</tr>
<tr>
<td>it</td>
<td>4.232 × 10⁻²</td>
<td>3.162</td>
<td>9.438</td>
</tr>
<tr>
<td>he</td>
<td>3.049 × 10⁻²</td>
<td>3.490</td>
<td>9.110</td>
</tr>
<tr>
<td>they</td>
<td>1.850 × 10⁻²</td>
<td>3.990</td>
<td>8.610</td>
</tr>
<tr>
<td>company</td>
<td>1.652 × 10⁻²</td>
<td>4.103</td>
<td>8.497</td>
</tr>
<tr>
<td>we</td>
<td>1.112 × 10⁻²</td>
<td>4.499</td>
<td>8.101</td>
</tr>
<tr>
<td>I</td>
<td>1.020 × 10⁻²</td>
<td>4.585</td>
<td>8.015</td>
</tr>
<tr>
<td>U.S.</td>
<td>8.342 × 10⁻³</td>
<td>4.786</td>
<td>7.814</td>
</tr>
<tr>
<td>you</td>
<td>6.357 × 10⁻³</td>
<td>5.058</td>
<td>7.742</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>type of w</th>
<th>avg. p(w)</th>
<th>avg. I(w)</th>
<th>avg. Ut(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty category</td>
<td>2.457 × 10⁻¹</td>
<td>1.368</td>
<td>11.23</td>
</tr>
<tr>
<td>pronoun</td>
<td>3.257 × 10⁻²</td>
<td>3.836</td>
<td>8.764</td>
</tr>
<tr>
<td>other noun</td>
<td>1.317 × 10⁻³</td>
<td>6.632</td>
<td>5.968</td>
</tr>
</tbody>
</table>

Provided that $Ut_0 = 12.6$

If the succeeding utterance $U_{i+1}$ includes a pair anaphors $w_1$ and $w_2$, the pair has two possible mappings: (A) and (B) in Figure 5.4. For the pairs that have a large difference $(EU(\text{refmap}_A(U_{i+1})) - EU(\text{refmap}_B(U_{i+1})))$, $S$ and $R$ should strongly prefer (A) over (B). Consequently, the ratio of positive samples (which comply with the preference) should reach almost 100%. On the other hand, for pairs that have a small $EU(\text{refmap}_A(U_{i+1})) - EU(\text{refmap}_B(U_{i+1}))$, $S$ and $R$ should weakly prefer (A). In this case, $S$ and $R$ can select (B) a little less frequently than (A). Hence, the ratio of the positive samples should be a little greater than 50%.

Figures 6.9 and 6.10 show the results on both corpora. For $EU_A - EU_B \geq 3$, the ratios of the positive pairs to all pairs, were 0.825 in the Mainichi
corpus and 0.822 in the \textit{WSJ} corpus. For $\text{EU}_A - \text{EU}_B < 0.5$, the ratios of the positive pairs were 0.564 in the \textit{Mainichi} corpus and 0.529 in the \textit{WSJ} corpus, that is, the greater difference ($\text{EU}_A - \text{EU}_B$), the stronger preference. This proves our prediction.

Moreover, Spearman’s rank correlation coefficients between $\text{EU}_A - \text{EU}_B$ and the ratio of positive samples were 0.833 in the \textit{Mainichi} corpus and 0.981 in the \textit{WSJ} corpus. These values indicated that the restriction of Preference 1a is strongly associated with $\text{EU}_A - \text{EU}_B$. This empirically proves our conjecture that the strength of the restriction of Preference 1a can be measured by $\text{EU}(\text{refmap}(U_{i+1}))$ in both corpora. Therefore, the verification strongly proves the hypothesis that Preference 1a (and Rule 1 of centering

![Graph showing the relationship between EU Difference and paired ratio]

Figure 6.9: Preference 1: Ratio of positive pairs is proportional to $\text{EU}(\text{refmap}_A(U_{i+1})) - \text{EU}(\text{refmap}_B(U_{i+1}))$ (\textit{Mainichi}).
CHAPTER 6. EMPIRICAL EVALUATION

Figure 6.10: Preference 1: Ratio of positive pairs is proportional to EU(refmap\textsubscript{A}(U_{i+1})) − EU(refmap\textsubscript{B}(U_{i+1})) (WSJ).

theory) can be reduced to the principle of expected utility in both Japanese and English.

6.3.3 Verification of Preference 1b

We verified the language universality of Preference 1b by using the \textit{Mainichi} and \textit{WSJ} corpora.

Preference 1b, the positive correlation between \(\text{Pr}(e|U_{i+1})\) and \(\text{Ut}(w)\), is a further generalization of Preference 1a. It can be interpreted as the tendency that the more salient \(e\) is, the less costly \(w\) referring to \(e\) will be.

Before the verification, we calculated Spearman’s rank correlation coefficients for the following cases:

- \(\rho_A\): Correlation in (A) samples; i.e., positive samples complying with
6.3. VERIFYING MEANING-GAME-BASED CENTERING MODEL

Table 6.14: Preference 1b: Spearman’s rank correlation between Pr(e|pre( Ui )) and Ut(w) (Mainichi)

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>95% confidential interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρA</td>
<td>+0.540</td>
<td>[+0.512, +0.567]</td>
</tr>
<tr>
<td>ρB</td>
<td>−0.086</td>
<td>[−0.135, −0.037]</td>
</tr>
<tr>
<td>ρ</td>
<td>+0.377</td>
<td>[+0.363, +0.390]</td>
</tr>
</tbody>
</table>

Table 6.15: Preference 1b: Spearman’s rank correlation between Pr(e|pre( Ui )) and Ut(w) (WSJ)

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>95% confidential interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρA</td>
<td>+0.454</td>
<td>[+0.444, +0.463]</td>
</tr>
<tr>
<td>ρB</td>
<td>−0.120</td>
<td>[−0.134, −0.105]</td>
</tr>
<tr>
<td>ρ</td>
<td>+0.237</td>
<td>[+0.231, +0.243]</td>
</tr>
</tbody>
</table>

Preference 1a

• ρB: Correlation in (B) samples; i.e., negative samples not complying with Preference 1a

• ρ: Correlation in all samples.

It is clearly predicted that ρA should be positive and ρB should be negative (Figure 5.4). If Preference 1b is followed by the samples in the corpora, the following inequality can be predicted to occur:

|ρA| > |ρB|, ρ > 0

If Preference 1b is not followed in the corpora, the following equality can be predicted to occur:

|ρA| ≃ |ρB|, ρ ≃ 0

85
Table 6.16: Preference 2: Avg. EU(refmap($U_{i+1}$)) of each transition type (Mainichi)

<table>
<thead>
<tr>
<th>transition type</th>
<th>#sample</th>
<th>Avg. EU(refmap($U_{i+1}$))</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue</td>
<td>1,783</td>
<td>6.89</td>
<td>[6.68, 7.09]</td>
</tr>
<tr>
<td>Retain</td>
<td>84</td>
<td>5.07</td>
<td>[3.99, 6.16]</td>
</tr>
<tr>
<td>Smooth-Shift</td>
<td>2,704</td>
<td>1.59</td>
<td>[1.49, 1.68]</td>
</tr>
<tr>
<td>Rough-Shift</td>
<td>194</td>
<td>0.81</td>
<td>[0.57, 1.05]</td>
</tr>
</tbody>
</table>

Table 6.17: Preference 2: Avg. EU(refmap($U_{i+1}$)) of each transition type (WSJ)

<table>
<thead>
<tr>
<th>transition type</th>
<th>#sample</th>
<th>Avg. EU(refmap($U_{i+1}$))</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continue</td>
<td>13,384</td>
<td>5.90</td>
<td>[5.83, 5.98]</td>
</tr>
<tr>
<td>Retain</td>
<td>2,314</td>
<td>3.67</td>
<td>[3.53, 3.80]</td>
</tr>
<tr>
<td>Smooth-Shift</td>
<td>18,904</td>
<td>2.96</td>
<td>[2.91, 3.01]</td>
</tr>
<tr>
<td>Rough-Shift</td>
<td>5,628</td>
<td>1.17</td>
<td>[1.10, 1.23]</td>
</tr>
</tbody>
</table>

We calculated $\rho_A$, $\rho_B$ and $\rho$ to determine whether the first inequality was true. Tables 6.14 and 6.15 show the results for both corpora.

