

学位論文の要約

Empirical Analysis of the Polysemy of the Japanese Adjective *Atsui*
and the Chinese Adjective *Re*
(日本語形容詞「あつい」と中国語形容詞「热」に関する実証分析)

Wang Haitao

With the advent of Cognitive Linguistics in the late 1970s, the analysis of polysemous lexical items has been a significant focus in cognitive lexical semantics over the past decades. Early research on lexical polysemy networks was predominantly reliant on speaker intuition and interpretation. However, since the mid-1990s, the field has seen a burgeoning emphasis on empirical methods, strengthening its empirical foundations. The theoretical background of this empirical development in Cognitive Linguistics stems from a growing emphasis on its intrinsic identity as a usage-based linguistics.

Empirical methods refer to research approaches that do not rely on introspection and intuition but try to ground linguistic analysis on the objective observation of verifiable empirical evidence. Currently, empirical methods for studying polysemy are distinctly categorized within corpus-linguistic and psycholinguistic paradigms. In corpus linguistics, a method known as the Behavioral Profile (BP) approach was developed to clarify the complex, multidimensional nature of meaning. Until now, the BP approach has been extensively applied to study essential issues in lexical semantics, including polysemy, synonymy, and antonymy, primarily across Indo-European languages. However, few studies have employed the BP approach to investigate polysemy in Japanese and synchronic Chinese. In psycholinguistics, sentence production tests and sentence sorting tasks are common and effective methods for studying polysemy. However, previous experimental studies on polysemy primarily relied on manual calculations with paper cards and on SPSS, which offers a limited range of statistical techniques. Moreover, while combining corpus-based analysis with psycholinguistic experiments has proven to be fruitful for the study of lexical semantics, few studies on polysemy have integrated both the BP approach and psycholinguistic experiments.

In an attempt to fill these gaps, this dissertation utilizes the Japanese adjective *atsui* (hot) and the Chinese adjective *re* (hot) as case studies, exploring their polysemy using quantitative empirical methods, including the BP approach, sentence production tests and sentence sorting tasks. Additionally, the analysis employs R statistical software for data analysis, along with online platforms such as *Sojump* and *KardSort* for conducting experiments. The five main research questions addressed in this dissertation are as follows:

1. Can we identify a prototypical meaning of a polysemous lexical item empirically?
2. Can we measure (dis)similarity or semantic distances between the senses of a polysemous word accurately by quantitative means?
3. Can we use corpus data to describe the semantic relationships between senses and to

capture the distinctive usage characteristics of each sense?

4. What is the specific structure of a polysemous word's semantic network, and what variables determine its categorization in the mental representations of native speakers?
5. Moreover, do native speakers produce groups that resemble the clustering obtained from analysis of corpus-data?

This dissertation is composed of the following seven chapters to address the five research questions outlined above.

Chapter 1 outlines the development of empirical methods in Cognitive Linguistics and polysemy research, highlighting pivotal studies that have shifted the field from introspective methodologies towards empirical, data-driven approaches. It further identifies existing gaps in the literature, particularly concerning the application of empirical methodologies on polysemy research. Additionally, the chapter delineates five specific research questions as detailed above. It also explains the reasoning behind selecting the temperature adjectives *atsui* and *re* for case studies, underlining their significance in Cognitive Linguistics and sensory processing research. The chapter concludes with an outline of the dissertation, summarizing the content and structure of the following chapters.

Chapter 2 presents an extensive review of relevant literature, organized into five sections. The first section provides an overview of the empirical development in Cognitive Linguistics, highlighting pivotal achievements and paradigms that have defined the field. The second section traces the history of polysemy in linguistics. The third section presents the development of theoretical approaches to polysemy in Cognitive Linguistics. The fourth section discusses how corpus linguistics and psycholinguistics have addressed polysemy within Cognitive Linguistics, elaborating on the theoretical foundations that underpin this research by introducing key theories and concepts. The final section provides a critical overview of research on the polysemy of the temperature adjectives *atsui* in Japanese and *re* in Chinese.

Chapter 3 describes the methodology employed for this dissertation, introducing the BP approach and detailing the psycholinguistic experiments, which consist of a sentence production test and a sentence sorting task. The BP approach commonly involves four steps: (i) retrieving a sample of the target expression(s); (ii) annotating concordance lines with various features; (iii) converting these data into a table of percentage vectors detailing feature occurrence; (iv) analyzing the data with exploratory statistical techniques. The production test mainly involves having participants independently produce instances of the target polysemous word. In a sentence sorting task, participants are presented with a series of sentences containing the polysemous word in various contexts, and are asked to sort these sentences based on a specific criterion, generally semantic one, in polysemy research.

Chapter 4 adopts the corpus-based BP approach to investigate the semantic relationships and usage characteristics of the senses of *atsui* and *re*. Initially, it identifies the prototypes of these adjectives through various corpus data analysis. For *atsui*, diachronically, sense 1: (*of an object*) *having a high temperature*, is the earliest attested in historical records. Synchronically, this sense shows a high family resemblance and fewer formal constraints, and ontogenically, it is the earliest acquired and frequently used. Similarly, for *re*, the chapter reveals that diachronically, sense 1: (*of objects*) *having a relatively or noticeably high temperature*, most closely relates to the historically

earliest attested sense. Synchronically, this sense dominates in frequency and exhibits fewer formal constraints in the corpus data. Ontogenically, it is the earliest acquired and used sense by children. Subsequently, hierarchical agglomerative cluster (HAC) analyses are used to illustrate the (dis)similarity relationships between the senses of *atsui* and *re*, with the results presented in dendrograms. The HAC analyses reveal that senses derived from the same sensory or subjective experiences tend to have similar usage characteristics. Finally, computing *t*-values has identified the distinctive morphosyntactic properties of various clusters for both *atsui* and *re*. The usage patterns and characteristics of each sense within a given cluster have been summarized.

