

THE UNIVERSITY OF CHICAGO

COGNITION-DRIVEN DISTRIBUTIONAL AND DYNAMICAL MODELING OF
DOMAIN-ORIENTED LANGUAGE VARIATION AND EVOLUTION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

YU JI

CHICAGO, ILLINOIS

AUGUST 2023

Table of Contents

List of Figures	v
List of Tables	viii
Abstract	x
Acknowledgments.....	xii
Chapter I. Introduction.....	1
1.1. Overall Introduction.....	1
1.2. General research question and overall background	5
1.3. Overall preview of chapters	7
Chapter II: Typicality-contextuality model of basic-level categories (Model I)	12
2.1. Chapter introduction	12
2.2. Background.....	18
2.3. The "typicality-contextuality" embedding via hyperbolic coordinates.....	21
2.4. Empirical testing and applications in the domains of color and kinship semantics	28
2.5. Testing with a Mturk study indicates the measure-dependency of typicality	36
2.6. Comparison with basic-level color and kinship semantics in Modern Chinese.....	42
2.7. Further evaluations in the domains of place names and multiword expressions (MWEs)	48
2.8. Discussion	51
2.9. Chapter Summary	54

Chapter III. Applying the typicality-contextuality model to investigate the diachronic change of basic-level semantics of color, smell, and shape-based classifiers in ancient Chinese	56
3.1. Chapter introduction	56
3.2. Background	60
3.3. Methods.....	68
3.4. Results on basic-level basic color and smell semantics	71
3.5. Results on shape-based classifiers	80
3.6. Discussion	86
3.7. Chapter summary	89
Chapter IV: A singularity-synchronicity model of dynamic relationships between metacognition and semantic change (Model II)	91
4.1. Chapter Introduction	91
4.2. Background	93
4.3. A metacognitive framework of domain-oriented “conventionality” (synchronicity) and “expressiveness” (singularity).....	97
4.4. The dynamical-systems model of metacognition.....	100
4.5. Five metacognitive dynamical modes derived from the model	105
4.6. Chapter discussion and summary.....	112
Chapter V. Applying Model II to investigate how metacognition is relevant to semantic changes in color, kinship, and smell domains in written American English.....	115
5.1. Five empirical scenarios of meta-level and non-meta-level cognitive dynamics	115

5.2. Deriving the corpus-based empirical measures of singularity and synchronicity	121
5.3. Implementation of Model II on basic-level color semantics in American and British English from 1800 to 2020	124
5.4. Implementation of Model II on basic-level kinship semantics in American English from 1800 to 2020	137
5.5. Implementation of Model II on basic-level smell semantics in American English from 1800 to 2020	146
5.6. Discussion	150
5.7. Chapter summary	161
Chapter VI. Extending the Two Modeling Approaches to Other Semantic Domains	163
6.1. Chapter introduction	163
6.2. Methods.....	165
6.3. Plural suffixes of basic-level body parts, time, and length words (by Model I)	166
6.4. Pragmatic knowledge of country-naming (by Model I).....	175
6.5. Metacognitive awareness for semantic categories of political attitudes, religions, and professions (by Models I and II)	178
6.6. Discussion	192
6.7. Chapter summary	194
Chapter VII: Conclusion	196
References Cited	199
Appendix: Supplementary data of typicality scoring in Table 2.2 in Chapter II	215

List of Figures

- Figure 2.1. A geometrical representation of the typicality-contextuality mapping. 27
- Figure 2.2. Mapping basic-level color and kinship semantic categories in modern English. 32
- Figure 2.3. Mapping main basic kinship terms in English. 35
- Figure 2.4. Robustness testing of the training on poetry corpus. 36
- Figure 2.5. Mapping basic-level color and kinship semantics in standardized modern Chinese. 44
- Figure 2.6. Mapping shows a patrilineal character for kinship characters in modern Chinese. 46
- Figure 2.7. Mapping semantic relationships of country/city names and MWEs 49
- Figure 3.1. Mapping semantic changes and cultural dimensions by *word2vec* in Euclidean space 62
- Figure 3.2. Bifurcations of the Information bottleneck (IB) color categories. 63
- Figure 3.3. Mapping basic color terms in ancient Chinese across six periods. 75
- Figure 3.4. Mapping basic terms for fragrance in ancient Chinese across six periods. 78
- Figure 3.5. Mapping classifiers and shape-words in ancient Chinese across six periods. 83

Figure 3.6. Mapping classifiers and shape-words in modern Chinese.	86
Figure 3.7. Robustness checking on small corpus size for the ancient Chinese poetry corpus.	87
Figure 4.1. Diachronic evolution of word frequency ranks in English from 1700 to 2008.	95
Figure 4.2. The dynamical-systems mapping of five metacognitive modes.	110
Figure 5.1. The empirically-derived mapping of five dynamical scenarios.	116
Figure 5.2. Evolution of basic-level colors terms from 1800 to 2020.	125
Figure 5.3. Diachronic trajectories of basic-level color words in American English mapped in the singularity-asynchronicity space.	128
Figure 5.4. Diachronic trajectories of basic-level color words in British English mapped in the singularity-asynchronicity space.	133
Figure 5.5. Diachronic trajectories of basic-level kinship words in American English mapped in the singularity-asynchronicity space.	142
Figure 5.6. Diachronic trajectories of basic-level smell words in American English mapped in the singularity-asynchronicity space.	147
Figure 5.7. Theoretical optima of computationally partitioning the human perceptual color space to illustrate informative communication.	151
Figure 5.8. Modeling changes in the position of inflectional head (INFL) in Yiddish.	156
Figure 6.1. Typicality-contextuality mapping of basic-level body-part terms in English.	168

- Figure 6.2. Typicality-contextuality mapping of basic-level body-part terms in Chinese. 170
- Figure 6.3. Typicality-contextuality mapping singular and plural forms of basic-level time and length terms in English. 172
- Figure 6.4. Typicality-contextuality mapping of 66 country names in English. 177
- Figure 6.5. Singularity-asynchronicity mapping evolution of English political attitude adjectives. 180
- Figure 6.6. Singularity-asynchronicity mapping evolution of seven religion names. 183
- Figure 6.7. Typicality-contextuality mapping of seven religion terms. 187
- Figure 6.8. Singularity-asynchronicity mapping evolution of profession names. 189
- Figure 6.9. Typicality-contextuality mapping of profession names. 191

List of Tables

Table 2.1. Summary of the broad discoveries by Model I	17
Table 2.2. Semantic similarity extracted by Model I and by word embedding	29
Table 2.3. Comparing predictions on the typicality of main color and kinship words in English made by the model and by other theoretical and empirical works in the literature.	39
Table 3.1. Predictions on diachronic change of color terms in ancient Chinese in historical linguistics	72
Table 3.2. Ancient Chinese color terms of multiple periods extracted by Model I	74
Table 3.3. Three groups of ancient Chinese shape-based classifiers extracted by Model I	80
Table 4.1. Components of the metacognitive model	102
Table 5.1. Dynamical analysis of basic-level color terms in American English from 1820 -1920 and from 1920 to 2020	128
Table 5.2. Dynamical analysis of basic-level color terms in British English from 1820 -1920 and from 1920 to 2020	135
Table 5.3. Dynamical analysis of basic-level kinship terms in American English from 1800 to 2020	139

Table 5.4. Dynamical analysis of basic-level kinship terms in American English from 1820 -
1920 and from 1920 to 2020 144

Table 5.5. Dynamical analysis of basic-level color terms in American English from 1820 -1920
and from 1920 to 2020 149

Table A.1 215

Table A.2 216

Abstract

This dissertation develops a cognition-driven computational approach to investigate variation and historical change of basic-level semantic categories. While prior research on language change has started to incorporate cognitive factors influencing large-scale language dynamics, the two primary approaches—distributional and dynamical models—have been pursued predominantly independently, leading to different research focal points and frequently incompatible predictions. This dissertation proposes a cohesive approach by creating both a distributional and a dynamical model to examine how semantic cognition and semantic evolution may reciprocally interact, by shaping semantic knowledge in the domains of color, kinship, smell, and shape.

In the first computational chapter (Chapter II), I develop a two-dimensional embedding model that maps semantic typicality and contextuality of basic-level categories in hyperbolic coordinates, named the "typicality-contextuality" model (Model I). This model is predicated on the conditional entropy of word co-occurrence weighted by collocation diversity and conditional probability ranking. Model I is empirically evaluated in a domain-specific manner. The model examines the interplay between semantic cognition and relatively long-term (hundreds to thousands of years) semantic shifts in literary ancient Chinese in the domains of color, smell, and shape-based classifiers (Chapter III).

In the second computational chapter (Chapter IV), I present the "asynchronicity-singularity" model (Model II), a two-dimensional metacognitive dynamical-systems framework designed to investigate the potential interaction between metacognitive feedback on semantic knowledge and

diachronic shifts of semantic categories. This model is applied to diachronic change data from Google Ngram in modern written American English, exploring the interplay between semantic cognition and relatively shorter-term (decades to hundreds of years) semantic changes in the realms of color, kinship, and smell (Chapter V). In Chapter VI, the two models are further extended to include other semantic or linguistic domains such as plural suffixes, pragmatic knowledge of place-naming, and metacognitive awareness of semantic categories within political attitudes, religions, and professions.

Together, this dissertation underscores the importance of linking cognitive mechanisms of semantic categorization with large-scale socio-historical language variations and shifts, and of evaluating such linkages with domain-oriented focuses. More broadly, this work contributes to the understanding of how language cognition and language change intersect with cognition, culture, and society, a key research topic in many branches of contemporary cognitive, computational, and social sciences.

Acknowledgments

The journey towards completing my PhD at the University of Chicago has been an intricate tapestry of academic rigor, personal growth, and societal evolution. The global pandemic has defined the context of this journey, bringing a tragic personal toll with the loss of friends and family, and serving as a poignant backdrop to my scholarly pursuits. The interdisciplinary community both within and outside the University provided a robust intellectual environment that fostered dynamic dialogues, collaborations, and a shared resilience in the face of these adversities. My research endeavor, much like a strenuous hike, necessitated an exertion of physical and mental stamina, with the reward being an enriched understanding of important scientific questions and a profound sense of achievement. Concurrently, the exhilarating advancements in artificial intelligence and large language models have offered a glimpse into the potential and complexities of the scientific frontier, serving both as an inspiration and a reminder of the challenges they pose to our societal fabric. In writing this, I acknowledge these complex threads that have woven my academic journey, shaping it into a personal narrative of resilience, discovery, and ongoing learning.

My acknowledgments begin with my remarkable committee members. I am very fortunate to have Marc Berman serve on both my trial research committee and dissertation committee. With his characteristic big smiles, Marc often raises profound questions on the very foundations of my research, prompting a deeper understanding of my own work. I am also grateful to Allyson Ettinger, whose interest in understanding language through both the mind and computational modeling mirrored my own. Her challenging questions, ranging from overarching

research inquiries to minute details of testing and evaluating, continually spurred me to refine my thinking and my work. The origin of my scientific journey with this dissertation work traces back to an invaluable internship with Terry Regier at UC Berkeley, who introduced me to the fascinating intersection of perception and language through language modeling, and I owe a big thanks to him. I also owe a debt of gratitude to Howard Nusbaum, who vigorously challenged some very foundational aspects of my work and presented alternatives that greatly shaped my research trajectory.

Within the university, Haun Saussy, with his insatiable intellectual curiosity and extensive knowledge across fields such as literary studies, linguistics, psychology, and philosophy, aided me in drawing the link between scientific and humanistic understanding of language. John Goldsmith's generous allocation of hours for discussing research and personal development, coupled with his intellectual rigor required for investigating linguistic problems, have been instrumental in improving my thinking on linguistic questions. I am thankful to John Lucy, whose insights into language and thought significantly influenced my academic trajectory in the early stages of my PhD. Daniel Casasanto, with his rich teachings on language, body, and perception, welcomed me into his lab meetings and activities, and enriched my intellectual experience at the University. Michael Silverstein (1945 - 2020), Susan Gal, and Costas Nakassis introduced me to the exciting field of linguistic anthropology, opening a new path for me to connect computational modeling on language with understanding language in culture and society. I remember Michael, who sadly passed away from brain cancer in 2020, with deep respect and regret that he could not witness the fruition of this research, which was still at its very early stage back then. Several other faculty members from various disciplines and

departments, including Steven Shevell, Brian Prendergast, Yuan Chang Leong, Manyuan Long, Chris Kennedy, James Robinson, Jennifer Scappettone, Edward Shaughnessy, Rosanna Warren, and Dali L. Yang, have been also part of this intellectual journey, and for their generosity and contributions, I am also very thankful.

In my lab and department, as well as across the university, I have been fortunate to be surrounded by a vibrant group of friends and an intellectual community. I thank Boleslaw Osinski, who dedicated countless hours when I was a new PhD student to guiding me through neural modeling and engaging me in illuminating discussions about both science and life. I am also thankful for Andrew Sheriff, Rui He, and Amritpal Singh, who were great companions in a myriad of discussions spanning science, music, and a host of other topics. I extend my gratitude to Geoffrey Brookshire and Defu Yap, who lent their invaluable guidance on cognitive psychology and provided much-needed support during my initial years in the doctoral program. My research and academic growth have also benefited from Donald Frederick, Daisy Huibo Li, Shane Peace, Nisarg Mehta, Vivian Nguyen, Dana Beach, Abigail Stuart, Zac Leveroni, Ché Lucero, Tom Gijssels, Benjamin Pitt, and Andrew Stier, for which I acknowledge. I also would like to express my gratitude to Mimi Maduff, Kristi Schonwald, Anissha Warner, Mike McCarthy, Janet Capodagli, among other staff members at the Institute of Biology and Mind and the Department of Psychology - no academic progress is possible without their work and dedication.

I am indebted to the dynamic East Asian community at Hyde Park and across Chicago. Our widely interdisciplinary interests, ranging from Hegelian philosophy to ancient Chinese

classics, and joint participation in social activities have created an environment conducive to my personal and professional growth. This community, which includes Jianghong An, Jiun-Sheng Chen, Venerable Cetovimutti Cong (1989 - 2023), Danica Cao, Siying Cao, Xiaoyu Gao, Ji Gao, Yanxiao He, Jue Hou, Chiayu Fiana Hsu, Tong Ju, Hailu Kong, Xinyi Li, Zihao Lin, Biying Ling, Wan-Zi Lu, Jiarui Sun, Yenana Sun, Yang Shen, Zhenzhou Tan, Yuan Tian, Lucas Tse, Fanmei Xia, Chao Wang, Han Wang, Tingfeng Yan, Yuqian Yan, Yuchen Yang, Yukun Zeng, Wenxu Zhang, Houwei Zhang, Xinyue Zhang, Sam Zheyang Zheng, Boqun Zhou, Zhuowen Zhang, Yuhao Zhuang, Shuting Zhuang, Su-Chuan Lin, Xi-San Lai, among others, has been an immense source of comfort and inspiration, especially when we were all so far away from home. I also owe a great deal to my old friends, Ruoji Zhou and Dan Gilles, with whom I was able to reconnect in Chicago after nearly a decade. Special thanks go to Tom Chen and Winnie Wang, who serve as exemplars in balancing family and intellectual life, and Chen Liu, a dear friend since our high school years, who was there when I first stumbled upon Wittgenstein's philosophy of language when I was still a teenager. Over time, she saw my initial fascination evolve, from raw curiosity to the sophisticated, intricate computational models that I now build in my work. I would also like to extend my heartfelt thanks to David Bolotin, Susu Knight, Peter Pesic, Philip LeCuyer, and John Cornell, my cherished teachers and friends at St. John's College. Their constant support has been a critical part of my intellectual and personal growth well beyond my time at college.

A special recognition is due to Leslie Kay, my advisor, whose influence has left an indelible imprint on my academic and personal life. Leslie, your steadfast guidance has illuminated my path, leading me from an eager yet innocent PhD student to a researcher standing

at the threshold of exciting scientific discoveries. Your patience, as we dissected intricate scientific inquiries together, has taught me the value of endurance in the pursuit of knowledge. You have generously gifted me a free intellectual playground where my curiosity was not only welcomed but encouraged to thrive, which has been instrumental for my growth and has made the realization of this work possible. Your unwavering commitment to mentorship, education, and academic justice has not only shaped my understanding of science but also cultivated my empathy and sense of responsibility towards the broader community. This journey, arduous at times, yet profoundly rewarding, would not have been possible without you. It is under your tutelage that I have grown both intellectually and personally, and for that, I am forever grateful.

With warmth and love I remember my two grandfathers, Qi Hong and Guoliang Fei, who lived their long and full lives extending well into their 90s, but passed away, respectively, in the first and last years during my PhD journey. Their lives, rich with personal and historical experiences, were a beacon of resilience for me. My journey to earn a PhD, while intellectually rewarding, meant fewer moments shared with them at home—a sacrifice that I carry with a heavy heart.

My beloved wife, Lixin Zheng, along with her parents Zheng Yelong and Li Chenxia, deserves special mention. Lixin, you have been so much more than a partner to me throughout this journey. You have been my confidante, my anchor, and at times, my wisest counselor. When you moved to Chicago at the outset of my PhD journey, it marked the beginning of our new shared adventure. Our union during these formative years of my PhD has become a cherished touchstone, a blessing that reverberates through every facet of my life. Your keen intellect,

extraordinary generosity, vibrant energy, and deep-rooted kindness have not only provided a reliable pillar of support but have been instrumental in sculpting the scholar, and indeed the person, I am today.