The results match the first inequality. Moreover, the 95% confidential intervals show the statistical significance of the results. Therefore, the results empirically prove the validity of Preference 1b.

6.3.4 Verification of Preference 2

Rule 2 of centering theory is generalized as Preference 2, which is just the principle of expected utility. Below, we verify the hypothesis that Rule 2 of centering theory can be reduced to the principle of expected utility.

Rule 2 of centering theory is represented as the ranking of the four transition types: Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift. To verify the hypothesis, we investigated the consistency of EU(refmap($U_{i+1}$)) with the above transition ranking by using the following procedure:
6.3. VERIFYING MEANING-GAME-BASED CENTERING MODEL

1. Determine $C_b(U_i)$, $C_b(U_{i+1})$, and $C_p(U_{i+1})$ by using the reference probability instead of the $C_f$-ranking as follows:
   
   $C_b(U_i)$: The entity in $\bigcup_{k=1}^{i} C_f(U_k)$ which has the highest $Pr(e|\text{pre}(U_{i-1}))$
   
   $C_b(U_{i+1})$: The entity in $\bigcup_{k=1}^{i+1} C_f(U_k)$ which has the highest $Pr(e|\text{pre}(U_i))$
   
   $C_p(U_{i+1})$: The entity in $\bigcup_{k=1}^{i+1} C_f(U_k)$ which has the highest $Pr(e|\text{pre}(U_{i+1}))$

2. Determine the transition types (Continue, Retain, Smooth-Shift, or Rough-Shift) on the basis of $C_b(U_i)$, $C_b(U_{i+1})$, and $C_p(U_{i+1})$.

3. Perform Wilcoxon’s rank sum test for verifying consistency of the order of Rule 2 (Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift) with the order of average $EU(\text{refmap}(U_{i+1}))$ between transition types.

4. Calculate Spearman’s rank correlation coefficient between $EU(\text{refmap}(U_{i+1}))$ and the order of Rule 2.

Tables 6.16 and 6.17 show the average $EU(\text{refmap}(U_{i+1}))$ for each transition type. Figures 6.11 and 6.12 show the average and distribution of $EU$ for each transition type. The order of average $EU$ was consistent with the order of Rule 2 of centering theory in both corpora. Wilcoxon’s rank sum test showed consistency between the order of average $EU$ and the order of Rule 2 at a significant level ($<2.2 \times 10^{-16}$). Spearman’s rank correlation coefficients between $EU$ and the order of Rule 2 were as follows: $0.639$ (95% confidential interval: $[0.621, 0.655]$) in the Mainichi corpus and $0.482$ (95% confidential interval: $[0.474, 0.489]$) in the WSJ corpus.

These results indicate the consistency of $EU(\text{refmap}(U_{i+1}))$ with Rule 2, with statistical significance. This empirically proves that Rule 2 of centering theory can be reduced to the principle of expected utility in both Japanese and English.

Thus, the overall hypothesis that centering theory can be reduced to game theory was empirically verified in both Japanese and English corpora. In conclusion, $S$ and $R$’s cooperative preference of referential coherence does not depend on language because Japanese and English are quite dissimilar.
6.4 Discussion

6.4.1 Characteristics of Expected Utility as a Scale of Referential Coherence

We discuss whether the expected utility satisfies the characteristics of scale of referential coherence.
6.4. DISCUSSION

The verification of Preference 1a strongly indicates that $EU(\text{refmap}_A(U_{i+1})) - EU(\text{refmap}_B(U_{i+1}))$, the difference between expected utilities of candidates, can be regarded as $S$ and $R$’s criterion for selecting an expression and an interpretation of the succeeding utterance. Concretely, Spearman’s rank correlation coefficients between $EU_A - EU_B$ and the ratio of the positive samples (A) were 0.833 in Mainichi and 0.981 in WSJ. This proves the language universality of our hypothesis because Japanese and English are quite dissimilar. We consider that various discourse compilation systems can take the expected utility as the criterion to select the expression of the succeeding utterance from its candidates. Moreover, we suppose that selection of a low $EU(\text{refmap}(U_{i+1}))$ potentially indicates the topic shift between $U_i$ and $U_{i+1}$.

We will verify this conjecture in the future.

The verification of Preference 2 indicates that $EU(\text{refmap}(U_{i+1}))$ can be regarded as the smoothness of the transition type between $U_i$ and $U_{i+1}$. Concretely, Spearman’s rank correlation coefficients between $EU(\text{refmap}(U_{i+1}))$ and the transition ranking (Continue $\succ$ Retain $\succ$ Smooth-Shift $\succ$ Rough-Shift) were 0.639 in Mainichi and 0.482 in WSJ. Figures 6.11 and 6.12, however, show that slight “Rough-Shift” samples can distribute in the greater $EU(\text{refmap}(U_{i+1}))$ range than partial “Continue” samples do. This problem is because these distributions were found by using a mixture of different samples that have different preceding contexts. We consider that if we canonicalize the difference of the preceding context, we may be able to get a stronger correlation. We will try to find a method of canonicalization in the future.

6.4.2 Quantitative Comparison of Japanese and English Corpora

Here, we quantitatively compare Mainichi and WSJ from the viewpoint of Preference 1b. Although the correlation coefficients were significantly positive in both corpora, the coefficient for WSJ was less than for Mainichi. Figure 6.13 indicates the reason. It represents the correlation between the
Figure 6.13: Reference probability and ratio of pronominalized entities

Pr value and the ratio of pronominalized entities, i.e., high-Ut entities. In the range of $\Pr < 0.75$, the ratio of pronominalized entities increased with $\Pr$ in both corpora. In the range of $\Pr > 0.75$ (i.e., the range of salient entities), the correlations were, however, different between corpora. In Mainichi, the pronominalization ratio smoothly increased with $\Pr$ in this range, whereas WSJ, the pronominalization ratio did not increase in this range.

We investigated this difference in the range of the salient entities. In Mainichi, only 17.6% samples were not pronominalized in this range; in WSJ, 55.3% (11,367) samples were not pronominalized. In particular, 41.7% (4,735) samples in the non-pronominalized samples were referenced by the proper nouns in WSJ.