Chapter 5 employs psycholinguistic experimentation to investigate the category structures of *atsui* and *re* in the cognitive representations of native speakers. Initially, this chapter adopted sentence production tests to identify their prototypes. For *atsui*, sense 1 was identified as the prototype based on its high frequency of activation among participants and its early activation in a substantial proportion of participants. For *re*, sense 1 was identified as the prototype since it was the most frequently activated sense and the earliest activated sense in most participants. Next, the chapter employed sentence sorting tasks to further investigate the semantic structures. The results of the HAC verified and supported the sense identification of *atsui* and *re*. Through the results of multidimensional scale (MDS) analysis, the (dis)similarity of senses was quantified. For *atsui*, sense 4 (*marked by intense emotion and excitement*) and sense 6 (*furiously, raging*) emerged as the most similar, whereas sense 1 and sense 9 (*very popular*) were identified as the most dissimilar. For *re*, sense 1 and sense 2 (*feeling or producing an uncomfortable sensation of body heat*) exhibited the highest similarity. In contrast, sense 2 and sense 7 (*popular, highly concerned*), as well as sense 2 and sense 9 (*overheated [economics], too prosperous and active in commerce*) showed the highest dissimilarity. Moreover, this chapter employed the results of MDS and HAC to identify the variables that native speakers used to categorize the senses of *atsui* and *re*. For *atsui*, these variables include ‘concrete’, ‘abstract’, ‘weather’, ‘tangible temperature’, ‘emotion’, ‘self-perception’, ‘interpersonal relationship’, ‘objectiveness’, and ‘exciting situations’. Meanwhile, for *re*, the variables comprise ‘emotion-involvement’, ‘objectiveness’, ‘sensory experience’, ‘visual’, ‘tactile’, ‘interpersonal description’, and ‘lively-scene description’. Finally, this chapter created cognitive spatial structure plots illustrating the category structures of *atsui* and *re* in the psychological spaces of native speakers.

Chapter 6 provides an integrated discussion to our empirical exploration of *atsui* and *re*. This chapter conducted a comparative analysis of the semantic relationships of *atsui* and *re* identified by the BP approach and psycholinguistic experimentation. Firstly, the consistency between the BP approach and production tests in identifying the prototypes of *atsui* and *re* showed that psycholinguistic experimentation corroborates the findings of the BP analysis. Secondly, statistical measures including cophenetic correlation and Baker's gamma index, combined with visual insights from the tanglegram and Bk plot, were employed to assess the similarity between the dendrograms derived from the BP analysis and the sorting task. The results indicated that for *atsui*, the dendrograms exhibited significant structural resemblances, whereas for *re*, the dendrograms are notably dissimilar. The chapter further explored the convergence and divergence between the BP approach and experimentation. In terms of variations in the results, this chapter identified potential factors such as limited sample sizes, individual differences, and method issues. To achieve

methodological consistency and converge empirical evidence, the chapter suggested enhancing the dataset, employing more fine-grained ID tags, refining stimuli selection, and conducting post-task interviews to reconcile divergent results.

Chapter 7 presents comprehensive conclusions to our empirical exploration of *atsui* and *re*. Initially, it offers a detailed discussion on how the findings from both corpus-based BP analyses and psycholinguistic experiments address the research questions outlined at the beginning of this dissertation. Specifically, Chapter 4 utilizes various corpus data to empirically identify the prototypes of *atsui* and *re*, addressing Research Question 1, and applies the BP approach to describe the semantic relationships between senses and determine the distinctive usage characteristics of each sense, addressing Research Question 3. Chapter 5 employs sentence production tests to identify prototypes, addressing Research Question 1, uses sentence sorting tasks to quantify semantic distances between senses, addressing Research Question 2, and illustrates the semantic structures and categorization variables of native speakers, Research Question 4. Chapter 6 undertakes a comparative analysis of the semantic relationships of *atsui* and *re* as identified through the BP approach and psycholinguistic experimentation, addressing Research Question 5. This chapter highlights the novel contributions our study makes to the field of cognitive lexical semantics. Also, it critically assesses the limitations encountered during our research. Lastly, this chapter proposes possible directions for future research to validate and complement the findings presented in this dissertation.

This empirical research project contributes to the current literature in the following ways: (i) This dissertation represents the first attempt to apply the BP approach to Japanese polysemy and synchronic Chinese polysemy. The ID tags and levels proposed in this study, along with the exploration of statistical techniques, can serve as a reference for future lexical analysis in these languages that seek to use the BP approach. (ii) The use of various online software tools has enhanced calculation speed and accuracy, and facilitated the conducting of remote and large-scale experiments. (iii) This dissertation employed R statistical software, integrating psycholinguistic experimentation with customized computational methods. Notably, this research presents the first attempt to integrate sorting tasks, HAC, and Average Silhouette Widths to verify sense identification results. (iv) The findings regarding the senses of *atsui* and *re* provide valuable references for Japanese and Chinese language teaching and lexicography. (v) By synthesizing findings from both the corpus-based BP approach and psycholinguistic experiments, this dissertation presents a pioneering comparative analysis of their convergence and divergence. The research methodology and findings from this integration offer significant insights and serve as a foundational reference for the field of cognitive research on polysemy.