Lastly, I wholeheartedly dedicate this dissertation to my parents, Yun Fei and Qi Hu. Despite the vast physical distance between us, your influence and love have always been close to my heart. The sacrifice you have made, allowing me to pursue a PhD in a far-off land, speaks volumes of your unconditional love and unflinching support. In the face of any challenge or difficulty, I have only to think of you as my role models to find the strength and inspiration I need. Each of my accomplishments, this dissertation included, is a testament to the invaluable life lessons and nurturing guidance you have provided. I am incredibly privileged and eternally grateful to be your son.

We feel as if we had to see right into phenomena: yet our investigation is directed not towards phenomena, but rather, as one might say, towards the "*possibilities*" of phenomena.

Ludwig Wittgenstein (90, *Philosophical Investigations*)

In psychology, there are experimental methods and *conceptual confusion*. [...] The existence of the experimental methods makes us think that we have the means of getting rid of the problems which trouble us; but problem and method pass one another by.

Ludwig Wittgenstein (371, *Philosophy of Psychology*)

Chapter I. Introduction

1.1. Overall Introduction

Recent developments in natural-language semantic modeling have been significantly shaped by large-scale statistical and distributional methodologies utilized in computational linguistics and natural language processing (NLP). This is notably seen in the use of Deep Neural Networks (DNNs), exemplified by models such as BERT (Devlin et al., 2018) and GPT-3 (Brown et al., 2020). Despite various DNN models being employed to model and test language cognition hypotheses (Chersoni et al., 2021; Wu & Ettinger, 2021), the design, training, and testing of these models are generally not driven by cognitive principles. These models necessitate vast quantities of textual data, often in the range of billions of word tokens. Furthermore, recent research indicates that these statistical and DNN models may not be effective in modeling

various linguistic phenomena that are extensively studied in cognitive science and other linguistics subfields (*e.g.*, Misra et al., 2021).

Arguably, language cognition and processing by humans are still far more effective and efficient compared to the most advanced language models in computational sciences (*e.g.*, Ouyang et al., 2022). Human language learning and linguistic and symbolic capacities utilize significantly smaller amounts of linguistic input, yet demonstrate remarkable flexibility and interpretative capacity (Geertz, 1973). Consequently, for some researchers, an essential future research direction in language model development should be the incorporation of insights gleaned from human language learning and processing. The aim of language modeling should be to create methodologies and models that are increasingly driven and inspired by cognitive principles (*e.g.*, Ettinger et al., 2016).

On the other hand, cognitive scientists typically adopt a different methodology to modeling linguistic phenomena, viewing the relationship between language and cognition as foundational. Many cognitive scientists further consider the fundamental role that non-linguistic cognition, such as sensation and perception, plays in "grounding" language cognition (Barsalou, 2008). The central questions guiding this perspective, which seeks to intertwine language, perception, and cognition, can be summarized as follows: How do language, cognition, and perception interact both cognitively and culturally? How can this interaction be characterized computationally or otherwise? And what mechanisms underlie these interactions?

Research across various disciplines has presented diverse frameworks and methods to tackle these questions. Neuroscientists tend to concentrate on specific physiological and

computational attributes of stimulus or perceptual spaces as well as neurophysiological connections between sensory and linguistic circuits in the brain. Psychologists and anthropologists, however, underscore the significance of the diversity of perceptual experiences in both the individual's interaction with the environment and the culture in which individuals live and use language. For example, recent research in the realm of the language of smell underscores tensions between these perspectives. It is generally accepted that humans in many societies struggle with identifying and discriminating odors linguistically, with the vocabulary for describing olfactory experiences often being scarce or imprecise (particularly compared to vocabulary for describing the visual world). Olofsson & Gottfried (2015) argue that this observation has a hardwired biological foundation, pointing to the sparse connectivity between the olfactory and language areas in the human brain. Yet, empirical studies in psychology and anthropology focusing on individual indigenous languages suggest that language pertaining to smell can be rich and precise in specific cultures and experiences. Speakers of languages with rich vocabulary for smell and rich cultural needs for using such vocabulary are also usually more proficient at identifying odors than those speaking modern and predominantly western languages (Majid & Burenhult, 2014; Majid et al., 2018).

The problem concerning the language of smell is not unique. The above cognitive studies on the relationship between language, cognition, and perception share a fundamental assumption: human linguistic processing is intrinsically cultural and embodied. For researchers who share this basic assumption, it is impossible to productively explore language and linguistic processing without considering the sensory and perceptual domains (Barsalou, 2008) or the sociocultural aspects (Geertz, 1973; Silverstein, 2022). However, many researchers in psychological and

cognitive sciences do not conceive perceptual domains, sensory embodiment, or sociocultural factors as fundamental aspects for inquiring about language. For these researchers, language is a unique and self-sufficient cognitive or computational module, and perceptual, bodily, sociocultural factors only influence language, but do not play a determining role about what language meaning or language cognition fundamentally is (Chomsky, 1966; Langacker, 2008; Poeppel & Monahan, 2008).

Hence, one fundamental challenge in studying the relationship between language, perception, and cognition is reconciling differing conceptions of these relationships across various disciplines such as computational sciences, cognitive psychology, neuroscience, linguistics, and anthropology (Goldsmith & Laks, 2019). An additional challenge lies in resolving the tensions between language modeling in contemporary mainstream NLP, which typically does not focus on cognitive, perceptual, or sociocultural questions in any direct way, and language modeling in cognitive science and social sciences, which instead typically regards language as fundamentally grounded in cognition, bodily sensation and perception, or society and culture.

In recent years, there has been a growing research trend towards integrating insights from NLP into the modeling of human language cognition and language processing in the brain (Tikochinski et al., 2021; Goldstein et al., 2022; Kumar et al., 2022; Antonello & Huth, 2022). Moreover, the use of human language processing and cognition as a means to inform and enhance NLP language models has also gained traction (Ettinger et al., 2020; Ouyang et al., 2022). Yet, for both cognition- and computation-driven language models, there are substantial

opportunities and challenges concerning the applicability and relevance of these models to human psychological, cultural, and social realities at large (Hovy & Spruit, 2016; Tamkin et al., 2021). The work presented in this dissertation represents my attempt to respond to these opportunities and challenges.

1.2. General research question and overall background

The primary objective of this dissertation is to establish two new semantic modeling methodologies, both of which are firmly anchored in cognitive science and offer robust interpretability in both cognitive and computational contexts. This dissertation, nonetheless, draws motivations from two largely separate fields of literature: one being domain-specific cognitive research on basic-level semantic categories, and the other being domain-generic language modeling research with dynamic strategies. Importantly, these two literature groups propose distinct micro-level cognitive mechanisms that could potentially influence large-scale language variation and change. Cognitive mechanisms proposed by the first set of literature include domain-specific efficiency in communication, which predicts domain-specific patterns of language change, such as the evolution of basic-level color semantics (*e.g.*, Regier et al., 2015). Cognitive mechanisms developed from the second set of literature, however, usually presuppose domain-general mechanisms such as cognitive optimization in language learning and processing, modeling phenomena like diachronic changes of domain-generic morphological and syntactic properties (*e.g.*, Niyogi, 2008). Additionally, the general research scopes and methodologies of

both sets of literature remarkably diverge from the statistical learning and distributional learning paradigm, which is most commonly adopted in contemporary NLP-driven language modeling.

In response to these two sets of literature and their tensions with the statistical learning framework in NLP, this dissertation proposes two models and two main groups of applications. The first model integrates a domain-focused cognitive approach with the statistical-learning computational approach, developing a domain-oriented distributional semantic embedding. The second model develops a dynamical-systems model of domain-oriented language change that hinges on the metacognitive awareness of the change, an underexplored cognitive mechanism in dynamical and cognitive modeling literature as well as in the literature of language change modeling. Notably, the primary evaluations and applications of both models are geared towards integrating the modeling of domain-oriented diachronic patterns of semantic categories with the modeling of semantic cognition.

The general methodological motivation of this dissertation is as follows: operating on broad social, cultural, and historical scales and contexts, cognitive mechanisms are crucial drivers of language variation and change, and elucidating the cognitive underpinnings of language variation and change can reciprocally provide insights for understanding language cognition. Historical linguistics and sociolinguistics have long identified language variation and change as a significant area of linguistic research. In the increasingly digitally-networked age, large-scale corpora across multiple languages have become readily available to researchers. These extensive datasets offer unprecedented opportunities to evaluate and test existing hypotheses about language cognition and language variation as well as generate novel ones,

particularly in research areas where traditional experimental methods in cognitive psychology and other social sciences face empirical and practical limits or constraints.

In general, this dissertation aims to develop novel cognitive-driven and domain-oriented semantic models to investigate and characterize how cognition may drive large-scale synchronic and diachronic variations of semantic categories, and how modeling macro-level variations and changes of semantic patterns can shed light on semantic cognition. Studies in this dissertation are also accompanied by new empirical frameworks for testing and evaluation, implemented across diverse domains and types of corpora spanning various socio-cultural histories.

1.3. Overall preview of chapters

Chapter II introduces a novel semantic embedding model (Model I) mapping basic-level categories onto two cognitively interpretable dimensions, which I term typicality and contextuality. In cognitive psychology, quantifying typicality empirically or computationally remains a significant challenge, with no single measure proven to be definite or superior. In contrast, contextuality is a fundamental assumption in much current NLP research and cognitive science, with word co-occurrence matrices and high-dimensional vectors typically used to develop algorithms or capture linguistic properties and relationships. However, typical methods based on distributional contextuality have notable constraints, such as thin representational capacity and compromised interpretability (Tenenbaum, 2011; Steyvers & Tenenbaum, 2005; Sajjad et al., 2022). The "typicality-contextuality" embedding model, on the other hand, captures

both hierarchical ("typicality" versus "idiosyncrasy") and similarity ("contextual dominancy" and "contextual dependency)" relations of words and semantic categories within a simple metric in two-dimensional hyperbolic space, offering an innovative way to embed and interpret multiple cognitive-linguistic dimensions at the same time. Evaluation in this chapter focuses on the domains of basic-level color and kinship semantics, and further testing is performed by a Mturk study as well as comparisons with empirical benchmarks on semantic typicality judgment in the literature.

Chapter III applies Model I to basic-level semantic categories in the domains of color, smell, and shaped-based classifiers over a span of more than 1500 years in ancient Chinese history. Historical competitions between multiple basic-color domains in ancient Chinese are examined by the typicality-contextuality mapping. Results indicate that, different from basic-level color semantics in modern English and Chinese, which predominantly align along the "typicality-idiosyncrasy" axis, historical competitions within basic-level color semantics mainly occur in the axis of contextuality. Further, five basic terms for pleasant smells are identified to stably overlap with several main color words throughout ancient Chinese history, indicating that the language of smell in a large portion of ancient Chinese history is stable yet dependent on the domain of color. Lastly, the study in this chapter also investigates a classical question in Chinese historical linguistics: the evolution of shape-based classifiers in ancient Chinese (Chao, 1966; Lucy & Gaskin, 2001). The typicality-contextuality mapping clusters ancient Chinese classifiers into three groups and indicates that classifiers mainly generated from one or two dimensional shapes are increasingly distributed in the contextuality dimension with high typicality, whereas

classifiers mainly evolved from plant and utensil shapes increasingly align over time along the typicality-idiosyncrasy dimension with wider distributions of idiosyncrasy.

While the first computational model is primarily for proposing a new metric and topological method to map domain-oriented semantic properties, Chapter IV proposes a causal and dynamical model (Model II) designed to characterize how metacognition (Yeung & Summerfield, 2012; Bruno et al., 2023; Fleming & Daw, 2017) of diachronic semantic patterns interacts with collective diachronic changes of basic-level semantic categories, and how these changes also reciprocally shape cognitive knowledge of semantic categorization. The foundation of this model rests on the hypothesis that the metacognition of semantic change is a crucial force steering the actual semantic change, which in turn provides cognitive feedback to metacognition. Two essential linguistic measures, "singularity," which measures linguistic "expressiveness," and "synchronicity," which measures linguistic "conventionality," are defined as two basic variables. A dynamical-systems model based on these two variables is developed, generating different cognitive scenarios which can be applied to diachronic corpus data.

Chapter V begins by identifying five empirical scenarios of metacognitive dynamics emerging from actual diachronic data. Then relating them with the theoretically identified dynamical scenarios of metacognition developed by Model II in Chapter IV. Three domains of basic-level semantic categories are examined through the lens of the American English corpus from the last 200 years as extrapolated from the Google Ngram Viewer. The analysis aims to simulate how metacognition could interact with basic-level semantic shifts within these domains based on real diachronic corpus data. In the domain of color semantics, the analysis based on

simulations indicates the relevance of the metacognitive awareness of the sociocultural history of race to the semantics of several basic-level color terms in American English, but less so in British English. In the domain of basic-level kinship semantics, multiple metacognitive scenarios are captured, and dynamic interactions between diachronic changes of sociocultural connotations of kinship terms with kinship semantics in American English are considered. Within the realm of smell words, the analysis raises the possibility of a potential overstability of basic-level smell semantics in the sociocultural history of American English. The chapter also embarks on a comparative discussion highlighting the advantages of the metacognition-based model over other dynamics-driven approaches, and further explores potential conceptual and formal connections between dynamical approaches to neurobiological systems and to sociocultural phenomena.

Chapter VI further demonstrates the capacity of the two models presented in this dissertation as integrative tools for generating hypotheses for investigating how distributional and dynamic properties of other semantic or task domains are cognitively significant. In the first study of this chapter, Model I is employed to examine plural suffixes in English, illustrating its potential to map the cognitive significance of plural markers in semantic domains like body parts, time, and length in American English. In the second study, Model I is leveraged to scrutinize pragmatic knowledge within the context of place names, suggesting that corpus-based categorical knowledge of place naming is also pragmatically salient. In the third study of this chapter, both Models I and II are employed to assess the metacognitive sensitivity of various semantic domains or categories (political attitudes, religions, and professions) based on corpus-based evidence of language change, extending this methodology to domains rich in pragmatic knowledge and metacognition. Collectively, this chapter underscores the prowess and potential

of these two models in dissecting a wide range of semantic domains beyond color, kinship, smell, and shape, stimulating further explorations and applications in domain-oriented approaches of studying semantic categories for future computational and empirical research in cognitive science.

Chapter II: Typicality-contextuality model of basic-level categories (Model I)

2.1. Chapter introduction

This chapter introduces a novel cognition-inspired model of basic-level categories, centered around the generation of rich and interpretable semantic representations at the word-embedding level. The model encompasses two fundamental semantic dimensions: *typicality* and *contextuality*.

The rationale behind selecting these two cognitive variables as core components stems from cognitive psychology. Typicality (or prototypicality) has been the subject of rigorous and extensive research on semantic categorical knowledge since the 1970s (Rosch et al., 1976). However, quantifying typicality remains a challenge both theoretically and empirically. Although the frequency of a single word in a corpus or its exposure appears to be an obvious metric for measuring typicality, it has proven to correlate poorly with semantic typicality (Uyeda & Mandler, 1980). A variety of other measures have been proposed, including word production frequency, rankings from elicitation tasks, and age of acquisition (Uyeda & Mandler, 1980; Plant et al., 2011; Banks & Connell, 2022). However, there is currently no consensus in the literature supporting the superiority of one measure over others, either theoretically or empirically.

In the realm of natural language processing (NLP), Kanishka et al. (2021) demonstrated the limited effectiveness of existing deep learning algorithms in distilling typicality from a

corpus. They contended that corpus data alone might not be sufficient to model typicality accurately. However, their study's validation was primarily based on one particular cognitive rating methodology (Rosch, 1975; Rosch et al., 1976); other studies using alternative methods generate different results on typicality judgment (Holmes & Ellis, 2007; see Table 2.3 for detailed comparisons between different studies on typicality).

Corpus-based contextuality, on the other hand, is a core basic assumption for contemporary NLP research, which treats distributional statistical patterns of word co-occurrence in corpuses or other types of language inputs as foundational for generating proper measures of contextuality. In cognitive science, distributional and contextual word co-occurrence is also a fundamental assumption of the statistical learning theory of natural language (Tenenbaum, 2011; Steyvers & Tenenbaum, 2005). Psycholinguists have demonstrated that infants have the capacity to learn certain aspects of phonology, syntax, and semantics of natural language from statistical patterns of sound or word co-occurrence in the language input from the environment (Aslin et al., 1998; Saffran, 2003).

A pivotal research enquiry within the domains of computational linguistics and cognitive science encompasses the generation of proper measures and interpretable representations to capture the linguistic attributes of semantic categories. Though the measurement of typicality incites a heterogeneous debate in both cognitive science and computational linguistics, the practice of measuring contextuality through word co-occurrence matrices and its representation via high-dimensional vectors is a prominent protocol for developing algorithms, such as in classical work of latent semantic analysis (*e.g.*, Dumais, 2005) and of word embedding (Mikolov

et al., 2013; Pennington et al., 2014; Brown et al., 2020). Almost simultaneously, cognitive scientists have conceived methodologies to encapsulate the statistical leaning of contextual co-occurrence within high-dimensional vector spaces (Steyvers & Tenenbaum, 2005; Jackson & Bolger, 2014). Further, recent scholarly contributions demonstrate promising possibilities relating these two research strands. For instance, Ettinger et al. (2016) leverages word embedding algorithms originating from NLP in the field of psycholinguistics, which demonstrates immense potential for fostering NLP-inspired statistical learning models to enhance research of language acquisition and processing in cognitive psychology.