In Japanese, a salient entity is frequently referenced by a zero pronoun. In English, especially in the newspaper articles, a salient person tends to be comparatively referred to by his last name (e.g., “Dr. Talcott” instead of “he”) as in the following example:

“We have no useful information on whether users are at risk,” said James A. Talcott of Boston’s Dana-Farber Cancer Institute. Dr. Talcott led a team of researchers from the National Cancer Institute.
6.5 Conclusion

The definition of perceptual utility has still room for improvement in this respect. At the same time, this difference empirically indicates that Japanese discourse has a stronger tendency to reduce perceptual load under the influence of discourse salience than English discourse does. We consider that such quantitative analysis about referential coherence can be regarded as an effect of our quantitative formulation.

6.5 Conclusion

This chapter empirically evaluated and verified our formulations by using corpora. Firstly, we evaluated the calculation method of $\text{Pr}(e|\text{pre}(U_i))$ as a degree of discourse salience. Secondly, we verified language universality of the meaning-game-based centering model.

Evaluating Calculation of Reference Probability

We empirically evaluated our proposed calculation method of the discourse salience, i.e., the regression-based calculation of $\text{Pr}(e|\text{pre}(U_i))$ that incorporates the window functions. As a preparation, we selected the optimal window function from candidate window functions. This resulted in the scientific clarification of the characteristics of recency effect in discourse. We found out that the inverse window function was suitable for handling the recency effect. The result also indicated that the decay curve of recency effect is similar to inverse function because the window functions are optimized to inverse window functions. Collaterally, we found that the grammatical function has wider influence extent than part-of-speech and simple occurrence. We also found that the influence extent in spontaneous conversation is wider than that in newspaper. Moreover, we found out that handling spoken language needs the integrating features more than handling written language does.
CHAPTER 6. EMPIRICAL EVALUATION

Verifying Meaning-Game-based Centering Model

We empirically verified language universality of hypothesis by using large Japanese and English corpora. This resulted in the scientific clarification of the universal principle to select a referentially coherent utterance. That is to say, the result showed the empirical proof of the language universality of the hypothesis that referential coherence can be accounted for by the meaning game. Thus, we clarified that the principle behind the referential coherence is to maximizing expected utility (i.e., game-theoretic behavioral principle). The results also showed the theoretical validity of the reference probability and the perceptual utility.

From these results, we consider that our formulation has a potential to be a criterion to automatically select the succeeding utterance from candidates in various languages. Furthermore, our formulation will provide a method of game-theoretical and statistical analysis about referential coherence. These results will enable us to develop a mechanism to cooperatively produce utterances in the future.
Chapter 7

Application to Information Provision

This chapter mentions developing prototype systems that provide contextual information. These are toward the realization of “context-grounding support systems”, i.e., the systems to help discourse participants to share their understanding of context. Concretely, we developed the following elemental technologies for this purpose.

1. Providing a graph representing time-series overview of discourse
2. Providing text Information related to users’ discourse context

7.1 SalienceGraph: Providing Time-Series Overview of Discourse

We developed a prototype system that visualizes transition of $\Pr(e|\text{pre}(U_i))$, which represents the flow of the target discourse. Figure 7.1 is an example of visualizing an overview of a conference minute. Here, we term this graph the “SalienceGraph”. The abscissa axis means the discourse progress, i.e.,
the sequence of utterance. The ordinate axis means the discourse salience, i.e., the reference probability of each word. The target conference minute is included in corpus of public debate, i.e., conference minutes of Yodo-river committee[5, 98]. We describe the specification of the corpus in Section 7.3.

The prototype system requires dictation of the discourse yet. The reference probabilities in Figure 7.1 are calculated with a logistic regression model obtained from the CSJ corpus. We suppose that the SalienceGraph helps users to grasp an overview of long discourse. The target words can
7.1. SaliencGraph: Providing Time-Series Overview of Discourse

Figure 7.3: Comparing SaliencGraph and TF-based visualization

be automatically selected on the basis of their salience. After grasping the overview, users can zoom into demanded area. For example of Figure 7.1, if a user wants to see the discussion about water shortage, she can zoom into the detail as Figure 7.2. Moreover, the SaliencGraph should enable her to add a new target word. We aim to apply it to browsing long discourse (e.g., conference minutes and judicial records). It can be a visual interface satisfying the Visual Information-Seeking Mantra: “Overview first, zoom and filter, then details on demand.” [2].

Additionally, Figure 7.3 shows comparison between the SaliencGraph based on $\Pr(e|\text{pre}(U_i))$ and a graph based on TF-IDF windows by rectangular window function. The superiority of the SaliencGraph has not been evaluated yet. We have to evaluate the effectiveness of the SaliencGraph in the future.
7.2 Providing Information Related to Users’ Discourse Context

We developed a prototype system providing text information (e.g., past discussion) related to the users’ current agenda. We suppose that providing information related to the current context of debate helps the participants to consider diverse viewpoints and to avoid disproportion of discussion. In this thesis, we call such information provision as the “contextual information provision.”

7.2.1 Representing Discourse Context as Salience Vector

The contextual information provision requires a representation of discourse context that dynamically changes with each utterance unit. Hence, let us formulate a representation of the discourse context. We can formulate such “context representation” on the basis of discourse salience, instead of Bag-of-Words as a “text representation” based on TF-IDF. We assume that the discourse context can be represented by the joint attentional state of the discourse participants. The attentional state can be represented as a vector consisting of salience of the discourse entities. In other words, we simply formulate the discourse context as the vector consisting of salience. We term the vector the salience vector.

To formulate the salience vector, it requires extracting \( \{e_1, \ldots, e_N\} \), a set of \( N \) words as discourse entities, from a corpus. Each extracted word is respectively assigned to one dimension in the \( N \)-dimension term space. Then, we can formulate a representation of the discourse context at the moment when a target utterance \( U \) is conveyed as follows:

\[
v(U) = \left[ \Pr(e_1|\text{pre}(U)), \ldots, \Pr(e_N|\text{pre}(U)) \right]^T
\]

Figure 7.4 shows an example of salience vectors, which change along with the discourse progress.
7.2. PROVIDING INFORMATION RELATED TO USERS’ DISCOURSE CONTEXT

Figure 7.4: Example of salience vectors changing with the discourse progress

7.2.2 Formulating Context Similarity

This section formulates the context similarity on the basis of the salience vector. Here, let \( \mathbf{v}(U) \) be the salience vector representing the discourse context at the moment when a target utterance \( U \) is conveyed between discourse participants (e.g., the users of information provision system). Let \( \mathbf{v}(S) \) be the salience vector representing the discourse context at a target sentence \( S \) (e.g., a sentence in a candidate text for provision). We formulate the context similarity between \( \mathbf{v}(U) \) and \( \mathbf{v}(S) \), as the cosine similarity between \( \mathbf{v}(U) \) and \( \mathbf{v}(S) \).

\[
\text{sim}(\mathbf{v}(U), \mathbf{v}(S)) = \frac{\mathbf{v}(U) \cdot \mathbf{v}(S)}{||\mathbf{v}(U)|| ||\mathbf{v}(S)||}
\]

This similarity deals with the dynamic change of discourse context because the salience vector comprises \( \Pr(e|\text{pre}(U_i)) \), which deals with the dy-
namic transition by incorporating the recency effect.

7.2.3 Developing Information Provision System

Figure 7.5 shows an example of the users’ discourse (in the CSJ corpus) and the provided text by the system. In this case, the sentence surrounded by the red box is regarded as the current sentence, i.e., the latest spoken one.

The procedure for the contextual information provision is as follows:

1. **Obtain regression models for text and for conversation**: As a preparation, train regression models for texts to provide and for users’ conversation, respectively.

2. **Indexing each sentence in a set of document to provide**: As a preparation, create salience vector $v(S)$ as an index for the target sentence $S$. The index represents the discourse context influenced by the preceding context $\text{pre}(S)$.

3. **Create query from users’ conversational context**: From users’ conversation $[U_1, \ldots, U_i]$, create salience vector $v(U_i)$ as a query. The query represents the users’ current conversational context.

4. **Search contextually similar sentence**: Find $k$ candidate sentences $S$, which have the maximum $\text{sim}(v(U_i), v(S))$.

5. **Provide the sentence**: Select $S$ to provide from the found $k$ candidates, and provide it (with its adjacent sentences). Here, we used only $\text{sim}(v(U_i), v(S))$ as the selection criteria.

It is a kind of query-free information retrieval because the query $v(U_i)$ is automatically generated from users’ conversation. We simply implemented the prototype system on the basis of PostgreSQL[99, 100]. The prototype system does not run in real time yet. It requires dictation of users’ conversation yet.
7.2. PROVIDING INFORMATION RELATED TO USERS’ DISCOURSE CONTEXT

Figure 7.5: Contextual information provision based on salience vector (Pr(ε|pre(Ui))): Providing text related to the current context of users’ discourse

Figure 7.6: Example of a generated salience vector as a query representing users’ discourse context

An example of the contextual information provision is shown in Figure 7.5. This is provided by our prototype system based on the salience vector. We used the dictation of spontaneous dialogue (D03F0040 in CSJ, which consists of 495 IPUs) as the users’ conversation. As a preparation, we collected 18,429 web pages (which comprise 397,089 sentences) as candidate texts to provide. We automatically annotated the web pages with the salience vectors for each sentence by using a regression model obtained from Mainichi.
the situation of Figure 7.5, the last utterance surrounded by the oval box is regarded as the current utterance $U_i$. The system automatically generates a query $\mathbf{v}(U_i)$ that represents the discourse context at the current utterance. The elements of $\mathbf{v}(U_i)$, the generated query, are shown in Figure 7.6. It included the influences by the preceding context $\text{pre}(U_i)$. The system found the candidate $S$ that has maximum $\text{sim}(\mathbf{v}(U_i), \mathbf{v}(S))$ by the query $\mathbf{v}(U_i)$ as shown in Figure 7.5. In this example, the salience vectors $\mathbf{v}(U_i)$ and $\mathbf{v}(S)$ mutually included “関西人” and “東京” with high salience. The provided text, however, included different viewpoint about these topics. This is desirable to help users to know diverse viewpoints related to the current context.

On the other hand, Figure 7.7 shows an example provided on the basis of Bag-of-Words, which consists of TF-IDF. We have to evaluate the superiority of our method.

Although we used spontaneous dialogue in CSJ here, we will further apply the technology to supporting public debate. For instance, providing information (e.g., past discussion) related to the current agenda is likely to be
useful to avoid disproportion of discussion because it can help the debate participants to know diverse viewpoints about the current agenda.

7.3 Corpus of Conference Minutes

The above example of the SalienceGraph represents an overview of a particular conference minute in the Yodo-river committee\[5, 98\]. The Yodo-river committee, a lower organization of Ministry of Land, Infrastructure and Transport, posts the corpus on their website. We automatically annotated 226 minutes with syntactic GDA tags by using CaboCha\[86\], toward supporting the Public Involvement (PI) process. We call the corpus as the “Yodo-river corpus” in this thesis.

7.4 Conclusion

We developed the elemental technologies (i.e., the prototype systems of the SalienceGraph and the contextual information provision) to help the users to share their understanding of the discourse context. We presented an example of the SalienceGraph and discussed its usages to browse long discourse. We presented an example of the contextual information provision.

Moreover, we described the specification of the Yodo-river corpus. It can be used for research and development of debate support systems. We describe some perspectives about potential applications in Chapter 8.
Chapter 8

Perspectives and Future Work

This chapter discusses major contributions of our work and difference from related studies. Furthermore, we described our perspective for the future, that is, the potential applications of our elemental technologies and future works to practically realize the applications.

8.1 Major Contribution

We formulated the salience-based model of discourse context. The major contributions of our formulation are as follows:

[Technical Contribution 1] Statistical basis to design the calculation method of salience: Our formulation of the discourse salience, Pr(e|pre(U_i)), enabled us to statistically integrate the influencing factors of salience on the basis of a corpus. Moreover, our formulation of the evaluation criteria, evalSal(m), also enabled us to optimize the calculation method of the discourse salience. Through the formulations, we found out that the window function was suitable for handling the recency effect with an engineering approach. Incorporating the window function enabled us to handle the dynamic change of the discourse context. As a result, evalSal(m) of our method became higher than
a naive term weighting scheme (i.e., TF windowed by optimized rectangular window). In particular, we found out that the effectiveness of our method in CSJ, spoken dialogues, is more significant than that in Mainichi, newspaper articles. This indicates that handling spoken language needs the integrating features more than handling written language does.

[Technical Contribution 2] Cooperative principle to select a referentially coherent candidate based on game theory: Our formulation of the expected utility $\text{EU}({\text{refmap}}(U_{i+1}))$ provided the quantitative principle to cooperatively select a referentially coherent $\text{refmap}(U_{i+1})$. Furthermore, its corpus-based design enabled us to apply the formulation to various languages by using a corpus of the target language.

[Technical Contribution 3] Time-series context similarity incorporating dynamic transition of the discourse context: Our formulation of the salience-based representation enabled us to handle the dynamic change of context on the basis of $\text{Pr}(e|\text{pre}(U_i))$. As a result, our formulation of the context similarity also enabled us to handle the dynamic change of the context similarity.

[Scientific Contribution 1] Empirical clarification of the characteristics of recency effect in discourse context: Our definition of the evaluation measure $\text{evalSal}(m)$ enabled us to empirically clarify the importance of the candidate influencing factors of the discourse salience. We clarified what shape the decay curve of the recency effect has in the discourse context. Concretely, we found out the decay curve of the recency effect was close to inverse function in the discourse context.

[Scientific Contribution 2] Empirical clarification of cooperative principle behind the referential coherence with language universality: We empirically proved the universality of the meaning game hypothesis. That is to say, we proved that the referential coherence
can be explained by a simple and general principle of expected utility, that is, the principle of expected utility. Our formulation of the referential coherence enabled us to explain the behavioral principle behind the cooperative process of the referential coherence on the basis of game theory. Moreover, its statistical approach enabled us to empirically verify the hypothesis. As a result, we empirically confirmed that the game-theoretic principle can universally explain the referential coherence in Japanese and English corpora.

[Applicative Contribution 1] **Sequential visualization of discourse overview**: We developed the system visualizing time-series discourse overview based on the reference probability. Our discussion about the sequential visualization of discourse that the technology potentially helps the users to share their understanding about the dynamic flow of the discourse context.

[Applicative Contribution 2] **Salience-based retrieval of contextually related text**: We developed the system providing a contextually related text based on the salience vector. Our discussion about the contextual provision suggested that the technology potentially helps the users to consider diverse viewpoints and to avoid disproportion of discussion.

8.2 Difference from Related Studies

This section discusses difference of the salience-based model of discourse context from related studies.