Conventional statistical embedding frameworks, however, exhibit several notable constraints. The first pertains to thin representational capacity: traditional embedding algorithms typically generate linguistic representations within a 2-D Euclidean space, where word relations are entirely predicated on *similarity* relations between vectors (Mikolov et al., 2013; Pennington et al., 2014; Brown et al., 2020). More recent developments have utilized alternative embedding spaces, such as hyperbolic space, to construct representations for *hierarchical* relations among words (Nickel & Kiela, 2017; Sala et al., 2018; Dhingra et al., 2018). Yet, the features that can be captured by the Euclidean vector space frequently diminish within the alternative hyperbolic embedding space. Additionally, there lacks a robust methodology to discern how linguistic features captured by each of these embedding spaces are related to or correlate with one another¹.

¹ In the NLP literature, hyperbolic embedding proposed by Nickel & Kiela (2017) can rank hypernyms and hyponyms under a particular baseline of WordNet (Fellbaum, 1998). But this approach cannot in principle rank typicality for words at the same hierarchical level (such as "red" and "green"), and the capacity of measuring semantic similarity is lower than standard word-vector embedding algorithms in the Euclidean space can usually achieve.

In the broader literature on hyperbolic embedding, there have been recent proposals for embeddings that aim to encapsulate two dimensions (such as hierarchy and similarity) and even multiple dimensions (Jiang et al., 2022; Kovacs et al., 2022). However, these works seldom establish any realistic linguistic or cognitive relevance.

The second limitation involves compromised or opaque interpretability. Interpretation primarily relies on baselines established from a specific set of research literature, which usually assumes a confined interpretive and evaluative awareness among researchers within a single discipline (Steyvers & Tenenbaum, 2005; Sajjad et al., 2022). Particularly, a significant portion of these interpretive and evaluative baselines are predominantly semantics-oriented. Pragmatic dimensions, notably those encompassing social, cultural, and historical characteristics of texts and contexts, are usually undervalued in the interpretive and evaluative processes (Wagner et al., 2015; Tanweer et al., 2021; Nguyen et al., 2021).

In this chapter, I propose a "typicality-contextuality" embedding model. The significance of the model is as follows. Firstly, the model demonstrates an expansive and integrated representational capability, capturing *both* the hierarchical relation (via typicality) *and* similarity relation (via contextuality) of words and semantic categories. Crucially, the model produces these two fundamental linguistic dimensions through the same metric, resulting in their co-representation within the same embedding space. This yields a novel methodology for investigating the relationally interpretive and evaluative aspects of multiple linguistic dimensions, in ways that are well computationally grounded and motivated. Secondly, the model exhibits transparent interpretability that is firmly grounded in cognitive principles: the two

embedding dimensions are cognitively constructed and interpreted with transparent cognitive meanings generated from the cognitive science literature, and the evaluation and interpretation processes are anchored in related work in both NLP and cognitive science. Thirdly, the model exhibits transparent interpretability that allows feasible empirical mapping between algorithm and performance. Many hyperbolic embedding works in the literature either require sophisticated hyperbolic representation, such as the Poincare-disk language embedding models (Nickel & Kiela, 2017; Sala et al., 2018; Dhingra et al., 2018), or contain cognitively opaque procedures of optimization, such as the popularity-similarity optimization (PSO) model (which is proposed in network theory for generalized hyperbolic networks: Papadopoulos et al., 2012; Kovacs et al., 2022). This model, instead, converts the embedding from a two-dimensional Euclidean space into straightforward two-dimensional hyperbolic coordinates, and this transformation utilizes solely cognitively-motivated weighted functions and avoids any other computationally sophisticated or cognitively opaque optimization procedures.

The principal findings derived from testing and evaluating the typicality-contextuality embedding model are as follows. First, the results reveal that local statistical patterns of word co-occurrence (such as bigrams) tend to produce what psychologists understand as basic-level categories (for instance, in the semantic domain of color, "red" and "blue" are considered basic-level categories, whereas "azure" and "vertigo" are not). However, the degree to which these basic-level categories are encoded in local word co-occurrence patterns is shown to be specific to different semantic domains, such as color, kinship, and smell. Second, model evaluations suggest that typicality judgment varies across semantic domains and depends on the empirical measures utilized to elicit this judgment in actual experimental tasks. These evaluations are produced by

comparing the typicality mapping generated from the model with multiple empirical studies on typicality found in cognitive psychology literature, along with a Mechanical Turk study conducted for the purpose of evaluation. Third, the embedding approach indicates that typicality and what I term "idiosyncrasy" are highly correlated linguistic dimensions with opposite cognitive valences. This implies that the phenomenon of typicality and the phenomenon of multiword expressions (such as "Hong Kong" or "blue whale"), usually regarded as two largely independent research questions in cognitive psychology and semantic modeling, can be highly relevant to each other. Fourth, the embedding approach indicates that two contrasting linguistic valences of semantic contextuality, which I separately term as "contextual dominancy" and "contextual dependency," should be at least partially accounted for by pragmatic factors such as genre and register. If this observation proves accurate in further inquiry, it could facilitate the integration of sociolinguistic theories into semantic embedding interpretation, fostering the future development of language models that amalgamate both semantics and pragmatics.

Noticeably, all the aforementioned discoveries are derived from the same embedding structure and representation, underscoring the model's capacity of generating rich and interpretable representations of corpus-based semantic knowledge.

Main Contributions	Descriptions
(i). Relating domain specificity of basic-level categories with distributional linguistic patterns	The degree to which basic-level categories are encoded in local word co-occurrence patterns is shown to be specific to different semantic domains.
(ii). Demonstrating domain-sensitivity and measure-dependence of typicality	Typicality judgment varies across semantic domains and depends on the empirical measures utilized to elicit this

judgment	judgment in actual experimental tasks.
(iii). Demonstrating that typicality and idiosyncrasy are highly related linguistic phenomena with contrasting cognitive valencies	The phenomenon of typicality and the phenomenon of multiword expressions, usually regarded as two largely independent research questions in cognitive psychology and semantic modeling, can be highly relevant.
(iv). Demonstrating contrasting pragmatic characteristics of semantic contextuality	Two contrasting linguistic valencies of semantic contextuality, which I separately term as "contextual dominance" and "contextual dependency," should be at least partially accounted for by pragmatic factors such as genre, domain, and sociocultural dynamics.

Table 2.1. Summary of the broad discoveries from testing and evaluating the typicality-contextuality model in Chapter II and III.

2.2. Background

In this background section, I will further survey three pertinent bodies of literature and relate them together to my current study: criticisms regarding the "thin" representational nature of distributional language modeling methods in cognitive psychology, debates concerning the domain-specific and domain-generic nature of semantic knowledge in cognitive science, and tensions on employing distributional methods to incorporate pragmatic, social, and cultural contexts in contemporary NLP literature.

A substantial critique of distributional learning theory within cognitive psychology literature is the assertion that any statistical learning mechanism would be too "generic" to learn the full breadth of the human language repertoire, as the high-dimensional representation of

language knowledge is too "thin" to encapsulate the richness and complexity of human semantic knowledge (Steyvers & Tenenbaum, 2005). For many psychologists, semantic learning and cognition are also fundamentally grounded upon embodied sensation and perception (Barsalou, 2008), as well as societal and cultural contexts (Geertz, 1973, Silverstein, 2004). Recognizing these criticisms, I aim to demonstrate that cognitive-driven language models, under suitably tuned metrics and mapping strategies, can effectively offer more "thick" descriptions (Geertz, 1973) of semantic relationships than many current DN and cognitive models can manage to achieve.

A related debate on the nature of semantic knowledge concerns the extent to which semantic knowledge is specific to semantic or perceptual domains. Research on typicality serves as a representative research area for such discussions. Cognitive psychologists typically survey a large number of semantic categories without considering domain specificity (Rosch, 1975, Uyeda & Mandler, 1980, Plant et al., 2011, Banks & Connell, 2022). This method focuses on ranking the typicality of assigned semantic categories using a generic empirical approach. However, an alternative research method is domain-specific, which often categorizes semantic typicality according to the most typical semantic categories being considered as the basic-level ones (Berlin & Kay, 1969). Usually, the domain-specific approach focuses on semantic relationships within one or two particular domains, underpinned by rigorous theoretical foundations (*e.g.*, information bottleneck for basic-level color categories in Zaslavsky et al., (2019) and cross-language and cross-culture comparisons in Regier et al. (2015) for color words and Majid (2021) for smell words). The model developed in this chapter seeks to strike a balance between these two cognitive approaches: while differentiating and mapping typicality

relationships within some of the most well-studied domains of basic-level categories (such as color and kinship), the model does not presume a domain-specific position and can map typicality relationships in a cross-domain and generic fashion.

Also relatedly, a recent tendency in computational social science and digital humanities is to employ distributional modeling approaches and algorithms to explore the dynamics between language, culture, and society (Hamilton et al., 2016; Li et al., 2018; Koslowski et al., 2019). Despite their productivity, these works face many of the challenges inherent to standard distributional modeling algorithms, such as developing rich and interpretable representations of the dynamic relationship between language and society (See Section 3.2.1 in Chapter III for more detailed discussions on this literature). Furthermore, situating language within society and culture is deeply rooted in theoretical debates in cognitive science, linguistics, and philosophy of language since the 1950s, centering around the relationship and difference between semantics and pragmatics. From a strong semantic perspective, it is fundamental to investigate the core of language meaning as an internally coherent structural system, and pragmatic factors only add further complexity to the largely intrinsically self-sufficient system of semantic knowledge (Chomsky, 1966). Conversely, from a strong pragmatic perspective, it seems implausible to investigate semantic knowledge without taking into account the pragmatic dimension, as language use is considered to intrinsically determine language meaning (Silverstein, 1993). Whether explicitly or implicitly, the standard distributional methodology adopts a strong semantic perspective by focusing on co-occurrent contextual patterns in corpus or language inputs. Initial work in NLP in the 2010s that applies distributional algorithms to model the dynamics between language and society also assumes such a self-sufficiency for semantic

knowledge. However, some more recent work in NLP begins to take a more pragmatic perspective on language modeling, but this line of research has been too nascent to be productive as the first group (Tamkin et al. 2021). The model developed in this chapter aims to rigorously and effectively incorporate both the semantic and pragmatic perspectives into semantic modeling.

2.3. The "typicality-contextuality" embedding via hyperbolic coordinates

The "typicality-contextuality" embedding model introduced in this section provides a novel distributional method for measuring and generating a unified representation of both semantic *similarity* and *hierarchical* relations, derived from co-occurrence statistical patterns.

Below is the formal description of the model. Given a target word, the conditional entropy of the target word upon the occurrences of its context words is defined with two weighting parameters: the target word's collocation diversity and the ranking of the conditional entropy of each pair of word occurrences for every target word. Formally, let w_i be any target word and w_j be any context word, both of which belong to the whole vocabulary W of the corpus. Given the target word $w_{i=a}$, the collocation diversity d is equal to $count(w_j)/f(w_a)$, where $f(w_a)$ is the frequency of the target w_a in a given corpus or language input data, and $count(w_j)$ is the count of the type number of each context word w_j that co-occurs in a given length of the context window adjacent to the target word w_a . The second parameter r is equal to

the rank of conditional probability $f(w_{i=a}|w_{j=b})$ among all $f(w_{i=a}|w_j \in W)$, where $w_{i=a}$ is the target word, w_j is the context word of w_a , and W is the whole vocabulary. It is convenient to express d and r collectively as a single weighting parameter q (as seen in Equation 2.1). The development of weighting parameters d and r for the single-word conditional entropy is fundamentally influenced by the notion of a semantic network. In this framework, the understanding of word semantics is determined by the network properties of semantics themselves. From the viewpoint of network theory, collocation diversity is considered synonymous with the concept of link degree. This concept has been employed in prior semantic modeling work (Vincze, 2015). In addition, probability ranking has been shown to be related to the hierarchical and geometrical characteristics of complex networks (Krioukov, et al., 2010). From a computational standpoint, the use of collocation diversity and probability ranking as weighting parameters for embedding is equivalent to the construction of a co-occurrence-based semantic network weighted by node degree and link ranking. Section 2.3.3 offers further computational and cognitive justifications for these constructions.

The weighted entropy-based metric H for the embedding is defined as below:

$$\begin{aligned}
 H(w_{i=a}|w_{j=b}) &= -q_{ab}f(w_a, w_b) \log\left(q_{ab} \frac{f(w_a, w_b)}{f(w_b)}\right) \\
 &= -q_{ab}f(w_a, w_b) \log(q_{ab}f(w_a|w_b)) \tag{2.1}
 \end{aligned}$$

$$\begin{aligned}
 \text{where } q_{ab} &= d_{i=a} \frac{1}{(r_{i=a, j=b}+1)} \\
 &= \frac{\text{count}(w_j)}{f(w_a)} \cdot \frac{1}{1+\text{rank}[f(w_{i=a}|w_{j=b}) \text{ in } f(w_{i=a}|w_j \in W)]} \tag{2.2}
 \end{aligned}$$

Given a corpus, Equation (2.1) is the formula of the weighted conditional-entropy metric H for measuring the weighted co-occurrence relationship between the target word w_a and any context word w_b . Equation (2.2) is the formula of the weighting parameter q_{ab} , which is a linear function of d and r . Both w_a and w_b are extracted from the vocabulary of the corpus W .

The model then defines the distributional metric C for each target word $w_{i=a}$ as the expected value of a sigmoid function of H , characterized by the word co-occurrence relationships between the target word $w_{i=a}$ and its context. Let S be the sigmoid function of H , for a given target word $w_{i=a}$, $C(w_{i=a})$ can be expressed as the following form:

$$\begin{aligned} C(w_{i=a}) &= \langle S(H(w_{i=a}|w_j)) \rangle \\ &= \langle \frac{1}{1+\exp(-H(w_{i=a}|w_j))} \rangle \end{aligned} \quad (2.3)$$

Here, $C(w_{i=a})$ is the metric used for extracting the similarity and typicality measures for the target word $w_{i=a}$ and w_j is any contextual word. The weighted word co-occurrence matrix M^C for the whole vocabulary is defined by C , where each element of the matrix M^C is $C(w_i|w_j)$. In this study, the model also sets the threshold of minimum frequency for the target word w_i as 15 and the threshold of minimum frequency for the context word w_j as 5. In order to make simulations and testing both cognitively relevant and computationally convenient, the typical

corpus size used for the testing and evaluating of the model in this chapter is set to be around millions².

2.3.1. Generating similarity measures

Let $l(w_i)$ be the number of non-zero values of $C(w_i|w_j)$ for w_i . For any chosen target word w_a , let $l^{[w_a]}(w_i)$ = the number of non-zero values of $C(w_i|w_j)$, and let $C(w_a|w_j)$ be conditioned by every context word w_j that is commonly shared by the chosen target word w_a and any target word w_i from the whole vocabulary of W . The extraction strategy for producing the set $W_i^{[w_a]}$ in which w_i s are all linguistically closest to w_a is as follows:

$$W_i^{[w_a]} = \{w_{i \max}^{[w_a]}\},$$

$$\text{where } w_{i \max}^{[w_a]} = \operatorname{argmax}_{w_i}(L^{[w_a]}(w_i)), \text{ in which } L^{[w_a]}(w_i) = \frac{2 \log l^{[w_a]}(w_i)}{\log l(w_i) + \log l(w_a)}$$

(2.4)

The set of $W_i^{[w_a]}$ that is produced from the procedure in Equation (2.4) is the set of words that are determined to be semantically closest to w_a , where the similarity ranking between each w_i and w_a is determined by $L^{[w_a]}(w_i)$. Empirically, it is demonstrated that using the medium-range context size (*e.g.*, around 5 word tokens) can optimize the general performance of the procedure

² According to research in developmental psychology (Hart & Risley, 1995; Schneidman et al., 2013), it is estimated that a child hears over 7 million words per year. To approximate similar semantic input conditions as experienced by humans, the modeling work in this dissertation intentionally utilizes corpora with token sizes in a similar scale.

described in Equation (4.1) for the purpose of extracting semantic similarity, but testing also shows that optimal context size is also sensitive to genre, domain, language, as well as individual goals of semantic modeling (see *Results* section in the chapter and the ancient Chinese poetry corpus modeling in the next chapter).

2.3.2. Four linguistic features captured in the two-dimensional hyperbolic embedding

In a two-dimensional Euclidean space, for each target word $w_{i=a}$, let

$$\begin{aligned}
 x &= \left\langle \frac{1}{1 + \exp(-H(w_{i=a}|w_j))} \right\rangle \\
 &= \langle S(H(w_i|w_{i=a})) \rangle ; \\
 y &= \left\langle \frac{1}{1 + \exp(-H(w_{j=a}|w_j))} \right\rangle \\
 &= \langle S(H(w_j|w_{i=a})) \rangle . \quad (2.5)
 \end{aligned}$$

In Equation (2.5), the x -axis is the expected value of the sigmoid function of each target word's weighted conditional entropy, and the y -axis is the expected value of the sigmoid function of the context's weighted entropy conditioned on the target word w_i . Euclidean coordinates (x, y) are then transformed into the hyperbolic coordinates (u, v) , where

$$\begin{aligned}
 u &= \ln \sqrt{\frac{x}{y}}, \\
 v &= \sqrt{xy}. \quad (2.6)
 \end{aligned}$$

Empirically, it is demonstrated that for words that are extracted through the procedure shown in equation (2.4), using the local-range context (*e.g.*, bigrams), u and v (light blue dotted lines and orange dotted hyperbolic lines in Figure 2.1) respectively constitute the effective measures of similarity (u in light blue) and hierarchical relations (v in orange). Specifically, the model generates a common two-dimensional representation in which four interpretable linguistic features and their relations can be mapped. Canonically, on the u dimension of the hyperbolic coordinates, strong typicality is mapped towards the origin, and I term the direction opposite to strong typicality in the same hyperbolic axis as *idiosyncrasy*. On the v dimension of the hyperbolic coordinates, contextuality is mapped towards directions orthogonal to the typicality-idiosyncrasy dimension, and I term one direction of the v dimension as *contextual dominancy* and the other direction *contextual dependency*. Section 2.4.1 along with later sections in this and next chapters will further provide empirical motivations and justifications for these canonical relations.