**Difference from Centering Theory**  Our salience-based modeling of discourse context is inspired by centering theory. Although our model is consistent with centering theory, it is technologically extended and theoretically simplified.
Firstly, we quantified the discourse salience and the referential coherence by integrating salience factors. On the other hand, centering theory had not been quantified because it was rule-based theory. Centering theory deals with only one salience factor, that is, grammatical function. Our model is more easy to use computing than centering theory due to the quantification. The definition of context similarity in Chapter 7 is one of the instances. Furthermore, our model is more suitable to represent discourse salience than centering theory due to the integration of salience factors. Hence, we regard our model as technologically extended as compared to centering theory.

Secondly, our meaning-game-based centering model is explained by the simple and general principle, i.e., principle of the expected utility. On the other hand, centering theory consists of complex rules. Additionally, the different versions of centering theory are proposed by different researchers. Hence, we regard our model as theoretically simplified as compared to centering theory.

**Difference from Term-Weighting Schemes** The term-weighting schemes such as \( \text{TF-IDF} \) is basically based on the term frequency. Our method also deals with frequency by calculating summation of features that is weighted by a window function. As shown in Chapter 6, our formulation of reference probability is more specialized in dealing with the dynamic transition than the term-weighting schemes. Additionally, in Chapter 6, we found that the term weighting schemes can be improved to deal with the dynamic transition by using suitable window function, i.e., the inverse window function.

**Difference from Spreading Activation Theory** The “activation” in spreading activation theory plays a similar role to the discourse salience. Although it is suitable to represent priming effect, it does not deal with the grammatical function and part-of-speech that our model dealt with. On the other hand, our model does not deal with the indirect priming effect. We should investigate the quantitative difference between them in the future.
8.2. DIFFERENCE FROM RELATED STUDIES

**Difference from TextTiling and Topic Sequence**  There is a similar point between our work and TextTiling[35], that is, both studies use window function and handle transition of discourse context. The aim of our work, however, is different from that of TextTiling. TextTiling is an algorithm to determine boundaries between text segments. On the other hands, our aim is to represent the dynamic flow of discourse context and to apply it to information provision. Additionally, the usages of window functions in these studies are also different. TextTiling uses two rectangular windows that are adjacent to each other for segmentation. Our model uses one inverse window for handling the decay curve of the recency effect.

Let us discuss difference between visualizing approaches based on these models. Figure 8.1 shows SalienceGraph visualizes discourse flow on basis of the discourse salience. It deals with the transition of the discourse context for each utterance. Users can easily grasp the flow of discourse from the SalienceGraph because the meanings of the axes are clear: the abscissa axis means the sequence of utterance and the ordinate axis means the salience of each word.

On the other hand, Topic Sequence[47] visualizes discourse flow on the basis of text segmentation like the TextTiling. It does not deal with the transition for each utterance because it is based on the text segmentation. Although it represents rich information about the relationships between terms,
the meanings of the axes are not clear.

Both type of visualization can be applied to user interface of discourse browsing system. It is important to use the suitable one according to the user’s situation and demand.

**Difference from ThemeRiver** ThemeRiver visualizes chronological transition of themes in a collection of documents[45, 46]. Although it is useful to understand chronological change, it is not for applying to discourse browsing system.

### 8.3 Potential Applications

Our elemental technologies for contextual information provision are toward developing the systems that help discourse participants to share their context. We suppose that such systems are desired because debate participants often fail to share their understanding due to difference of their backgrounds (e.g., in PI process). We call the desired systems “context-grounding support systems”. The elemental technologies have a potential to be applied to the following fields.

#### 8.3.1 Discussion Analysis Support System

Discussion analysis is required to support the Public Involvement (PI) process. PI is a citizen participation process in the decision making of public policy[3, 4, 5]. To support the public debates in PI, we have to investigates the characteristics of appropriate debate. For this investigation, a discussion analysis support system is required. Concretely, it will include the debate browsing system based on the SalienceGraph. Additionally, we will use the Yodo-river corpus to develop the discussion analysis support system.
8.3.2 Debate Support System

Debate support systems are also likely to be helpful for the PI process. The elemental technologies that we developed can be extended toward a debate supporting system as follows:

- Real-time visualization of SalienceGraph
- Real-time provision of contextually related information

The real-time SalienceGraph is likely to help the debate participants to recognize agenda in the preceding discourse. We conjecture that it also help facilitators of the debate to organize an appropriate flow of discussion. For example, it is possible that the facilitator perceive a shortage of discussion from a SalienceGraph in order to determine how to manage the debate effectively.

Furthermore, The real-time provision of contextually related information is likely to help the debate participants to avoid disproportion of discussion. We conjecture that it will provide diverse viewpoints related to the current agenda. Although errors in automatic speech recognition will be a problem in developing the debate support system, it is worth trying to apply the public debates in order to support appropriate decision making.

8.3.3 Call Center Support System

The elemental technologies that we developed can also be applied to a call center support system. The contextual information provision, which provides hints related to the conversational context between a service agent and a client, is likely to help them to solve the client’s problem. Firstly, we will aim to develop a system supporting a chat-based call center service. We conjecture that the chat-based call center service will become common because it has the following merits.

- Chat-based service can be provided with visual information as hints to solve the client’s problem.
• The chat-based service has higher affinity to our technologies than the conventional call center because it is free from automatic speech recognition errors.

• Advices in text tend to not be forgotten.

• One service operator can concurrently respond to multiple clients.

8.3.4 Conversation-Targeted Advertisement

The contextual information provision can be applied to “conversation-targeted advertisement”. That is to say, the system automatically creates query from users’ conversational context and provides related advertisement. It has a potential to be used for a business.

Hoshi, a Japanese novelist, published a short-short story about a advertisement system targeted for phone conversation[101]. In this story, the system provides spoken advertisements related to the conversational context. Such conversation-targeted advertisement can be regarded as a variation of content-targeted advertisement. Although the spoken advertisements disturbed the conversation, we consider the idea worth applying to the recent communication tools on the internet (e.g., chat).

8.4 Future Works

We have the following remaining issues for the technical and application layers.

8.4.1 Remaining Issues for Technical Layer

Reduce labor for manual annotation: Although the practical calculation phase of Pr(e|pre( Ui )) does not need anaphoric annotation, the learning phase requires it. We should try to reduce the labor for manual annotation. We consider the following two approaches to this issue.
The first one is to use automatic anaphora resolution. We did not use anaphora resolution in the feature extraction for calculating $Pr(e|pre(U_i))$ because the accuracy of automatic anaphora resolution is not sufficient yet. We will incorporate it with the feature weighting according to confidence degree of automatic anaphora resolution.

The second one is to formulate salience as just occurrence probability instead of the reference probability. Calculating occurrence probability does not need the anaphoric tags. At the present stage, however, we conjecture that the reference probability is better than the occurrence probability as the scale of discourse salience. We have to verify this presumption in the future.

**Incorporate indirect priming effect to predict indirect anaphora:**
Although we did not deal with indirect priming effect, it is required to predict anaphora more accurately. Predicting indirect anaphora is specifically conjectured to require dealing with the indirect priming effect. We can deal with the priming effect by rotating base vector of term as mentioned in Appendix of this thesis. Furthermore, we can also use PLSA and LDA (mentioned in Chater 2) to deal with the priming effect. Although we also omit the indirect anaphora in this thesis, we have to investigate the relationship between the indirect anaphora and the indirect priming effect in the future.

**Formulate each participant’s subjective salience:** The reference probability that we formulate represents objective discourse salience, which is shared by discourse participants. We further need to formulate each discourse participant’s subjective salience because it is likely to be useful for discussion analysis for public debates. Analyzing the flow of discussion requires visualizing the each participant’s subjective context. To formulate their subjective salience, we will expand the formulation of $Pr(e|pre(U_i))$ by incorporating their turn-taking[102, 103].

**Incorporate prosodic information:** Prosodic information influence to
CHAPTER 8. PERSPECTIVES AND FUTURE WORK

discourse salience. We should investigate effective features in prosodic information.

**Evaluate** $\Pr(e|\text{pre}(U_i))$ by using WSJ: Although the calculation of $\Pr(e|\text{pre}(U_i))$ was evaluated on the basis of CSJ and Mainichi, it has not been evaluated on the basis of WSJ. We will further evaluate it by using WSJ in the future, in order to investigate the characteristics of discourse salience in English newspaper.

**Verify** EU(refmap($U_{i+1}$)) by using CSJ: Although the formulation of EU(refmap($U_{i+1}$)) was verified on the basis of Mainichi and WSJ, it has not been evaluated on the basis of CSJ. We will further verify it by using CSJ in the future, in order to investigate the characteristics of referential coherence in spontaneous dialogue.

**Evaluate** the context similarity based on the salience vector: Although the formulations of reference probability and referential coherence were evaluated, the formulation of the salience-based formulation of context similarity has not been quantitatively evaluated. We will further formulate a evaluation scale for the calculation method of context similarity. Then, we will quantitatively evaluate it in the future.

8.4.2 Remaining Issues for Application Layer

**Improve** the prototype systems toward practical use: Our systems do not run in real time because these require dictations of a target discourse. We will deal with the result of automatic speech recognition in the future. Moreover, the contextual information provision is too heavy and slow to practically use because we implemented it with naive programming. We have to improve the implementation to speed up them.

**Evaluate** the effectiveness of information provision: We have to evaluate the effectiveness of our systems through demonstration experiment.
Develop a method for rearranging utterance units: $EU(\text{refmap}(U_{i+1}))$, which represents the referential coherence between $\text{pre}(U_i)$ and $U_{i+1}$, can be used as criteria to determine the easy-to-understand order of utterance units. The rearrangement is likely to be useful to determine “how to say (provide)” for the contextual information provision system. To improve understandability of provided text, we will further develop a method for rearrangement of utterance units on the basis of $EU(\text{refmap}(U_{i+1}))$.

Develop a discourse browsing system: As discussed in this chapter, we will develop a discourse browsing system on the basis of salience-based

Figure 8.2: Future work: required interface of discourse browsing system
CHAPTER 8. PERSPECTIVES AND FUTURE WORK

visualization shown in Figure 8.2, which will satisfy the Visual Information-Seeking Mantra: “Overview first, zoom and filter, then details on demand.”

For instance, Figure 8.2 shows an example of required interface and user’s operation on the discourse browsing system. We will use this to analyze the flow of discussion in public debates for the PI process.

Develop a discussion analysis support system: As discussed in this chapter, we will develop a discussion analysis support system in order to support investigating the characteristics of appropriate debate. It will include the discourse browsing system based on the SalienceGraph.

Develop a debate support system: As discussed in this chapter, we will develop a debate support system on the basis of the two elemental technologies that we developed. Concretely, we will develop a real-time SalienceGraph and a real-time contextual information provision to apply to the debate support system. Furthermore, we will also develop a visualizer of relationship between debate participants after formulating each participant’s subjective salience.

Develop a call center support system: As mentioned in this chapter, we will apply the contextual information provision that we developed to call center support system. Firstly, we will aim to develop a system supporting a chat-based call center service because it has higher affinity to our technologies than the conventional call center.
Chapter 9

Conclusion

In this thesis, we established a salience-based model of discourse context that can be used to develop information provision systems to support consensus building (e.g., for the PI process). We divided this study into the two layers: the technical layer and the application layer. For the technical layer, we established the formulation of the salience-based model of discourse context. For the application layer, we developed prototype systems that provide contextual information on the basis of salience-based model of discourse context.

The goal for the application layer was to develop information provision systems that can be used to help users to share their understanding of the discourse context. Our aim in this thesis was to develop the two applications:

**Applicative Goal 1** Provide a time-series overview of discourse

**Applicative Goal 2** Provide information related to discourse context

The goal for the technical layer was to formulate a computational model of discourse context that changes along with the discourse progress because the targets of participants’ attention change with each utterance unit. To establish such model, we needed to deal with the three technical issues:

**Technical Issue 1** Formulate discourse salience
CHAPTER 9. CONCLUSION

Technical Issue 2 Formulate referential coherence

Technical Issue 3 Formulate context similarity

Technical Issue 1 is essential to deal with the dynamic change of the discourse context. Technical Issue 2 is required to determine “how to say”, that is, the discourse processing systems need to produce referentially coherent sentences according to the cooperative preference to produce and to interpret the referring expressions. Technical Issue 3 is required to determine “what to say”, that is, the information provision systems need to select contextually consistent contents. To solve these technical issues, we also needed to deal with the following scientific issues:

Scientific Issue 1 Clarify the influencing factors of discourse salience

Scientific Issue 2 Clarify the behavioral principle of referential coherence

We solved these technical and scientific issues through the following approaches:

Solution 1 Probabilistic formulation of discourse salience

Solution 2 Game-theoretic formulation of referential coherence

Solution 3 Vectorial formulation of context similarity

The probabilistic formulation of discourse salience is the basis of our model. In other words, we formulated the referential coherence and the context similarity on the basis of the discourse salience.

To arrive at Solution 1, we formulated the discourse salience as the reference probability \( \Pr(e|\text{pre}(U_i)) \), that is, the probability of the discourse entity \( e \) being referred to in the succeeding utterance unit \( U_{i+1} \). This is because a salient entity tends to be referred to successively. This formulation contributed to the integration of the influencing factors of the discourse salience. We developed the calculation method of \( \Pr(e|\text{pre}(U_i)) \) from the influencing
factors; i.e., the referential features of \( e \) in the preceding discourse \( \text{pre}(U_i) \). In particular, the evaluation result showed that the most important factor is the recency effect of the reference to \( e \). We employed the inverse window function to take into account the recency effect. Doing this enabled the model to handle the dynamic change of the discourse context. Furthermore, we established the evaluation criteria of the method used to estimate salience. This enabled us to optimize the method for calculating salience. We empirically evaluated our formulation of the discourse salience by using corpora. The experimental result showed that our formulation of \( \text{Pr}(e|\text{pre}(U_i)) \) enabled us to estimate the discourse salience more accurately than a naive term-weighting (i.e., TF windowed by an optimized rectangular window).

To arrive at Solution 2, we formulated the referential coherence on the basis of the behavioral principle of expected utility, which is behind the cooperative preference between the discourse participants. This formulation is based on the hypothesis that centering theory can be derived from game theory. Specifically, we formulated the referential coherence between \( \text{pre}(U_i) \) and a candidate \( U_{i+1} \) as the expected utility of the candidate. To ensure the language universality of the formulation, we statistically formulated the language-dependent parameters, i.e., the discourse salience and the perceptual utility, on the basis of corpus. Doing this enabled us to apply the model to various language by using a corpus of the target language. Hence, we empirically verified the language universality of the hypothesis by using Japanese and English corpora. As Japanese and English are quite different, the experimental result indicated that the principle of expected utility is universally behind the referential coherence.