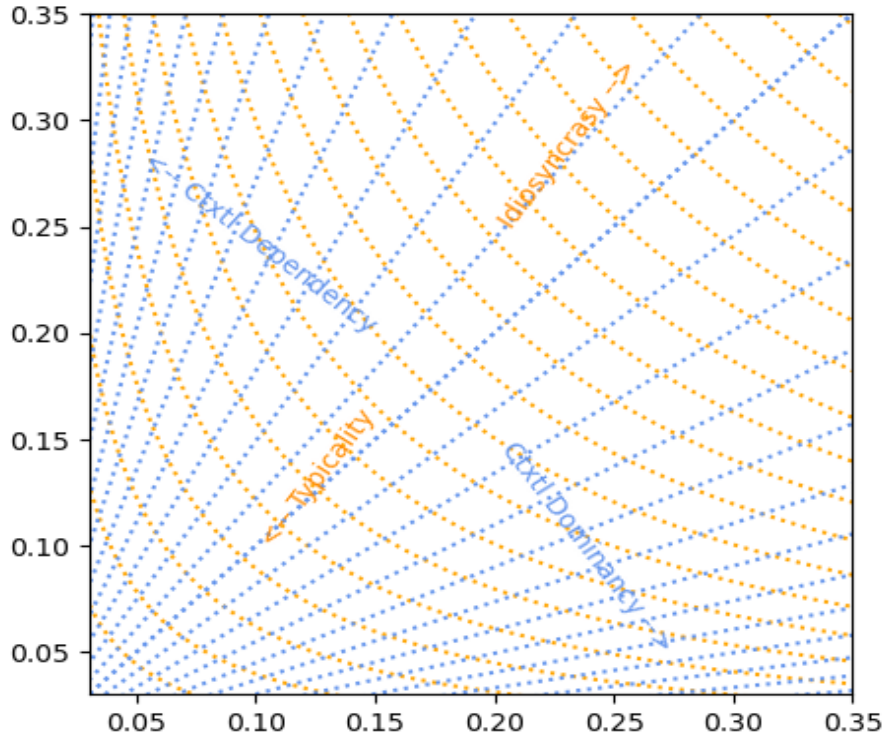


Figure 2.1. A geometrical representation of the canonical mapping of the typicality-contextuality embedding model. The u dimension maps semantic similarity relations (light blue) and the v dimension maps semantic hierarchical relations (orange). Testing, evaluation, and applications in later sections in this and next chapters provide empirical justifications that the similarity dimension (u in blue) captures semantic typicality and idiosyncrasy in its two opposite directions, and hierarchy dimension (v in orange) captures contextual dominancy and contextual dependency in its two opposite directions. See equations (2.5) and (2.6) for the definitions of (x, y) and (u, v) coordinates.

2.3.3. Computational and cognitive motivations of the basic metrics

The two basic weighting functions that are used to generate the embedding metrics are collocation diversity and probability ranking. Collocation diversity is intrinsically a concept of link degree in network theory, which is occasionally used in some work on semantic modeling (Vincze, 2015). In some literature on network theory and computational neural modeling, link

degree and ensemble synchronization are related to each other (Barabási, 2009), indicating the possibility that collocation diversity could be a quantitative measure of how linguistic ensemble properties such as similarity or dissimilarity among semantic categories are learned through the developmental stages of the human. Probability ranking, on the other hand, can be formally understood as link ranking, which is demonstrated to be relevant to hierarchical properties in complex networks, as well as to curvatures of hyperbolic space in a power relation (Krioukov et al. 2010). These properties provide further computational justifications for embedding the weighted conditional-entropy metric in Equation (2.1) into a hyperbolic space to generate meaningful semantic representations. Broadly, the justifications of hyperbolic embedding in this work can be further observed within the context of other hyperbolic embedding work in NLP and computational neuroscience (Nickel & Kiela, 2017, Sala et al., 2018, Dhingra et al., 2018, Sharpee, 2019). But this embedding model arguably provides richer, more well-grounded, and more interpretable representations than other existing embedding models in the literature.

2.4. Empirical testing and applications in the domains of color and kinship semantics

2.4.1 Mapping semantic similarity and hierarchy in a unified representation

The first experiment tests how the procedure described in Equation (2.4) extracts similarity relationships from the corpus data, with a comparison with the *word2vec* method in

word embedding (Mikolov et al., 2013)³. Testing is performed for the domains of color, kinship, and country/city names. The data shown here is averaged across two corpuses of English Wikipedia texts randomly chosen from the whole English Wikipedia data. Each of the corpuses contains 8M word tokens.

Target word	15 words closest to the target word, modeled by the similarity ranking algorithm of the typicality- contextuality model	15 words closest to the target word, modeled by <i>word2vec</i> algorithm (Mikolov et al., 2013)
red	<u>red</u> , <u>blue</u> , <u>white</u> , <u>green</u> , <u>black</u> , <u>yellow</u> , <u>orange</u> , <u>pink</u> , <u>purple</u> , colors, color, <u>brown</u> , sox, dark, <u>gray</u> , star	<u>red</u> , <u>white</u> , <u>blue</u> , <u>green</u> , <u>black</u> , <u>brown</u> , cross, color, <u>yellow</u> , star, similar, <u>orange</u> , head, gold, dark
blue	<u>blue</u> , <u>white</u> , <u>red</u> , <u>green</u> , <u>yellow</u> , colors, <u>purple</u> , <u>orange</u> , color, <u>pink</u> , dark, <u>brown</u> , colour, <u>gray</u> , bright, stars	<u>blue</u> , <u>green</u> , dark, <u>brown</u> , <u>yellow</u> , <u>white</u> , <u>red</u> , color, appearance, cover, <u>black</u> , wild, colour, bear, cat
purple	<u>purple</u> , <u>green</u> , <u>yellow</u> , stripes, <u>orange</u> , <u>blue</u> , colors, bright, <u>red</u> , colour, <u>brown</u> , <u>white</u> , color, dark, wearing, <u>gray</u>	<u>purple</u> , <u>pink</u> , pale, frog, flower, darker, skull, helmet, candy, bean, honey, ash, pig, coloured, feathers
brother	<u>brother</u> , <u>son</u> , <u>daughter</u> , <u>sister</u> , sir, married, Henry, reign, succeeded, duke, Mary, king, Philip, Albert, prince, <u>sons</u>	<u>brother</u> , <u>daughter</u> , <u>wife</u> , friend, married, prince, <u>son</u> , <u>sons</u> , <u>father</u> , <u>husband</u> , <u>mother</u> , <u>sister</u> , duke, throne, <u>brothers</u>

Table 2.2. Semantic similarity extracted by the typicality-contextuality model, compared with results by a word embedding method in the literature. Left column: the similarity ranking algorithm of the typicality-contextuality model in this paper. Right column: word similarity mapped by the original word2vec algorithm (Mikolov et al., 2013). Results for three basic color words in modern English as target words ("red," "blue," "purple") are shown in the table, and the last row shows results for one kinship term ("brother"). Basic color and kinship terms that are properly mapped are underlined. Overall, for measuring semantic similarity relationships, the typicality-contextuality model performs slightly better than the word2vec algorithm on the three

³ Since this model takes a cognitive position for modeling semantics, the testing does not consider comparisons with most of the NLP models that have to be trained on by DN models with corpuses usually in the size of billions of word tokens.

color terms, and slightly worse on the one kinship term (see text for comparisons). Corpus: averaged on two randomly chosen English Wikipedia dataset, each with 8M word tokens. Context size = 4-word tokens (**Table 2.2**).

The color and kinship domains are among the most intensively studied domain-specific approaches in cognitive science (Kemp & Regier, 2012; Zaslavsky et al., 2019). The testing here models 11 basic color words and 12 kinship words using the typicality-contextuality embedding, and then compares these results with those from the word2vec embedding. The color words include "red," "yellow," "green," "blue," "white," "black," "brown," "orange," "pink," "purple," "gray," while the kinship words include "father," "mother," "son," "daughter," "brother," "sister," "husband," "wife," "uncle," "cousin," "grandfather," "grandmother". The typicality-contextuality model outperforms the Word2vec embedding marginally on basic terms and varies on kinship terms (see Table 2.2 for sample results). More specifically:

- (i). The Word2vec and the typicality-contextuality method exhibits similar performance when extracting basic color words that are semantically close to the most typical color words (e.g., "red" and "blue").
- (ii). The typicality-contextuality method excels in handling less typical basic color words (e.g., "purple" and "gray"). It also models three generic color terms in English ("color," "colour," "colors") effectively for both typical and less typical basic color words, a task that the word2vec method does not perform equally well with the same corpus and context sizes.

(iii). The performance of the two models on kinship terms is more varied. For instance, for the target word "brother," the weighted entropy method is more effective at extracting three basic kinship words as the semantically closest ones ("son," "daughter," "sister") compared to the word2vec method. However, the word2vec method is capable of extracting four additional basic kinship terms ("father," "husband," "mother," "sons") that the typicality-contextuality method does not capture within 15 words (the 4th column, Table 2.2).

Overall, for embedding word-level similarity relations in these two semantic domains, the typicality-contextuality model performs at least as well as the standard word-vector method.

2.4.2. Mapping basic-level categories of color and kinship in hyperbolic coordinates

Equations (2.5) and (2.6) map words that are selected or extracted from the previous procedures to the two-dimensional Euclidean space and then to the hyperbolic coordinates (u, v) , where a geometrical mapping for both similarity and hierarchical relations emerges, where the v coordinate maps hierarchy (typicality and idiosyncrasy) and the u coordinate maps similarity

(contextual dominance and contextual dependency).

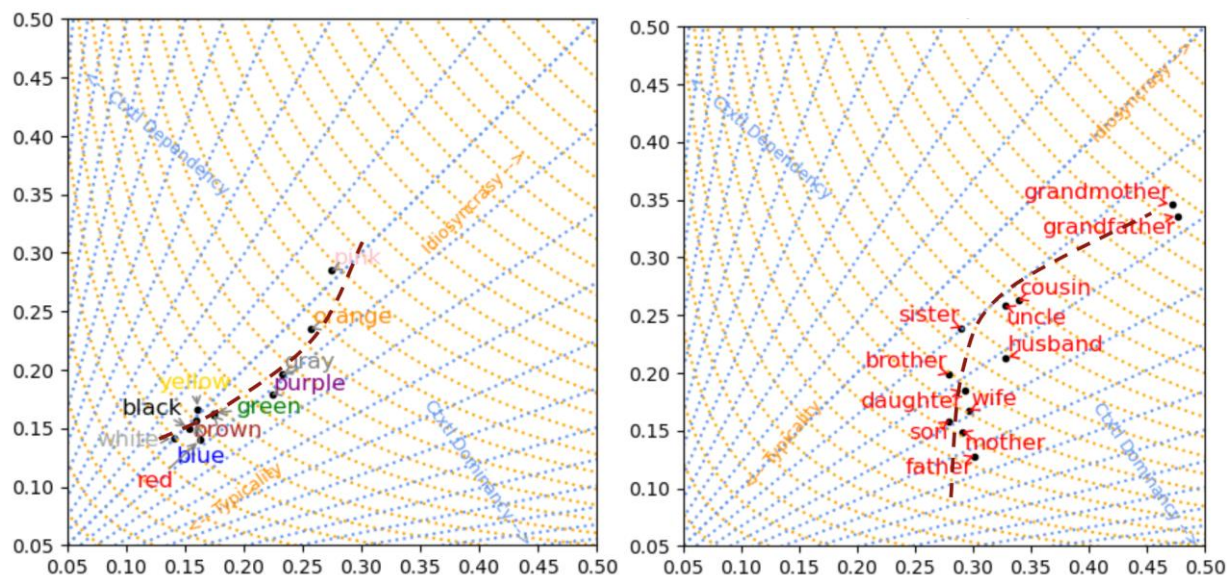


Figure 2.2. Mapping basic-level color and kinship semantic categories in modern English. The dark red dotted lines show the overall direction in which the words being mapped are distributed. Left: A representation of the mapping of 11 basic color words in English in the hyperbolic coordinates. From bottom left (more typical) to upper right (less typical): "white," "red," "blue," "black," "green," "yellow," "gray," "purple," "orange," and "pink". Right: from bottom left (more typical) to upper right (less typical), mapping of 12 basic kinship words in English: "father," "mother," "son," "wife," "brother," "daughter," "sister," "husband," "uncle," "cousin," "grandfather," and "grandmother". The u dimension maps semantic similarity as contextual dominance and contextual dependency (light blue) and the v dimension maps semantic typicality/idiosyncrasy (orange). See Equations (2.5) and (2.6) for the definitions of x and y axis and u and v coordinates. Corpus: a randomly chosen Wikipedia English corpus of 8 M words. Context size = 4 word tokens.

The first evaluation of the results is based on the domain-specific approach of studying basic-level categories in cognitive science. Domain-specific cognitive research proposes a rigorous theory about the basic color terms and their rankings, and comprehensively surveys more than one hundred languages for testing the theory. In English, 11 basic color words are identified and ranked in the following order, which can be treated as one reference of typicality

ordering: dark and light colors> "red" > "yellow," "green" > "blue" > "brown" > "purple," "pink," "orange," "gray" (Berlin & Kay, 1969, Regier et al., 2015)⁴.

Figure 2.2 shows mapping results of the 11 basic color words in English by the typicality-contextuality model. Specific observations are as follows:

- (i). All 11 basic color words are mapped closely to one another in the similarity axis, with "red," "yellow," "green," "blue," "white," "black" being mapped as most typical and similar to one another.
- (ii). "white," "black," "red," "blue," "yellow," "green," "brown" form a cluster of the more typical and typical color words.
- (iii). "gray," "purple," "orange," and "pink" are mapped as less typical, with the order of "purple" > "gray" > "orange" > "pink".

Overall, the typicality-contextuality model distinguishes the most typical color words from other less typical ones, which follows the general prediction made by the domain-specific method, without one exception of the word "brown". Nonetheless, the model identifies a high degree of homogeneity among the most typical color words ("white," "black," "red," "blue," "yellow," "green," "brown") yet not for the less typical words ("purple," "gray," "orange,"

⁴ Dark and light colors are ranked prior to "red", but as Berlin and Kay (1969) note that the terms for them may not necessarily be words like "black" and "white" in English. Following the usual practice in many NLP algorithms, the similarity algorithm here does not differentiate letter cases, and all words are coded as lowercases. For words of "brown" and "gray," after the algorithm selects and ranks candidate co-occurrent contextual words, English names related to "brown" and "gray" in their selected contexts are manually excluded.

"pink"), whereas the domain-specific method predicts the opposite (see also rows 1 and 2 of Table 2.2).

The testing also investigates the domain of kinship, another domain that the domain-specific literature in cognitive science mostly focuses on (e.g., Kemp & Regier, 2012). The main results are summarized as follows:

- (i). The typicality-contextuality model successfully predicts 12 basic-level kinship terms: "father," "mother," "son," "wife," "brother," "daughter," "sister," "husband," "uncle," "cousin," "grandfather," and "grandmother" in English; "aunt," "niece," and "nephew" are not extracted due to their low frequencies. Among the 12 predicted terms, "mother," "father," "son," "daughter," "brother," "sister," "husband," "wife," "uncles" and "cousin" form a cluster of typical kinship terms more loosely organized than the most typical color terms, and "grandfather" and grandmother" form the least typical cluster (Figure 2.2 right).

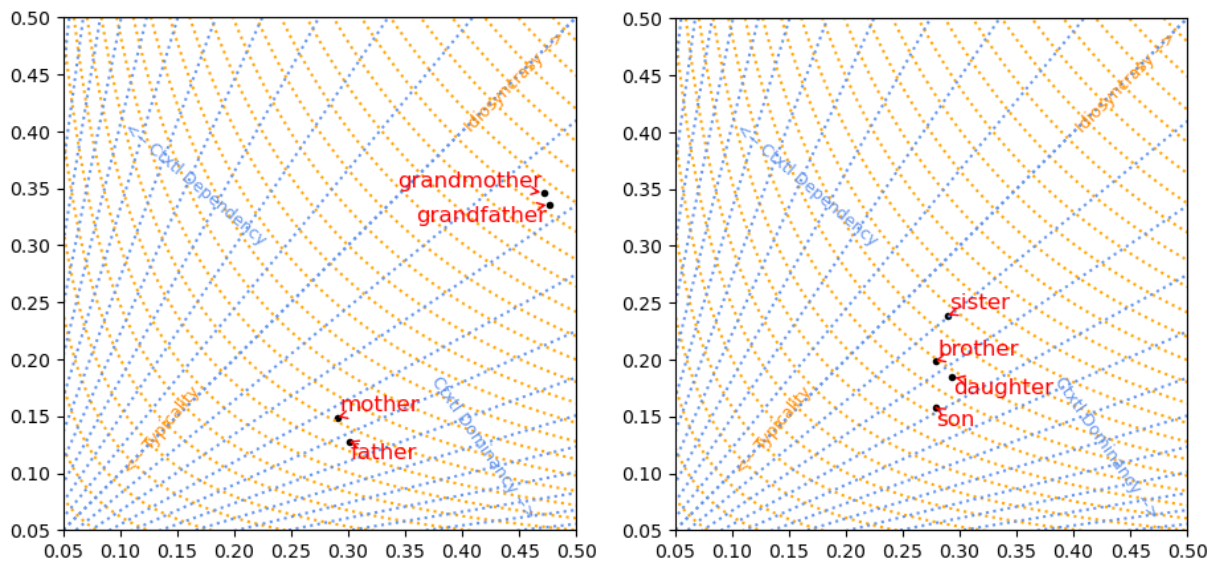


Figure 2.3. Mapping main basic kinship terms in English. The semantic relationships between "father" and "mother" are mapped as equally typical, and "grandfather" and "grandmother" as equally atypical and much less typical than the pair of "father" and "mother". Right: The respective semantic relationships between "son" and "daughter," and "brother" and "sister" are mapped as almost parallel to one another in the hyperbolic space. On the other hand, the mapping predicts that "son" and "brother" are respectively more typical than "daughter" and "sister" to a similar extent. This result does not assume or follow the hyponym-hypernym relationships formulated and annotated in the simplified cognitive and computational settings such as WordNet.