To arrive Solution 3, we formulated the context similarity on the basis of the salience vector, that is, the vectorial representation of the discourse context, which is the vector consisting of salience of the discourse entities. This formulation is based on the assumption that the context similarity between an utterance in a particular discourse and another one in different discourse can be reduced to the similarity of attentional targets. As we formulated the
salience vector by regarding $\Pr(e|\text{pre}(U_i))$ (changing along with the discourse progress) as the discourse context, the formulation enabled us to calculate the context similarity reflecting the dynamic change of the context.

Our contributions are thus summarized as follows:

**Technical Contribution 1** We provided the statistical basis needed to design the calculation method of salience. Through formulating $\Pr(e|\text{pre}(U_i))$, we provided the statistical criteria needed to integrate the salience factors. Moreover, through formulating $\text{evalSal}(m)$, we also provided the statistical criteria to optimize the calculation method. As a result, we found out that handling spoken language needs the integrating features more than handling written language does.

**Technical Contribution 2** We provided the quantitative principle needed to cooperatively select a referentially coherent candidate based on game theory. Furthermore, the corpus-based formulation of $\text{EU}(\text{refmap}(U_{i+1}))$ enabled us, by using a corpus of the target language, to apply the formulation to various languages.

**Technical Contribution 3** We provided the time-series context similarity that incorporates the dynamic transition of the discourse context. Our formulation of the salience-based representation of discourse context enabled us to handle the dynamic change of context on the basis of $\Pr(e|\text{pre}(U_i))$.

**Scientific Contribution 1** We empirically clarified the characteristics of the recency effect in the discourse context by optimizing the window function according to the evaluation scale $\text{evalSal}(m)$. We found out the decay curve of the recency effect was close to the inverse function in the discourse context.

**Scientific Contribution 2** We empirically clarified the cooperative principle behind the referential coherence with language universality. That is to say, we empirically proved that the referential coherence can be
explained by a simple and general principle of expected utility, that is, the principle of expected utility. Furthermore, the universality of the meaning game hypothesis is also indicated. The empirical verification was done by using the corpus-based formulation of EU(refmap(U_{i+1})).

**Applicative Contribution 1** We developed SalienceGraph, a simple graph visualizing time-series discourse overview that is based on the reference probability. This elemental technology can help the users to share their understanding of the dynamic flow of the discourse context.

**Applicative Contribution 2** We developed a prototype system providing a contextually related text that is based on the salience vector. This system can help the users to consider diverse viewpoints and to avoid the discussion becoming dominated by one particular viewpoint.

We summarize each chapter as follows. In Chapter 1, we introduced our motivations, our goals, and the key issues. We needed to establish a computational model of discourse context, which gradually changes along with the discourse progress, in order to realize advanced processing of discourse.

In Chapter 2, we surveyed the literature related to our goals. We described the contemporary condition that various salience factors proposed in different research fields have not been integrated.

In Chapter 3, we described the development of corpus based on Global Document Annotation (GDA). It consists of APIs for processing GDA, automatic annotation system with GDA, and simple viewers of GDA.

In Chapter 4, we provided a formulation of Pr(e|pre(U_i)) as the scale of discourse salience. We developed a probabilistic method to integrate salience factors and to optimize the salience calculation.

In Chapter 5, we provided a game-theoretic formulation of the referential coherence. To verify the language universality of a hypothesis that the referential coherence can be explained by game theory, we statistically formulated pronominalization and the expected utility.
In Chapter 6, we empirically evaluated and verified our model. To deal with the dynamic transition of discourse context, we empirically determined the optimal window functions. We found out that our approach to formulate salience was more effective for spontaneous conversation than for newspaper articles. Furthermore, we found out empirical evidences of the hypothesis that the referential coherence can be explained by game theory by using large Japanese and English corpora.

In Chapter 7, we developed the elementary technologies used to provide contextual information. We developed prototype systems to support sharing discourse context: one is the system visualizing dynamic transition of salience, and the other is the system providing information related to current discourse context.

In Chapter 8, we discussed our major contributions and the potential applications of our model. Especially, we discussed our future perspectives to support the public involvement process.

We consider this work our first step towards developing “context-grounding support systems”, which encourage cooperative decision making. We will carry on our work toward developing them as the future plan discussed in Chapter 8.
Appendix: Salience-based Approach to Incorporate Indirect Priming Effect

As described in Chapter 2, the indirect priming effect can be dealt with by using the aspect model or the spreading activation model. Incorporating the indirect priming effect is a remaining issue in this thesis. We conjecture that this issue is important for predicting indirect anaphora.

Here, we describe an alternative approach to deal with the indirect priming effect. That is, an approach based on the idea that related two words can be represented as the near two vectors. More concretely, we deal with indirect priming by rotating base vectors of terms on the basis of term co-occurrence in salience vectors.

Assumption: Indirect Priming and Co-occurrence in Attentional State

We assume that two words, which tend to concurrently occur, raise the other’s salience. In other words, we assume that the indirect priming effect is caused by co-occurrence in attentional state. We can represent the attentional state at particular moment as the salience vector, which is defined in Chapter 7.
\( v(U) = [\Pr(e_1|\text{pre}(U)), \cdots, \Pr(e_N|\text{pre}(U))]^T \) \hspace{1cm} (9.1)

\( N \) denotes the number of terms in the target corpus. The salience vector \( v(U) \) is a sparse vector in the \( N \) dimension space. Its elements are the reference probability.

We incorporate the indirect priming on the basis of the above assumption. The influences among terms are acquired from the target corpus as co-occurrence in a salience vector (not in a document).

**Acquiring Average Attentional State \( b_{e_j} \) When Focusing \( e_j \)**

The influences among terms can be obtained from a target corpus. We suppose that an average attentional state \( b_{e_j} \) when the entity \( e_j \) is paid attention represents the influences of \( e_j \) to other related terms. Hence, we obtain \( b_{e_j} \) from corpus. Let us list up the procedure to acquire \( b_{e_j} \).

1. **Extract noun phrase:** Extract all noun phrases \( e_1, \cdots, e_N \) from a target corpus.

2. **Estimate \( v(U) \):** Estimate salience vectors \( v(U) \) for each utterance unit \( U \) in the corpus.

3. **Calculate \( b_{e_j} \):** Carry out the following procedure (3.a, 3.b, 3.c) for each noun phrase \( e_j (j = 1, \cdots, N) \).

3.a. **Extract attentional state with salient \( e_j \):** Extract a set of salience vectors when \( e_j \) is paid attention from the corpus as follows: \( S_{e_j} = \{v(U) \mid \frac{\Pr(e_j|\text{pre}(U))}{||v(U)||} \geq \theta\} \). Notice that \( \theta \) denotes threshold that represents a criteria whether \( e_j \) is salient or not.
3.b. **Calculate summation:** Calculate the summation of the salience vectors weighted by the reference probability of \( e_j \) as follows: 
\[
v_{e_j} = \sum_{u(U) \in S_{e_j}} \Pr(e_j | \text{pre}(U)) v(U) .
\]

3.c. **Normalize:** Normalize \( v_{e_j} \). That is to say, 
\[
b_{e_j} = \frac{v_{e_j}}{\|v_{e_j}\|}
\]
can be regarded as the average attentional state when \( e_j \) is paid attention.