- (ii). The model ranks "father" and "mother" as equally typical, and "grandfather" and "grandmother" as equally much less typical than "father" and "mother" (Figure 2.3 left). On the other hand, the model predicts that "son" is more typical than "daughter" and "brother" is more typical than "sister" to a similar degree. Unlike the domain-specific method which predicts the same degree of typicality for every pairwise kinship relationships (Hirschfeld, 1986, Kemp & Regier, 2012), the typicality-contextuality model predicts that some pairwise kinship relationships have unequal degrees of typicality, but the degrees of typicality asymmetry of certain isomorphic kinship relationships can be close, such as the typicality relationship like {"son": "daughter" :: "brother" : "sister"} (Figure 2.3 right).

2.4.3. *Convergence Testing*

Convergence testing shows that results are substantially robust after the training reaches 5 million tokens. Figure 2.4 shows that distances between all trained color words in both the similarity and typicality dimensions begin to converge after the 4th or 5th epoch. For the

similarity distance variation between the 5th and 6th epochs of all color words, the mean = 0.0091 and the variance = 0.0014; for the typicality distance variation between the 5th and 6th epochs of all color words, the mean = 0.0084 and the variance = 0.0037.

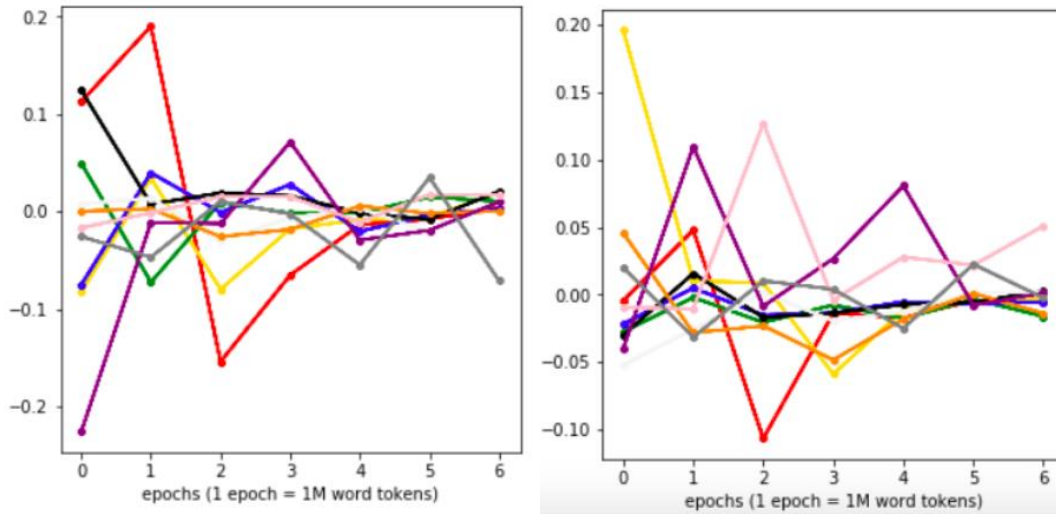


Figure 2.4. Robustness testing of the training. In the robustness testing, the mapping of each color word starts to converge during the 4th or 5th epochs, after the model being trained on 5 to 6 M word tokens. Each epoch runs on 1M word tokens. The 0th epoch = 1M words. Each color corresponds to the basic color word in English. In this figure, each color-coded line represents the training of the corresponding color word. Left: the y -axis is the distance difference on the dimension of similarity u between color words across every two epochs. It is shown here that on the similarity dimension u , every color word (except "gray") begins to converge during the 4th epoch (after being trained on 5M word tokens). Right: the y -axis is the distance difference on the dimension of typicality v between color words across epochs. It is shown here that on the typicality dimension v , every color word (except "pink") begins to converge around the 5th epoch (after being trained on 6M words) (**Figure 2.4**).

2.5. Testing with a Mturk study indicates the measure-dependency of typicality

Tests of the model in the previous section use the literature of the domain-specific method for studying basic-level categories in cognitive science, which emphasizes domain specificity and differentiating basic from non-basic types of categories. More generally, researchers in cognitive psychology who study the questions of typicality do not assume these two particular positions. These studies often perform typicality testing over many semantic categories in the same study and usually do not focus on developing clear differentiations between the basic terms and non-basic ones. More problematically, these studies have not agreed upon what are the most empirically reliable measures for typicality: main candidates for the measure of typicality include word production frequency, ranking in elicitation tasks, age of acquisition, among others; but these measures yield discrepant results in empirical studies (Rosch, 1975; Uyeda & Mandler, 1980; Plant et al., 2011; Holmes & Ellis, 2007; Banks & Connell, 2022)⁵.

This section describes an Amazon Mturk study designed within specific parameter ranges to further address the modeling results. Further, I compare the results with other studies on ranking typicality in multiple semantic categories in the extant literature. The Mturk study contains two sections, one for studying typicality of color terms and the other for typicality of kinship terms. In each sub-task, the participant is asked to name 7 color or kinship words, in the order of the "most typical" to the "least typical," according to their judgment. An instruction and sample task that illustrate the procedure is provided before the task. Data from 190 participants

⁵ Some literature in cognitive science and computational linguistics implicitly or explicitly treats single-word frequency in corpus as the measure of typicality. Though straightforward and convenient as it is, much research in psycholinguistics since the 1970s show that single-word frequency from corpus or passive exposure has low correlation with semantic typicality (e.g., Uyeda & Mandler, 1980).

were collected for both color and kinship domains. Data from participants that obviously did not follow the instruction was rejected, and the tasks were re-published for new participants. Responses in which more than half of the words are not in the 11 color words or 12 kinship words that this current study investigated are preserved (for example, "vertigo," "Hawaiian," "family"), but not included in the current analysis. In total, 141 responses for the color domain and 89 responses for the kinship domain were considered as both valid and relevant, and used for the analysis. (The Mturk project was reviewed and approved by the IRB of the Social Science Division of the University of Chicago: IRB21-0790) Besides the domain-specific studies on typicality (Berlin and Kay, 1969; Regier et al., 2015), results are also compared with three empirical studies on typicality that include results on the color or kinship domain: ranking through production frequency and through typicality rating in Banks & Connell (2022) and ranking through rating in Uyeda & Mandler (1980).

	COLOR	KINSHIP
Typicality-contextuality model*	<i>red, yellow, green, blue, brown > purple > gray > orange > pink</i>	<i>{mother = father} :: {grandfather = grandmother}</i> <i>{brother > sister} :: {son > daughter}</i>
Domain-specific approach (Berlin & Kay, 1969 for color; Kemp & Regier, 2012 for kinship)*	<i>red > yellow, green > blue</i> <i>> brown > purple, pink, gray, orange</i>	<u><i>{mother = father}</i></u> > <u><i>{son = daughter}</i></u> > <u><i>{brother = sister}</i></u> > <u><i>{grandfather = grandmother}</i></u>

Production frequency with word number limit (Mturk study in this chapter)	<i>red, yellow, green, blue</i> > <i>orange</i> > <i>purple, pink</i> > <i>brown, gray</i>	<i>mother</i> > <i>father</i> { <i>brother</i> > <i>sister</i> } :: { <i>son</i> > <i>daughter</i> } :: { <i>grandmother</i> > <i>grandfather</i> }
Pairwise first production frequency with word number limit (Mturk study in this chapter)	(Not applicable)	<u>{{<i>mother</i> = <i>father</i>}}::{{<i>brother</i> = <i>sister</i>}}</u> } > <u>{{<i>son</i> = <i>daughter</i>}}::{{<i>grandfather</i> = <i>grandmother</i>}}</u> }
Production frequency with time limit (Banks & Connell, 2022)	<i>red, yellow, green, blue, purple</i> > <i>orange, pink</i> > <i>gray</i> > <i>brown</i>	<u>{<i>brother</i> = <i>sister</i>}</u> > { <i>mother</i> = <i>father</i> } > <u>{<i>son</i> = <i>daughter</i>}</u> > <u>{<i>grandfather</i> = <i>grandmother</i>}</u>
Typicality rating (Banks & Connell, 2022)	<i>pink</i> > <i>red, blue</i> > <i>brown</i> > <i>yellow, green, purple, gray</i> > <i>orange</i>	<i>mother</i> = <i>father</i> = <i>brother</i> = <i>sister</i> = <i>son</i> = <i>daughter</i> = <i>grandfather</i> = <i>grandmother</i>
Typicality rating (Uyeda & Mandler, 1980)	<i>red, yellow, blue</i> > <i>green</i> > <i>orange</i> > <i>purple</i> > <i>pink</i> > <i>brown</i> > <i>gray</i>	(Not applicable)

Table 2.3. Comparing predictions on the typicality of main color and kinship words in English made by the model and by other theoretical and empirical works in the literature. Results of typicality ranking for the basic-level color and kinship words, predicted by the typicality-contextuality model and the domain-specific theory of semantic categorization (marked by "*"). The theoretical predictions are compared with empirical results from the Mturk studies in this paper as well as other three studies in the existing literature (rows three to six). Typicality predictions from the typicality-contextuality model are closest to the empirical results from Mturk studies of word production frequency with word number limit, as indicated in **bold**. The *italicized* words indicate one group of similar results, and the underlined words represent the

other group of similar results. "{}" indicates one pairwise kinship relationship, "=" and ">" represents equal or superior typicality relationships. See Appendix A for the details of scoring and re-weighting for the results of each study (**Table 2.3**).

Table 2.3 summarizes the results and comparisons (See Appendix A for the details of rating and weighting scores for each study). For the color domain, the typicality-contextuality model ranks "red," "yellow," "green," "blue" as the most typical basic color words, the pattern of which generally agrees with results from three empirical methods: this Mturk study, the domain-specific approach, and production frequency within a time limit (rows 2, 4, 5, 1st column, Table 2.3). Further, the domain-specific approach provides more refined ranking within these four color terms (rows 4, 1st column, Table 2.3), whereas the typicality-contextuality model and the production frequency within a time limit respectively indicates one extra word that should be included in the cluster ("brown" for the former and "purple" for the latter; (rows 1 and 5, 1st column, Table 2.3). On the other hand, results on typicality ranking of the other five words, "brown," "orange," "purple," "pink," "gray" show much higher variations, for which none of the modeling or empirical studies completely agree with one another, except that the typicality of "purple" is consistently ranked equal to or higher than at least two of these five words. Lastly, the two studies that use typicality rating methods do not agree with each other on the color words results (rows 6 and 7, 1st column, Table 2.3).

For the results in the kinship domain, choosing production frequency as the typicality measure provides results that are similar to what the typicality-contextuality model predicts, in which typicality of each pair of kinship terms is not symmetrical (e.g., the typicality of

"brother" > the typicality of "sister") , but such asymmetry has linearly proportional regularity (e.g., {brother > sister} :: {son > daughter}; rows 1 and 2, 1st column, Table 2.3). However, choosing word production frequency of first appearances in each pairwise relationship (e.g., whether "brother" tends to appear before "sister" in the Mturk data) instead gives results that are similar to alternative predictions made by the domain-specific method and production frequency within a time limit, in which words in each pair of kinship relationship show an equal degree of typicality (e.g., the typicality of "brother" = the typicality of "sister"; rows 4 and 5, 2nd column, Table 2.3). Additionally, one study that uses the method of typicality rating shows low resolution, without effectively differentiating typicality ranking of any kinship words (row 6, 2nd column, Table 2.3).

Importantly, the above analysis shows that typicality predictions from the typicality-contextuality model on both domains of color and kinship words are closest to the empirical results from Mturk studies of word production frequency with word number limit (studies in rows 1 and 3 in **bold**, Table 2.3). These results echo the hypothesis raised in multiple empirical studies since the 1980s, that typicality is highly correlated with word production frequency (Uyeda & Mandler, 1980, Plant et al., 2011, Banks & Connell, 2022). At the same time, the results are measure-dependent: word production frequency with time limit (as shown in row 5) shows lower correlation with typicality predicted by the typicality-contextuality model than word production frequency with word number limit.

In sum, results from the two domains demonstrate the capacity of the typicality-contextuality model, as well as measure dependency of modeling and evaluating typicality.

Though in both domains, mapping typicality is shown to have strong regularity, the results of the color words domain indicate relatively a lower degree of measure dependency and the results of the kinship words domain shows a higher degree of measure dependency.

Though Model I cannot directly address the discrepancy of measure dependency across these two domains in English, the metacognitive model (Model II) developed in Chapter IV and applied to color and kinship domains in Chapter V is able to provide a cognitive-computational account for it. See Sections 5.3 and 5.4 in Chapter V for details.

2.6. Comparison with basic-level color and kinship semantics in Modern Chinese

In this section, the typicality-contextuality model is also applied to basic-level color and kinship semantics in standardized modern Chinese. Interestingly, though only single-character words are mapped⁶, similar to modern English, the basic-level terms in both semantic categories are mapped to be predominantly distributed along the typicality-idiosyncrasy dimension (Figure 2.5). Similar to the mapping for color and kinship semantics in English, in both these mappings, the u dimension represents semantic similarity (shown in light blue), while the v dimension

⁶Different from words in English, Chinese words can be formed by single or multiple characters (usually up to three or four). An embedding model of an indo-European language with single words as units is equivalent to an embedding of a Chinese language with single characters as units. In this thesis, though all modeling work on Chinese focuses on single-character words, this is only for practical reasons. In principle, the typicality-contextuality model can map multi-character words and expressions in Chinese in a similar way as single-character/word cases with minor adjustments. This possibility is demonstrated in employing the typicality-contextuality model to map multiword expressions in English (See section 2.7 in this chapter and section 6.5 in chapter VI).

represents semantic typicality/idiosyncrasy (shown in orange). Each mapping is drawn from a randomly chosen Chinese Wikipedia corpus of 8 million words, with a context size of four character tokens.

For the domain of color semantics, 9 basic-level color words in modern Chinese are mapped within hyperbolic coordinates. In Figure 2.5 (left), these terms run from bottom left (more typical) to upper right (less typical) as follows: *bai* (white), *hong* (red), *huang* (yellow), *hei* (black), *lan* (blue), *lv* (green), *zi* (purple), *ju* (orange), and *zong* (brown). Though relatively more widely distributed in the mapping than the corresponding English color terms, the most typical color terms are the same: "white," "black," "red," "yellow," and "green". "Purple" and "orange," on the other hand, are mapped as relatively less typical and even more widely distributed on the typicality-idiosyncrasy dimension. These patterns are generally shared across both modern English and Chinese.

What differs in basic-level color semantics for modern Chinese from modern English are as follows: first, "brown" is ranked as the least typical basic-level term in the chosen color words, which is different from "brown" in modern English. This also does not agree with Berlin and Kay (1969)'s prediction that "brown" is the most typical term after "blue" and "green" across cultures and languages. Also, "pink" and "gray" are not extracted for the mapping. This agrees with empirical studies in the cognitive science literature that predict that the number of basic-

level color terms in modern Chinese range from 9 to 11, with "pink" and "gray" being not completely identified as basic-level ones (Sun and Chen, 2018; Gao and Sutrop, 2014)⁷.

For the domain of kinship semantics, 16 basic kinship single-character words in standardized modern Chinese are mapped. In Figure 2.5 (right), these terms run from bottom left (more typical) to upper right (less typical) as follows: *fu1* (husband), *bo* (fraternal elder uncle), *fu2* (father), *mu* (mother), *di* (younger brother), *shu* (fraternal younger uncle), *xiong* (elder brother), *gu* (young aunt), *nai* (grandmother), *zhi* (nephew), *qi* (wife), *ye* (grandfather), *mei* (younger sister), *yi* (elder aunt), *jie* (elder sister), and *jiu* (maternal uncle).