\( b_{j,k} \), the \( k \)-th element in \( b_{e_j} \) obtained from the corpus, represents the influence of \( e_j \) to \( e_k \). In other words, it represents how salient \( e_k \) tends to be when \( e_j \) is salient. We can regard \( b_{e_j} \) as a base vector rotated from \( e_j \)’s original base vector toward the axes of the related terms.

### Incorporating Indirect Priming by Using \( b_{e_j} \)

This section describe calculating salience by incorporating the indirect priming effect with \( b_{e_1}, \ldots, b_{e_N} \) obtained from the target corpus. Here, we regard the \( k \)-th element in \( b_{e_j} \) as the degree that \( e_j \) activates \( e_k \). We thus estimate the salience of \( e_k \), which deals with the indirect priming, as follows:

\[
sal(e_k, U) = \sum_{j=1}^{N} b_{j,k} \Pr(e_j | \text{pre}(U)) \tag{9.2}
\]

The salience vector considering the indirect priming effect, \( V(U) \), consists of \( \text{sal}(e_k, U) \).

\[
V(U) = [\text{sal}(e_1 | \text{pre}(U)), \ldots, \text{sal}(e_N | \text{pre}(U))]^T
\]

We can calculate \( V(U) \) from the original salience vector \( v(U) \) by using \( b_{e_1}, \ldots, b_{e_N} \), because of the equations 9.1 and 9.2.

\[
V(U) = 
\begin{bmatrix}
  b_{1,1} & \cdots & b_{N,1} \\
  \vdots & \ddots & \vdots \\
  b_{1,N} & \cdots & b_{N,N}
\end{bmatrix}
\begin{bmatrix}
  \Pr(e_1 | \text{pre}(U)) \\
  \vdots \\
  \Pr(e_N | \text{pre}(U))
\end{bmatrix}

= 
\begin{bmatrix}
  b_{e_1} & \cdots & b_{e_N}
\end{bmatrix} v(U)
\]

123
Appendix

The formulation shows that $V(U)$ is rotated from $V(U)$ into the nonorthogonal coordinate system which comprises the base vectors $b_{e_1}, \ldots, b_{e_N}$. $v(U)$ represents the attentional state estimated not by incorporating the indirect effect. On the other hand, $V(U)$ represents the attentional state by incorporating the indirect effect.

Experiment and Discussion

Experimental methodology We obtained the particular $b_{e_j}$ by determining $e_j$ as America-mura, a geographic name in Osaka. We regard a sentence unit as a utterance unit $U$ in this section. The experimental procedure is as follows:

0.a. Learn regression model: We obtained regression model for calculating the reference probability from the Mainichi corpus.

0.b. Collect corpus: We collect 69 HTML documents through searching by the keyword “America-mura” in Google. Then, we automatically annotate them by using GDASDK.

1. Extract noun phrases: We extracted $N = 6,293$ noun phrases from the 69 HTML documents.

2. Estimate $v(U)$: We estimated the salience vector $v(U)$ for each utterance unit $U$ by using the regression model obtained from Mainichi. Then, we annotated the 69 documents with $v(U)$ estimated automatically.

3. Calculate $b_{\text{America-mura}}$: We obtained $b_{\text{America-mura}}$ from the 69 documents. $b_{\text{America-mura}}$ represents the average attentional state when “America-mura” is focused. Notice that we tried two settings of the threshold $\theta$: $\theta = 0.1$ and $\theta = 0.2$. 

124
Experimental Result  The elements of $b_{\text{America-mura}}$ obtained for two settings ($\theta = 0.1$ and $\theta = 0.2$) are shown in the following examples.

- **Elements of $b_{\text{America-mura}}$ for $\theta = 0.2$:**
  America-mura:0.647, America:0.369, Osaka:0.258, mura:0.159, security camera:0.139, camera:0.139, check out:0.129, out:0.129, inside:0.128, woman:0.120, man:0.102, center:0.098, crime:0.092, human:0.087, Takoyaki:0.082, Shinsaibashi:0.075, Minami:0.074, police:0.073, ···

- **Elements of $b_{\text{America-mura}}$ for $\theta = 0.1$:**
  America-mura:0.549, America:0.432, Osaka:0.280, security camera:0.212, camera:0.212, woman:0.141, mura:0.134, human:0.119, center:0.114, crime:0.114, town:0.104, harm:0.093, man:0.092, Mitsu:0.086, this time:0.084, name:0.084, eradication campaign:0.082, campaign:0.082, Shinsaibashi:0.082, ···

The elements are sorted by the value.

Discussion  We found that $b_{\text{America-mura}}$ is rotated to the direction of the base vector of “Osaka”, “security camera” and “Shinsaibashi”. This shows validity because these terms are related to “America-mura”. In particular, “security camera” is an example that depends on the recent news articles. This indicates that corpus selection according to the purpose is important.

Furthermore, we found out the following new aspect about threshold $\theta$:

- In $b_{\text{America-mura}}$, the values of related terms in case of $\theta = 0.1$ are greater than that in case of $\theta = 0.2$. That is to say, $b_{\text{America-mura}}$ for $\theta = 0.1$ is rotated more than that for $\theta = 0.2$.

- When $\theta$ is set as greater, the influence of indirect priming become smaller because the set of salience vector is more narrowed down into the case that “America-mura” is particularly salient. At the same time, the acquired $b_{\text{America-mura}}$ become influenced by the bias of data.

- When $\theta$ is set as smaller, the influence of indirect priming become greater because the set of salience vector become representing more general cases.
The setting of small $\theta$ has the risk that the influence of indirect priming become excessive.

**Conclusion of Appendix**

This appendix described a salience-based approach to incorporate the indirect priming effect. We formulated an alternate calculation method of salience, which incorporates the indirect priming on the basis of $b_{e_1}, \cdots, b_{e_N}$ obtained from a target corpus.

We carried out the experiment to acquire $b_{\text{America-mura}}$. The experiment resulted in the qualitative validity of rotating into the direction of the related terms. Furthermore, we found out that we can control the influence of the indirect priming by varying the threshold $\theta$.

This approach has not been quantitatively evaluated. It has to be compared with the other models dealing with the indirect priming, e.g., aspect model and spreading activation model.
Bibliography


BIBLIOGRAPHY


[34] Daichi Mochihashi. Bayesian Approaches in Natural Language Processing (自然言語処理におけるベイズ統計). Technical Report NC2006-49,
The Institute of Electronics, Information and Communication Engineers (IEICE), 2006. (in Japanese).


BIBLIOGRAPHY


BIBLIOGRAPHY


List of Publication by the Author

Major Publications

Journal Papers


International Conference

1) Shun SHIRAMATSU, Kazunori KOMATANI, Kōiti HASIDA, Tetsuya OGATA, and Hiroshi G. OKUNO: Empirical Verification of Meaning-Game-based Generalization of Centering Theory with Large Japanese


**Lecture Note**


**Other Publications (All in Japanese)**


2) Shun SHIRAMATSU, Kazunori KOMATANI, Tetsuya OGATA, and Hiroshi G. OKUNO: Estimating Salience by Incorporating Association Based on Obtaining Related Words from Corpus (コーパスからの関連語獲得に基づく連想を加味した顕現性の推定). In *Proceedings of the*