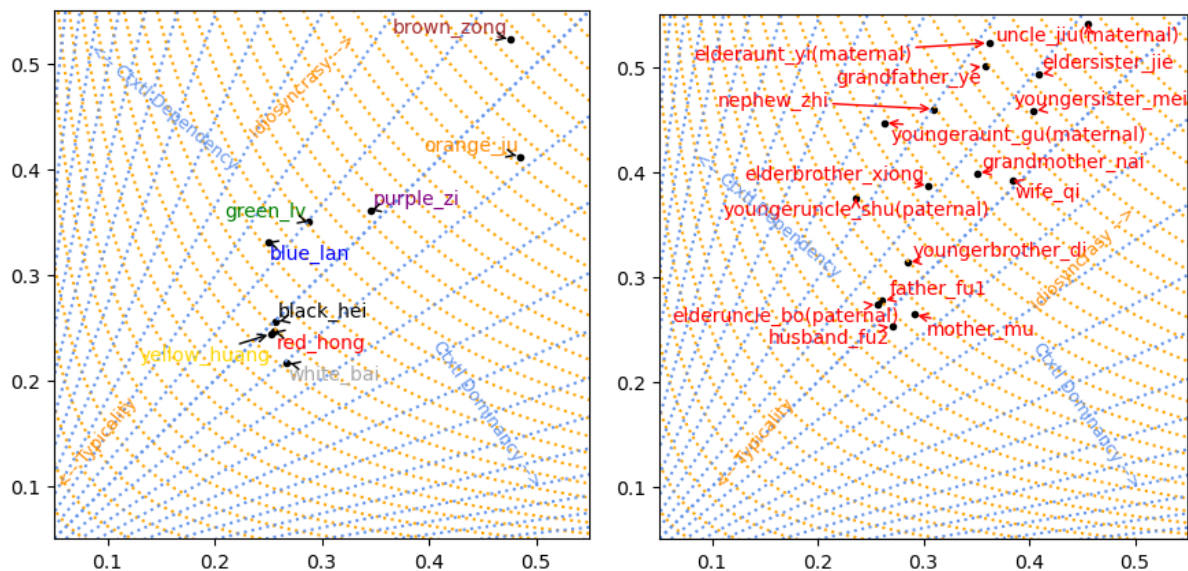


Figure 2.5. Mapping basic-level color and kinship semantics in standardized modern Chinese. Similar to color and kinship semantics in modern English, both groups of basic-level terms are predominantly distributed along the typicality-idiosyncrasy dimension. Left: A representation of

⁷ In modern Chinese, pink (*fen*) is either identified as one type of red (*fen-hong*: "pink red"), or the word still evokes a connotation of powder, while gray (*hui*) is often cognized together with the meaning of ash.

the mapping of 9 basic color words in modern Chinese in the hyperbolic coordinates. From bottom left (more typical) to upper right (less typical): *bai* (white), *hong* (red), *hei* (black), *lan* (blue), *lv* (green), *zi* (purple), *ju* (orange), *zong* (brown). Right: from bottom left (more typical) to upper right (less typical), mapping of 16 basic kinship single-character words in standardized modern Chinese: *fu1* (husband), *bo* (paternal elder uncle), *fu2* (father), *mu* (mother), *di* (younger brother), *shu* (paternal younger uncle), *xiong* (elder brother), *gu* (young aunt), *nai* (grandmother), *zhi* (nephew), *qi* (wife), *ye* (grandfather), *mei* (younger sister), *yi* (elder aunt), *jie* (elder sister), *jiu* (maternal uncle). The *u* dimension maps semantic similarity (light blue) and the *v* dimension maps semantic typicality/idiosyncrasy (orange). See Equations (2.5) and (2.6) for the definitions of *x* and *y* axis and *u* and *v* coordinates. Corpus: a randomly chosen Wikipedia Chinese corpus of 8 M words. Context size = 4-character tokens (**Figure 2.5**).

The overall distribution of basic-level kinship terms (single-character ones) in standardized modern Chinese is categorically different from kinship terms in modern English. Crucially, this difference appears to be not accounted for by a communicative or information account but is best accounted for by sociocultural reasons. While typicality relationships between many basic kinship terms in English more closely approximate sociobiological distances among these terms (Figure 2.2 right)⁸, the typicality and similarity relationships shown for basic-level kinship characters in modern Chinese more closely resemble a patrilineal order, in which a more typical term is close to a more dominant position in a patrilineal society. Figure 2.5 (right) shows that "husband" (*fu2*), "father" (*fu1*), and "paternal elder uncle" (*bo*) are ranked as the most typical/dominant kinship terms in Standardized modern Chinese. While in modern English, "father," "mother," and "son" are ranked as the three most typical ones (Figure 2.2 right).

⁸ Applications of Model II to kinship terms in American English, nonetheless, show that they are also subject to influences of sociocultural factors - albeit very different ones from those in Chinese: the sociocultural evolution of the American family in the 19th and 20th centuries. See Section 5.4 in Chapter V.

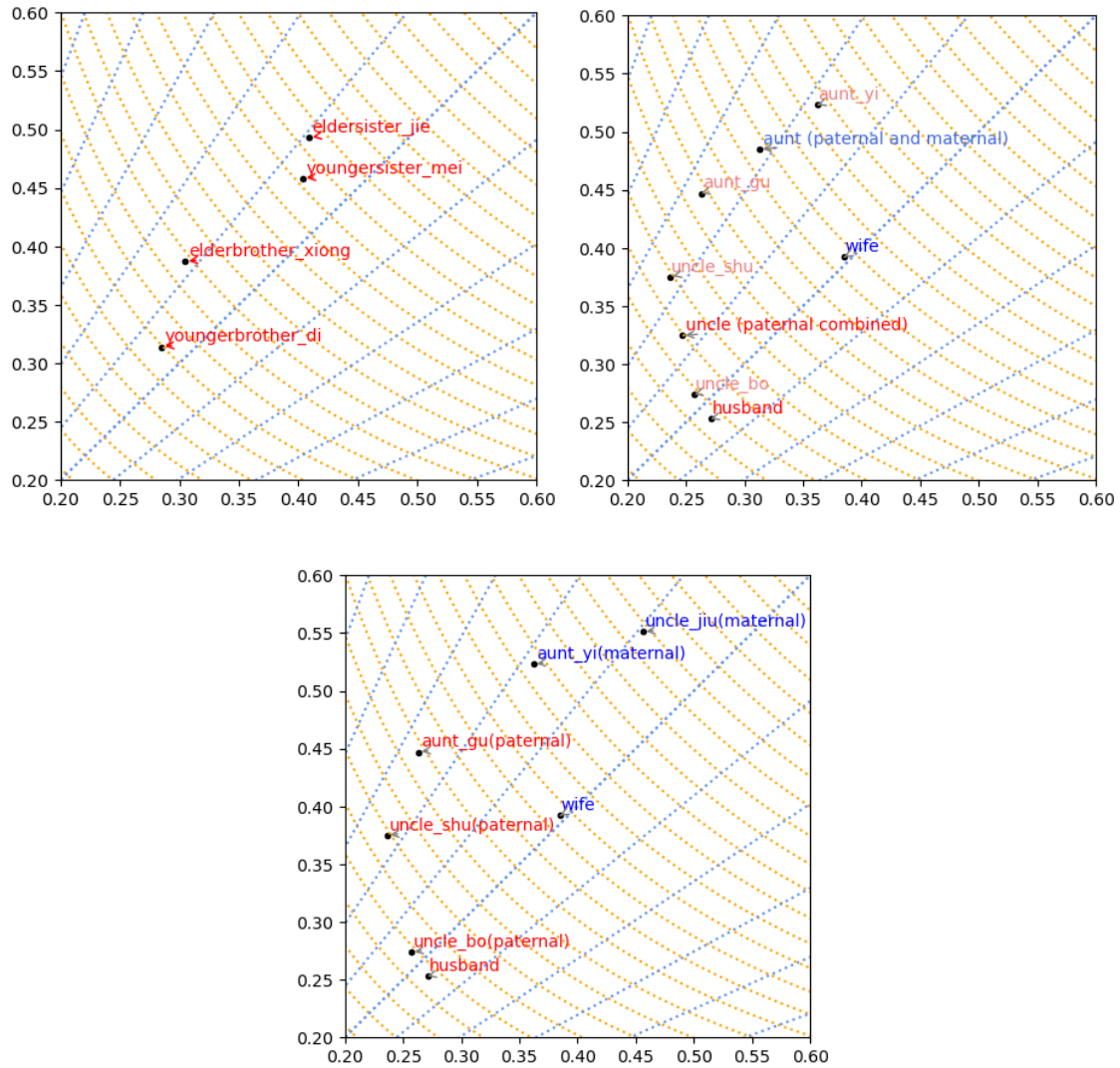


Figure 2.6. The hierarchical and similarity relationships shown for basic-level kinship characters in modern Chinese possess a patrilineal character. In the top right and bottom figures, paternal terms are coded in red, maternal terms are coded in blue, and maternal and paternal combined term is coded in dodger blue. In the top right figure, the mapping of aunt is averaged from the two terms for aunt (*gu* and *yi*, marked in light coral red), and the mapping of uncle is averaged from the two terms for paternal uncles (*bo* and *shu*, marked in light coral red).

Additionally, Figure 2.6 (top left) shows that while "younger brother" (*di*) and "elder brother" are similarly mapped as "younger sister" (*mei*) and "elder sister" (*jie*) (i.e., "younger

brother" : "elder brother" :: "younger sister" : "elder sister"), the two characters of brother are ranked as more typical/dominant than the two characters of sister. Figure 2.6 (top right) shows the relationship between paternal uncles combined (*bo* and *shu*) and paternal and maternal aunts combined (*gu* and *yi*) are similarly mapped as the relationship between husband (*fu2*) and wife (*qi*), and the terms of husband and paternal uncles are ranked most typical/dominant. Figure 2.6 (bottom) shows that all paternal terms are ranked more typical and predominant than all maternal terms. Furthermore, within paternal terms or maternal terms, the typicality/dominance ranking also follows a patrilineal logic ("husband" > "elder paternal uncle" > "younger paternal uncle" > "paternal aunt"; "wife" > "maternal aunt" > "maternal uncle"⁹).

Together, employing the typicality-contextuality model to map color and kinship terms in standardized modern Chinese further demonstrates the model's effectiveness in mapping domain-oriented basic-level semantic knowledge. The modeling shows domain robustness in the domain of basic-level color semantics in modern Chinese and English, as well as sensitivity to sociocultural difference in the domain of basic-level kinship semantics across these two languages.

⁹ That relation of "maternal aunt" > "maternal uncle" follows a patrilineal logic can be understood in a simple egocentric manner: a maternal uncle in a paternal society would be ranked in a higher social position in a different paternal line, and thus could be ranked lower socio-culturally than a maternal aunt according to one's own paternal family line.

2.7. Further evaluations in the domains of place names and multiword expressions (MWEs)

This section introduces another two semantic realms of testing and applications. The first group of tests is in the domain of country/city names, which forms one most classical benchmark for embedding algorithms in contemporary NLP (e.g., Mikolov et al., 2013; Pennington et al., 2014). Results show that while being capable of incorporating similarity relationships that many standard embedding methods in NLP can model to the mapping, the typicality-contextuality model is capable of also modeling hierarchical relationships between place names that standard embedding algorithms cannot easily map¹⁰. Compared to typical word-embedding algorithms which predominantly models similarity relationships, the typicality-contextuality model maps both typicality and similarity relationships: while relationships between each country name and capital name are shown to be almost parallel to one another in hyperbolic coordinates, the mapping also shows that country names in general have higher degrees of typicality than capital names, and some place names (e.g., Berlin, Paris) are more typical than others (e.g., Tokyo, Madrid) (Figure 2.7 left). Section 6.4 in Chapter VI will provide a more elaborated study using the typicality-contextuality mapping on place names, proposing that the modeling approach

¹⁰ The alternative hierarchical embedding methods in NLP are able to map such hierarchical semantic relationships, but they do not usually perform well on the similarity dimension (Nickel & Kiela, 2017; Sala et al., 2018; Dhingra et al., 2018). In the general literature of hyperbolic embedding beyond NLP, hyperbolic embedding that aims to embed two dimensions (hierarchy and similarity) and even multiple dimensions have been proposed (e.g. Jiang et al., 2022; Kovacs et al., 2022), but these works have little relevance with how their embedding and measures are linguistically or cognitively related. See the introduction and discussions sections in this chapter (Sections 2.1 and 2.8) for more details.

proposed in this chapter also surpasses other embedding models by integrating both semantic and pragmatic knowledge of country names in the mapping.

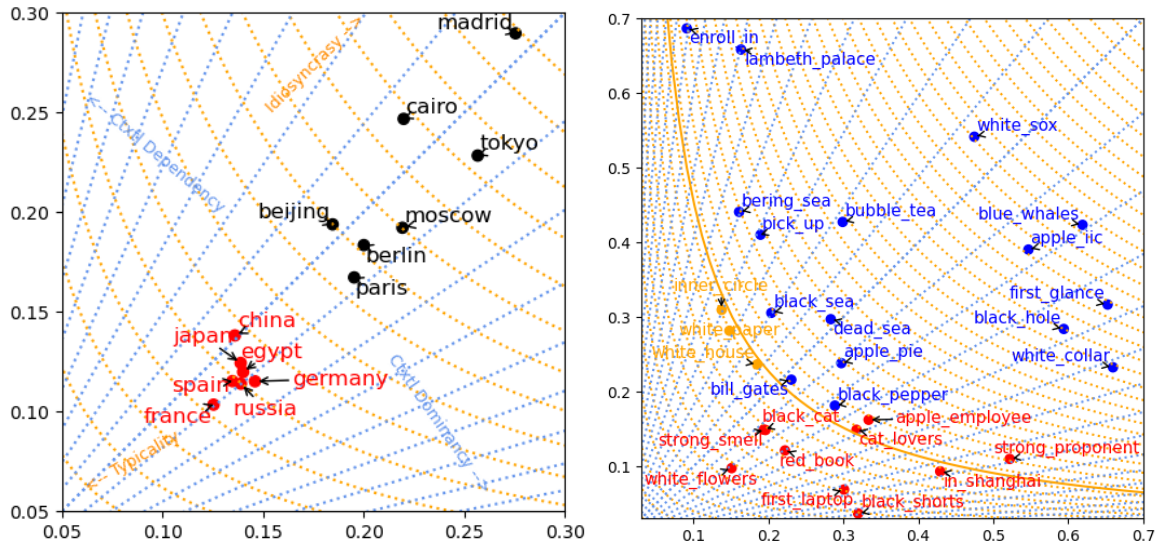


Figure 2.7. Left: mapping semantic relationships of several country/city names by the typicality-contextuality model. Country names are in red and city names are in black. Country names are ranked more typical (less idiosyncratic) than capital names, and all place names are measured as semantically close according to the contextuality metric. Context size = 2 word tokens. Right: mapping ordinary bigrams (in red) and bigram multi-word expressions (MWEs; in blue) by the typicality-contextuality model. The bold orange hyperbolic curve roughly classifies the two types of bigram expressions, and terms marked orange have dual meanings in English, which are mapped close to the orange hyperbolic curve. For both mapping, the corpus is randomly chosen from a *Wikipedia* English corpus of 8 M words.

MWE expressions also form a classical challenge in NLP (Sag et al., 2002; Petruck & Kordoni, 2015). It differentiates ordinary bigram expressions (such as "red car," "strong smell") from highly regular yet idiosyncratic expressions in natural language (such as "white collar," "blue whales"). Semantically, the importance of MWEs is that their meanings are typically not

derived from their components (i.e., the meaning of "white collar" cannot be derived simply from adding the meaning of "white" with the meaning of "collar").

Figure 2.7 (right) shows a mapping for randomly chosen bigram expressions by the typicality-contextuality mapping. Ordinary non-MWE bigrams (marked in red) and bigram MWEs (marked in blue) are approximately clustered in the hyperbolic space (marked by the bold orange hyperbolic curve). The mapping also successfully captures marginal bigram cases that can be both MWEs and non-MWEs ("white house," "white paper," "inner circle"; marked in orange). A thorough testing and application of the typicality-contextuality method to MWEs are not the task of this dissertation. There are two purposes in providing a preliminary example of mapping MWEs. The first is to illustrate the possibility that the modeling approach developed in this chapter provides a measure of semantic typicality and idiosyncrasy in the opposite direction of the dimension. Defining atypicality under a certain definition of idiosyncrasy has long been investigated and tested in empirical psychology, psycholinguistics, and sociolinguistics (Trousdale, 2008; Eckert, 2000), but a computational modeling approach for incorporating both typicality and idiosyncrasy is still lacking. Future work on testing the model with cognitive experiments on idiosyncrasy will provide more evaluation and evidence for the effectiveness of the model.

The second purpose for demonstrating an example of MWE mapping is to illustrate the potential of this modeling approach to move beyond the single-word level. In this dissertation, the empirical focus on single words is completely practical. As MWEs form more semantically independent units than non-MWEs for both semantic analysis and processing, a simple

procedure to differentiate MWEs from non-MWEs provides a way to recursively scale up for modeling N-grams. Scale-free properties are investigated in formal theories of hyperbolic embedding (Krioukov et al., 2010; Holme, 2019) and scientific investigations on brain activities (Kardan et al., 2020) and social phenomena (Holme, 2019; Barabási, 2009). Future testing and application will evaluate and explore possible scale-free properties of the typicality-contextuality mapping, as exemplified in its potential capacity of modeling MWEs.

2.8. Discussion

This discussion section addresses three aspects of the study that were described in the introduction and background sections: the assumption of domain specificity, the relationship between the cognition-driven modeling approach and modeling approaches in NLP and machine learning, and future directions of modeling and testing.

The reason that the current testing mainly focuses on the domains of color and kinship words is that research in cognitive psychology provides a rich theoretical and empirical groundwork for semantic cognition in these domains. Empirical baselines for evaluating the semantic relationships of basic color and kinship words are among the most developed in domain-specificity focused cognitive linguistic research. Much work, however, that applies the domain-specific approach in cognitive science to semantic categories faces difficulties in finding

common baselines for empirical evaluation beyond the color, kinship, and several other semantic domains¹¹. On the other hand, the methodology of the typicality-contextuality model is not limited to mapping semantic categories in a domain-specific way. This model can provide a useful tool for linking theoretical and empirical inquiries on domain-specific and domain-general semantic categories.

Further, it is acknowledged that in testing semantic typicality, other domain-specific factors may contribute or confound results. While being asked to list typical color words, the participant may mentally survey the color space rather than cognitive semantic knowledge in their mind. For testing typicality of kinship terms, many of these terms are morphologically or phonologically related ("grandfather" and "father," "father" and "brother"), which may also influence how participants order these terms in typicality testing. Recent literature, however, shows that even though other cognitive knowledge, such as embodied knowledge about the color space or perceptions of morphological and phonological similarities, may influence judgments of typicality in actual empirical tests, these factors appear to be minor or secondary (Banks and Connell, 2019).

In terms of the relation of this work with research in the NLP community, this work is related to current research in NLP that explores geometric spaces other than the Euclidean one

¹¹ In the literature of domain-specific cognitive psychology, the domains of shape and number also provide good data and baselines for further testing and applying the model, and computationally investigating cross-domain semantic relationships, which empirical cognitive studies also focus on. Testing on these two domains and on the cross-domain cases will be published in upcoming papers. One noticeable domain in current cognitive research that invokes tremendous research interests, but nonetheless faces empirical difficulties as discussed in this section, is the domain of the language of smell (Majid, 2021). Applying the model to the domain of smell language is also part of the ongoing work.

for developing alternative language embedding algorithms. In recent years, embedding hierarchical relationships in hyperbolic geometry has become increasingly popular in NLP as well as computational neuroscience (Nickel & Kiela, 2017; Sala et al., 2018; Dhingra et al., 2018; Sharpee, 2019). Yet those inquiries on language embedding face a general issue: training or testing in most of the work heavily relies on artificially or narrowly determined training data or evaluating benchmarks such as WordNet. In the broader literature on hyperbolic embedding, beyond just the NLP domain, there have been proposals for hyperbolic embeddings that aim to encapsulate two dimensions like hierarchy and similarity and even multiple dimensions (Jiang et al., 2022; Kovacs et al., 2022). However, these works seldom establish the linguistic or cognitive relevance of their embeddings and measurements, and in particular have little relevance with evaluating cognitive semantic typicality and similarity¹². Further, beyond hyperbolic embedding, nonlinear dimensionality reduction is an important research field in machine learning in computer science. Hence, testing with other nonlinear algorithms on semantic relationships, such as Isomap and spectral clustering, is also performed. Current results using alternative nonlinear geometric methods, however, indicate no evidence that any of these methods could generate meaningful mapping for hierarchical or similarity relationships.

Empirically, future tuning and testing will mainly focus on the following aspects: effects of varying text genres and data sizes, varying context window sizes, polysemy, and testing in different languages. Moreover, extensions to other languages will require certain modulations,

¹² Nonetheless, whether some of these multi-dimensional hyperbolic embedding models can be related to the typicality-contextuality model developed in this chapter and be modified to model cognitive-driven semantic typicality and similarity should be an intriguing future research direction.

especially for semantic categories that are represented in non-alphabetic languages. For example, words in Chinese and Japanese are character-based. In NLP, a step of word segmentation is usually required for embedding algorithms to run on Chinese and Japanese texts. Initial evidence in modeling multiword expressions in English demonstrates that word segmentation can be incorporated into the typicality-contextuality model for modeling character-based language, such as Chinese and Japanese. Further work is needed for verifying the capability of the model for comprehensively mapping typicality and similarity relationships in languages other than English.

2.9. Chapter Summary

In this chapter, a typicality-contextuality modeling framework for embedding domain-oriented basic-level semantic categories is introduced. In contrast to conventional distributional contextuality-based methods, the "typicality-contextuality" embedding model simultaneously encapsulates both hierarchical (typicality versus idiosyncrasy) and similarity-based (contextual dominance versus contextual dependency) relations. This approach furnishes a fresh perspective for embedding and interpreting multiple cognitive-linguistic dimensions concurrently. Evaluations within this chapter, utilizing the domains of basic-level color and kinship semantics, along with further testing via a Mechanical Turk study and comparisons with empirical benchmarks, demonstrate the efficacy of this model in addressing sophisticated questions on domain-oriented semantic relationships, underscoring its theoretical and empirical potential.

The modeling and empirical works in this chapter lead to several broader contributions. First, they show that under proper cognitive-driven extraction strategy and metric, domain-specific semantic knowledge can be richly generated from local word co-occurrence patterns. Second, this chapter also demonstrates that typicality/idiosyncrasy judgments are not only sensitive to the semantic domain, but also dependent on the empirical measures, and are influenced by different linguistic-sociocultural contexts. Third, these works suggest that typicality and idiosyncrasy, usually considered as separate issues, are closely related linguistic phenomena with contrasting cognitive valencies. Fourth, these studies also indicate that the two aspects of semantic contextuality (as distributional similarity) contextual dominancy and contextual dependency may have domain, dynamic, and pragmatic implications. The studies on ancient Chinese corpus data in the next chapter will further illustrate these contributions.

Chapter III. Applying the typicality-contextuality model to investigate the diachronic change of basic-level semantics of color, smell, and shape-based classifiers in ancient Chinese

3.1. Chapter introduction

In this chapter, I leverage historical corpus data in ancient Chinese and the typicality-contextuality model developed in the previous chapter (Model I) to further investigate the research question concerning the diachronic change of domain-oriented semantic categories. I will perform further research not only on the extent to which the distributional embedding of Model I can generate domain-specific and cross-domain semantic knowledge at the word level, but also on how the model can be utilized to further understand the typological and temporal dynamics of such knowledge. The domains I focus on in this chapter are color, smell, and shape-based classifiers in ancient Chinese, for which the rest of the introduction and background sections will provide further justifications.

Before further discussing the motivations for utilizing historical data to achieve the research objectives, a more in-depth analysis of the question of domain-specificity in cognitive semantics than what was presented in the previous chapter is required. In cognitive linguistics and cognitive anthropology, the study of basic-level semantic expressions in natural languages has a prolific yet contentious history, with one primary methodology being the domain-specific

approach. This approach is in particular well illustrated by investigations into basic semantic terms in the color domain (Berlin & Kay, 1969).

Proponents of the domain-specific methodology for studying semantic categories view it as a productive and integrative research program (Regier et al., 2015). However, doubts have been raised regarding whether semantic domains like color are too specific and narrow, potentially limiting their generalizability for addressing broader structural and functional questions about human language (Lucy, 1997). Recently, there has been a notable extension of the color-naming research program in cognitive psychology, focusing on the language of smell (Majid, 2021). Nonetheless, such extensions may encounter theoretical and practical obstacles akin to those faced by color naming research, only with greater intensity. Psychologists and cognitive anthropologists have spent decades amassing empirical data on color naming, creating a substantial and comprehensive database encompassing color-naming data across hundreds of cultures and languages. But practically, it is implausible to fully replicate this extensive, long-term data collection process, given that many other semantic categories, the olfactory one in particular, appear to be more heterogeneously definable across languages and cultures in a domain-specific manner. There is ongoing debate on the extent to which studying the language of smell should account for the empirical difficulties of identifying cultures whose speakers exhibit varying levels of capacity in speaking about smells, and how to explain such discrepancies both across and within cultures and societies (Majid, 2021). Alternatively, some researchers attempt to address the puzzle by suggesting that the domain differences across various semantic categories, which are observed at least in modern English among other languages across color and smell naming (Majid, 2021), are crucially driven by perceptual and

neurobiological factors. They attempt to explain the problem of the language of smell through the properties of olfactory circuitry in the brain or the high dimensional and non-topological characteristics of the olfactory space in sensory processing (Olofsson & Gottfried, 2015; Cleland & Borthakur, 2020). Overall, extending the color-naming research paradigm to broader domains presents significant theoretical and practical challenges. These challenges raise important questions regarding the validity and effectiveness of the domain-specific methodology for exploring the semantic categories of natural language.

This chapter presents a series of studies that employ the typicality-contextuality model (Model I) to map domain-sensitive semantic knowledge in historical data. The rationale for examining historical corpus data extends far beyond testing the model's applicability alone. It is also strongly motivated by related fundamental theoretical questions that arise in both cognitive science and NLP literature. In Berlin & Kay (1969)'s seminal study on color semantics, the evolution of basic color terms in culture formed the crux of their research and predictions. According to their framework, specific domain-specific cognitive assumptions have particular implications for both synchronic and diachronic variations in color naming within cultural and historical dynamics. Regier et al. (2009; 2015) and Zaslavsky et al. (2018) further expand this framework into an information and communication-based cognitive theory and provide formal predictions on the cultural evolution of color naming.

The typicality-contextuality model and ancient Chinese historical data offers an alternative avenue for exploring domain-specific semantic changes and potential cognitive mechanisms that underpin these changes. Crucially, I argue that the domain-specific language

changes observed in ancient Chinese cannot be fully explained by the classical domain-specific theory of color naming cognition and the derivative theory of color semantic evolution (Berlin & Kay, 1969). Nor can they be accounted for by the information and communication-based domain-specific computational models based on these theories (Regier et al., 2009, 2015; Zaslavsky et al., 2018). A notable feature of the ancient Chinese corpus is its long and relatively stable history of semantic categories, allowing for the observation of long-term, diachronic effects of semantic competitions or integrations. Notably, current domain-specific semantics literature in cognitive science would struggle to account for many relatively stable patterns of multiple basic-level categories coexisting with the same or closely similar meanings in ancient Chinese. For the domain of color semantics, Chinese historical linguists have been examining such phenomena in at least three color categories ("red," "green," and "black"). Instead of attempting to force-fit these phenomena into the traditional domain-specific research on color, the results of this study suggest the need for both a new theory and new empirical methodology to analyze diachronic linguistic phenomena as observed in the diachronic changes of semantic categories in ancient Chinese.

In addition to closely examining the color domain, this chapter will also explore a subset of the semantic domain of smell in ancient Chinese, further connecting the contemporary literature on the language of smell in cognitive science to the investigation. The third semantic domain under investigation pertains to diachronic changes in the domain of shape-based classifiers. Shape-based classifiers represent a rewarding yet challenging research field for both cognitive science and linguistics (Lucy & Gaskins, 2001). Moreover, as shape-based classifiers

are not prevalent across most languages, they provide a particularly intriguing and valuable realm for linking cognition with computation and socio-cultural history.

The main empirical discoveries of the studies in this chapter are as follows: first, different from the basic-level categories of color in modern English and Chinese, which are primarily distributed along the typicality-idiosyncrasy axis (Figures 2.2 and 2.5 in Chapter II), the basic-level categories of color and shape in ancient Chinese are primarily distributed along the contextuality axis. Second, a substantial subset of basic-level categories of smell in ancient Chinese (fragrant/fragrance) co-evolved with a substantial subset of basic-level categories of color, which are primarily distributed along the typicality-idiosyncrasy axis. Third, basic-level color semantics in standardized modern Chinese is primarily distributed along the typicality-idiosyncrasy axis (which is different from color semantics in ancient Chinese), but major basic-level shape-based classifiers are more saliently distributed along the contextuality axis as ancient Chinese evolved to modern Chinese after the 20th century. Together, these results strongly indicate that domain specificities of semantic categories should be measured along at least two different linguistic or cognitive axes, through which the dynamics of synchronic and diachronic variations of semantic categories can be richly represented and mapped. The studies together have significant empirical and theoretical implications for investigating the domain-oriented and cross-domain properties of semantic categories in both cognitive science and NLP.

3.2. Background

The background section provides a closer literature review on three groups of literature: Research on using distributional and domain-specific approaches to link semantic change and cognition, the cognitive-driven research on historical change of basic-level color and smell semantics in ancient Chinese, and research on shape-based classifiers in Chinese in historical linguistics and cognitive science.

3.2.1. Distributional and domain-specific approaches linking semantic change and cognition

Traditional quantitative research in historical linguistics has usually concentrated on statistical analysis of well-defined linguistic features over time, such as single-word frequency, word collocation, mutual information, and term frequency-inverse document frequency (tf-idf), to name a few. Interestingly, the recent application of distributional NLP algorithms to explore diachronic and historical questions has become a burgeoning research direction in various computational social science and digital humanities subfields. The latest advancements in NLP, predominantly propelled by distributional algorithms like word embedding and deep learning algorithms, offer new computational tools for modeling linguistic features. This has engendered an emergent paradigm for diachronic linguistics. For instance, Hamilton et al. (2016a) utilizes the word embedding approach on four languages' corpora from the past 200 years, proposing a positive correlation between the rate of semantic change and polysemy, and a negative correlation with word frequency. Furthermore, word embedding algorithms have been employed to investigate diachronic changes of certain important language-related sociocultural phenomena,

such as gender and racial bias represented in large-scale corpus (Garg et al., 2016; Hamilton et al., 2016b; Koslowski et al., 2018; Figure 3.1).

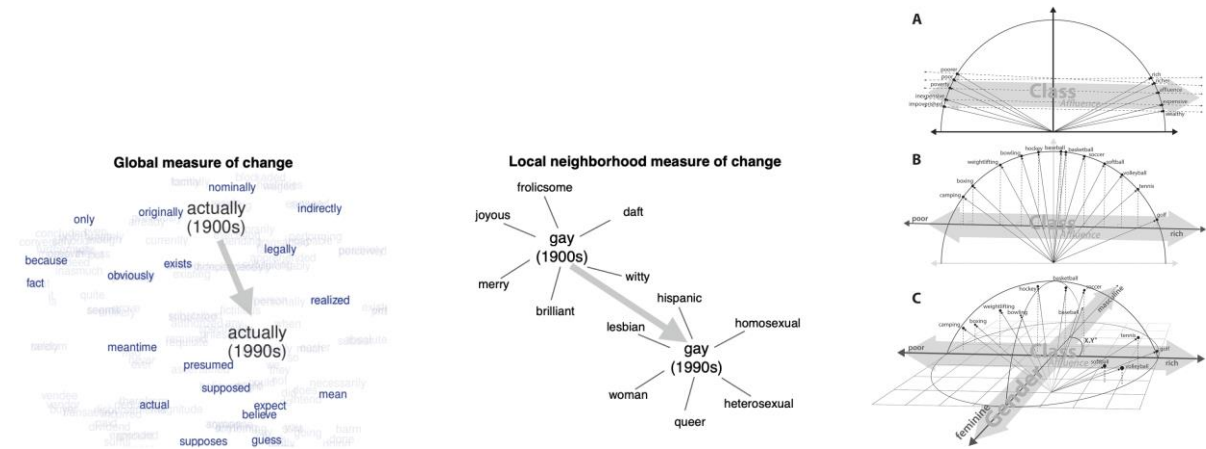


Figure 3.1. Left (adapted from Hamilton et al., 2016b): Hamilton et al. (2016b) demonstrate that the *word2vec* algorithm in the Euclidean space can capture two types of semantic change, which they argue as global and "linguistic" one and local and "cultural" one. For example, the top left figure shows an example of how the word "actually" underwent a global linguistic change during the 20th century, shifting from uses in objective statements about the world ("actually did try") to subjective statements of attitude ("I actually agree"). In contrast, the top right figure shows a local neighborhood measure of change measured by changes in a word's nearest neighbors, which they argue captures cultural shifts in core meaning, such as the word "gay"'s shift in meaning over the 20th century.

Right (adapted from Koslowski et al., 2018): Koslowski et al. (2018) show that the *word2vec* algorithm in the Euclidean space can capture multiple sociocultural dimensions such as class and gender. Bottom (A): the construction of a cultural dimension (class) in the Euclidean space via *word2vec*; bottom (B) the projection of words onto the Dimension; bottom (C) the simultaneous projection of words onto multiple dimensions (class and gender).

In another line of research, many psychologists consider the distributional framework to be beneficial for studying language cognition (Ettinger et al., 2016). However, it remains unclear how to integrate the distributional approaches used for examining language variations and changes and for modeling language cognition. Interestingly, it is the domain-specific research

that offers a relatively robust and sophisticated scientific framework linking categorization in cognition with its diachronic evolution (Berlin & Kay, 1969; Regier et al., 2015). According to this framework, language facilitates communicative efficiency through an "accuracy and complexity trade-off," allowing a language user to name and denote perceptual experiences to another language user. This hypothesis has been extensively validated by behavioral studies using color words from 110 languages, along with kinship and object shape naming patterns across multiple languages. More recently, Zaslavsky et al. (2018; 2020) employ the theory of the information bottleneck to further explore these predictions on the evolutionary patterns of color terms, initially proposed by Berlin and Kay (1969)'s classical research (Figure 3.2).

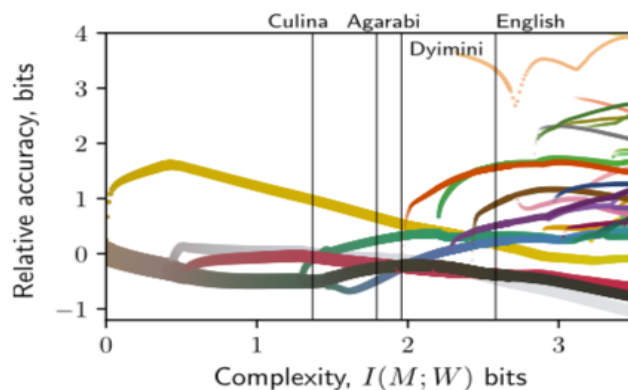


Figure 3.2. (Adapted from Zaslavsky et al., 2018) Bifurcations of the Information bottleneck (IB) color categories. The y axis shows the relative accuracy of each color category. Colors correspond to centroids and width is proportional to the weight of each category. Black vertical lines correspond to the IB systems. Zaslavsky et al. (2018) interpret that color systems in different languages correspond to different stages in this bifurcation process, such that the x axis has a historical implication.

Nonetheless, the aforementioned framework that connects color naming in culture with its historical dynamics assumes a specific functional aspect of the language -perception nexus. It

indirectly anchors cognition on language change: according to this framework, it is the synchronic variation in color terms across cultures that is mapped to diachronic predictions about their evolution over time. In this group of research, data from different languages in the contemporary world are used to assign each contemporary language to a certain "evolutionary stage" of color naming, which is expected to be further explained by the domain-specific hypothesis about cognition and communication. Moreover, critiques applicable to color naming research can also be applied to its derivative theory of historical dynamics, suggesting that such a research framework may presume an overly narrow functional conception of language meaning and empirical testing (Sanders, 1992; Lucy, 1997; see sections 2.2 and 2.8 in the previous chapter and section 5.5.1 in Chapter V for related discussions in this respect).

There are also other quantitative approaches to diachronic linguistics in the literature. Pioneering work addressing historical linguistic problems using mathematical and computational methods includes the application of techniques such as Bayesian statistics, dynamical systems, and game theory. Some of this work also implicitly assumes or explicitly proposes micro-level cognitive hypotheses for accounting for language change (Niyogi, 2006; Gell-Mann & Ruhlen, 2011; Nowak & Krakauer, 1999; Clark & Kimbrough, 2017; Griffiths & Kalish, 2007; see section 4.2.1 in Chapter IV for related discussions on this group of literature, and section 5.5.2 in Chapter V for further discussions on relating this literature to the current dissertation work). This chapter will focus specifically on merging the distributional algorithm developed in Chapter II (Model I) and domain-specific semantics to develop a new measuring methodology of semantic change from a cognitive-driven perspective. Chapter IV and V will develop a dynamical systems-based model (Model II) followed by another group of applications, examining how

metacognitive mechanisms may account for various dynamical scenarios of large-scale diachronic semantic changes.

3.2.2. cognitive-driven research on historical change of basic-level color and smell semantics in ancient Chinese

Remarkably, the evolution of color language in ancient China has emerged as a key research area in Chinese historical linguistics, an interest widely inspired by the domain-specific approach to semantics in cognitive science (Yao, 1988; Wu, 2012; Zhao, 2016). Ancient Chinese offers an enormously rich and relatively stable textual history spanning more than 2000 years. This forms a unique database for investigating the typological variations and diachronic changes of semantic categories, providing an invaluable corpus for scholars interested in the comprehensive study of semantic shift and development over time. It allows us to trace linguistic trajectories of individual words and semantic categories in socio-cultural contexts over time, as well as to investigate possible cognitive mechanisms that may shape or interact with these trajectories.

However, empirical research in Chinese historical linguistics on the language of color produces numerous observations that cannot be easily reconciled with aforementioned existing theories and models in the domain-specific framework of cognitive science. One such observation is that multiple groups of historical work on the diachronic changes of basic color words in ancient Chinese suggest the possibility of more than one basic term coexisting or competing to denote several prototypical colors (such as black and red) across different historical

periods (Yao, 1988; Wu, 2012). This finding challenges the conventional domain-oriented cognitive theories and models that typically operate under the assumption of a single dominant term for each prototypical color in a given linguistic community or culture.

Most cognitive-driven studies in Chinese historical linguistics on these matters, however, have not provided an advanced quantitative analysis beyond single-word frequency analysis supplemented by philological arguments. The richness and depth of the ancient Chinese corpus calls for more sophisticated methods and tools to quantitatively re-assess these linguistic phenomena.

Historical analysis on the semantic domain of olfaction in ancient Chinese is rare in the literature. But a few studies point towards certain intriguing properties of the smell language and its diachronic changes in ancient Chinese. Tong (1984) conducts a qualitative analysis arguing that the semantic system of smell in ancient Chinese shifted from a multipolar system in the early ancient period (before 2000 years ago) to a binary system around the early medieval period (before 500 CE). Given psychological work on studying the language of smell across cultures is heavily influenced by the color naming research, both synchronic and diachronic dimensions are important for such an investigation. This possible shift from a rich vocabulary in earlier times towards a sparser categorization system later on allows for a rare research opportunity in which the heterogeneity and variability of the languages of smell can be examined in a culturally and historically relatable condition within the same language family.

3.2.3. Shape-based classifiers in Chinese in historical linguistics and cognitive science

Intriguingly, shape-based classifiers in modern and ancient Chinese forms a special research area that interests both linguists (Chao, 1968, Tai & Wang, 1990, Chen et al., 2022) and cognitive psychologists (Lucy & Gaskins, 2001).

Shape-based classifiers are prevalent in most East Asian languages such as Chinese, Japanese, Korean, Vietnamese, Malay, Burmese, Thai, Hmong, and many South Asian languages, including Bengali and Munda. They also appear in Australian Aboriginal languages, specifically Yidiny and Murrinhpatha and languages of Mesoamerica like Classic Maya and its modern derivatives, such as Cuchumatán Mayan and Otomanguean (Aikhenvald, 2000; Hopkins, 2012)¹³.

Together, by incorporating the typicality-contextuality model developed in Chapter II to investigate domain-specific semantic questions in ancient Chinese corpus, we can glean valuable insights into important diachronic variations in a number of semantic domains. Quantitatively examining the long and rich language and cultural history of ancient Chinese can also inform and enhance the domain-specific methodology for studying semantic categories. This methodology will not only augment our understanding of important historical language changes in culture, but also shed light on the cognitive mechanisms that may drive such changes.

¹³ Further, languages that possess classifier-systems usually lack plural markers (Aikhenvald, 2000). Many European languages, in contrast, possess rich systems of plural markers but generally lack classifiers. These patterns raise an intriguing possibility that classifiers-systems and plural markers are linguistic phenomena that are in a cognitively tradeoff relation (Lucy & Gaskin, 2001). See Section 6.3 in Chapter 6 for modeling plural markers in modern American English using Model I.

3.3. Methods

The model that is used in this chapter is derived from the typicality-contextuality model proposed in chapter II. The core component of the model is a new distributional metric for measuring and mapping the correlative relationship of word co-occurrence patterns extracted from corpus. It is constituted by the target word's conditional entropy upon its co-occurrences with its context words, weighted by two extra parameters: the target word's collocation diversity and the ranking of the conditional entropy of each pair of word co-occurrences for every target word. The model first extracts words that are semantically most similar to the chosen target word. It then generates measurements of semantic typicality and similarity for extracted target words and maps these relationships in the hyperbolic coordinates. Results in Chapter II shows that based on corpus of relatively small size (~ millions of word tokens), the model is capable of extracting basic color terms in modern English and Chinese from the wikipedia corpus of each language, and of mapping the typicality relationships of these terms in the two-dimensional hyperbolic space (Figures 2.2 and 2.5) .

In order to address the possibility of multiple basic terms coexisting for "black," "red," and "green," as proposed by Chinese historical linguists (Yao, 1988; Wu, 2012), this study in particular focuses on the words of these color domains. For selecting basic color terms of each ancient period, the model first extracts words that are semantically closest to the chosen target

word. Similar to what is shown in Table 2.2 (1st column) for the modern English Wikipedia corpus, for the ancient Chinese corpus, 15 words that are mapped as most similar to the target word are chosen as potential basic color terms during each period. Further testing demonstrates the effectiveness of the algorithm for identifying semantic similarity on the ancient Chinese corpus, as the 15 chosen words for each target word in the results are predominantly either color words or related object nouns (or occasionally related verbs) that are semantically relevant to the target word during the specific historical period.

The typicality-contextuality model first embeds a chosen word in the two-dimensional Euclidean space, in which the x -axis represents the expected value of the sigmoid function of each target word's weighted conditional entropy, and y -axis represents the expected value of the sigmoid function of the context's weighted entropy conditional on the target word (see Section 2.3 In chapter II for details). The Euclidean coordinates are then transformed into the hyperbolic coordinates (u, v) , where $u = \ln \sqrt{\frac{x}{y}}$ and $v = \sqrt{xy}$ (Section 2.3.2). Empirical testing based on domain semantics in color and kinship of modern English and Chinese (Sections 2.4 and 2.6 of Chapter II) demonstrates that the u dimension forms an effective measure of semantic similarity in terms of contextual dominancy and contextual dependency, and the v dimension forms an effective measure of typicality and idiosyncrasy.

Under the hyperbolic coordinates of similarity and typicality, ancient Chinese words of "black," "red," and "green" in each historical period are compared across periods. Words of "brown," "blue," and "purple" are also compared, as previous work in historical linguistics

predicts different typicality rankings for these terms from the typicality sequence proposed in the color naming research in cognitive science.

For the language of smell, the model nominates five terms for "fragrance/fragrant" in ancient Chinese and maps them in terms of typicality and similarity relationships. Previous work in Chinese historical linguistics predicts that in ancient Chinese, the semantic system of smell developed from a multipolar system to a binary system before the early medieval period (Tong 1984). Due to the extremely small size of the textual corpus for periods before the early medieval era, the model cannot directly address this proposal about the significant historical transition of the language of smell in ancient Chinese. Nonetheless, focusing on the diachronic changes of the five words of "fragrant/fragrance" from the early medieval period to the early 20th century, this study can address how the binary system of smell language was strengthened and further demarcated across these times.

The early medieval periods before the Tang dynasty only contain poetry corpus with much smaller sizes than in later periods (around 0.32 M word tokens for the early Six Dynasties period (220 to 420), and around 0.45 M word tokens for the late Six Dynasties period (420 - 589)). Nonetheless, since these periods are treated as among the most important eras for understanding the change of ancient Chinese language in historical linguistics (Yao, 1988, Zhao, 2016), this study also models color and smell words in these periods despite the possible size issue. Nevertheless, perhaps because color and smell vocabulary was rich and complex during these two periods, the results echo many observations made by historical linguists on the change of color naming during these times. A robustness check with larger corpus size, however, is

performed on corpus in later periods (see Section 3.6 discussion), and the results show that substantially increasing corpus size alone does not affect most results in these studies.

3.4. Results on basic-level basic color and smell semantics

Informed by both the color naming and philological research on color words in ancient Chinese, Yao (1988) proposes a first general outline for the evolution of basic color terms from ancient to modern Chinese, in which there were more than one basic terms for black and red in certain early medieval periods (*xuan* and *hei* for black before the medieval period, and *hong* and *chi* for red during the Six-Dynasties period (220 - 589); Table 3.1). On the other hand, Zhao (2016) surveys 68 single-character words for black and 33 single-character words for red in medieval Chinese, but it does not differentiate them in terms of basic and non-basic color terms.

	Before the medieval period (220 CE)	Six Dynasties (220 - 589 CE)	Tang and Song Dynasties (618 - 1279 CE)	Up to the Modern period
Black	<i>xuan, hei</i>	<i>hei</i>	<i>hei</i>	<i>hei</i>
White	<i>bai</i>	<i>bai</i>	<i>bai</i>	<i>bai</i>
Red	<i>chi</i>	<i>chi, hong</i>	<i>hong</i>	<i>hong</i>
Yellow	<i>huang</i>	<i>huang</i>	<i>huang</i>	<i>huang</i>
Blue/Green	<i>qing</i>	<i>qing</i>	<i>qing</i>	
Green	<i>lv</i>	<i>lv</i>	<i>lv</i>	<i>lv</i>
Blue			<i>lan</i>	<i>lan</i>
Purple	<i>zi</i>	<i>zi</i>	<i>zi</i>	<i>zi</i>
Pink	<i>hong</i>	<i>hong</i>		
Gray		<i>hui</i>	<i>hui</i>	<i>hui</i>
Brown			<i>he</i>	<i>he, zong</i>
Orange				<i>cheng</i>

Table 3.1. In historical linguistics, Yao (1988) makes the first explicit proposal about the diachronic change of the basic color words in ancient Chinese. The proposal raises the possibility that in ancient Chinese, there could be more than one basic term of the same color categories that coexisted, such as black (*xuan* and *hei*) before the medieval period or red (*chi* and *hong*) during the early medieval (the Six-Dynasties) period. Most studies in cognitive science on color naming, such as Berlin & Kay (1969) among their most following works, do not address such a possibility. Results from the typicality-contextuality modeling aims to refine and further develop Yao’s proposal (See *Results*). The table is adapted from Yao (1988).

Using ancient Chinese poetry corpus from the early medieval to the last imperial dynasties (220 - 1911), the typicality-contextuality model characterizes the diachronic change of semantic similarity for multiple words as the candidates of basic color terms in ancient Chinese. Table 3.2 summarizes the results for the domains of black and red. The model predicts that there were respectively three words that coexist or compete with one another as basic terms for black (*hei, xuan, wu*) and for red (*hong, zhu, chi*), whereas Yao (1988) only indicates two for each (*hei* and *xuan* for black, and *hong, chi* for red; Table 3.1). Further, the model also shows different

timing for possible coexistences or competitions of the basic terms for the same color domains: in the poetry corpus, *hei*, the single basic term for black in modern Chinese, appears to not have become the predominant basic term for black even until the last imperial dynasty (1644 -1911), and *zhu* rather than *chi* appears to be the most salient competitor with *hong*, the single basic term for red in modern Chinese, the stable coexistence of which is shown to maintain at least until the Tang dynasty.

	Early Six dynasties (Wei and Jin; 220 - 420)	Late Six dynasties (Nanbei; (420 - 589)	Tang dynasty (618-907)	Song dynasty (960-1279)	Ming dynasty (1368-1644)	Qing dynasty (1644-1911)	Contemporary (wikipedia)
Black: <u><i>HEI</i></u>	<i>(The target word is not available during this period)</i>	<u><i>xuan</i></u> (7), <i>bai</i> (20)	<u><i>chi</i></u> (1)	NONE	NONE	<u><i>chi</i></u> (1)	<u><i>HONG</i></u> (1), <i>bai</i> (2), <i>lan</i> (3), <u><i>chi</i></u> (5), <i>huang</i> (6), <u><i>wu</i></u> (7), <i>lv</i> (14), <i>qing</i> (16)
Black: <u><i>xuan</i></u>	NONE	NONE	NONE	NONE	NONE	NONE	NONE
Black/ deep gray: <u><i>wu</i></u>	<i>jin</i> (2), <i>qing</i> (5), <u><i>zhu</i></u> (12), <i>bai</i> (17)	NONE	<u><i>HEI</i></u> (26)	NONE	<i>bai</i> (9)	NONE	NONE
Red:	<i>lv</i> (4),	<i>cui</i> (6),	<i>lv</i> (4),	<i>lv</i> (9),	<i>lv</i> (9),	<i>lv</i> (3),	<i>lan</i> (1),

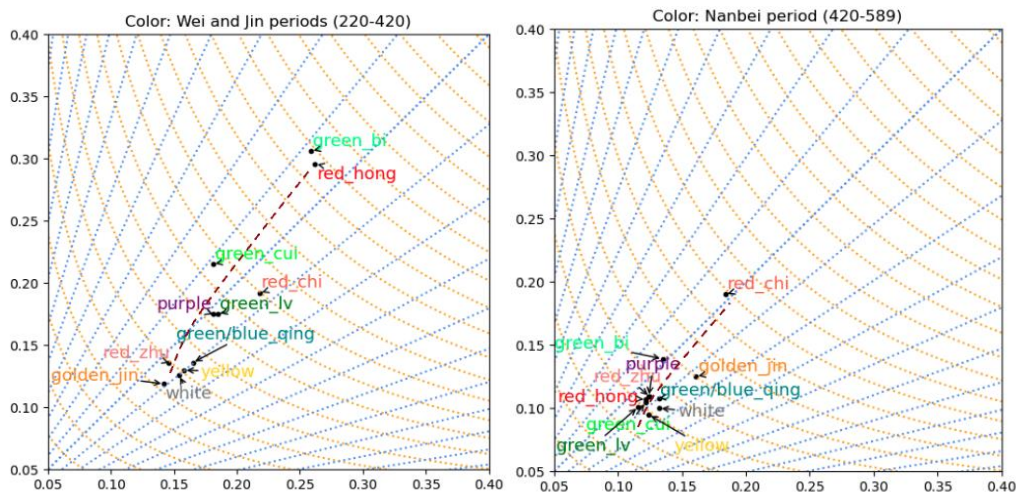
<u>HONG</u>	<i>qing(5), <u>zhu</u>(7), jin(12), huang(13)</i>	<i>lv(9), zi(18)</i>	<i>cui(6), bi(5), zi(20)</i>	<i>cui(14)</i>	<i>bi(12), cui(13), huang(19)</i>	<i>yin(12), bi(13), cui(20)</i>	<u>HEI</u> (3), <i>lv(6), bai(7), huang(9), qing(11), zi(20)</i>
Red: <u>zhu</u>	<i>jin(6), lv(14), bai(15), huang(24)</i>	<i>lv(1), cui(3), <u>HONG</u>(7), zi(19)</i>	<i>cui(4), yin(6), zi(7), <u>HONG</u>(14)</i>	<u>HONG</u> (9), <i>qing(18)</i>	<u>HONG</u> (10), <i>cui(21)</i>	<u>HONG</u> (9)	NONE
Red: <u>chi</u>	<i>bai(4), huang(19) qing(20)</i>	<i>zi(7)</i>	<u>HEI</u> (1)	<i>bai(10), <u>HEI</u>(13)</i>	<i>zi(10)</i>	<i>hei(1)</i>	<i>lv(19)</i>

Table 3.2. Based on ancient Chinese poetry corpus across six historical periods, candidate basic terms for black (*hei, xuan, wu*) and red (*hong, chi, zhu*) are extracted from corpus and ranked in terms of linguistic similarity by the typicality-contextuality model. Possible basic color words are selected manually as target words, based on the historical linguistic research on color terms in ancient Chinese (Yao, 1988; Wu, 2012; Zhao, 2016), then the algorithm extracts 20 words that are semantically closest to each target word. Possible basic words of black and red are underlighted. Words that become the basic terms of red (*hong*) and black (*hei*) in modern Chinese are both underlighted and capitalized. Candidates of other basic color terms in the table besides red and black are: *bai* (white), *huang* (yellow), *qing* (green-blue), *lv* (green), *bi* (green), *cui* (green), *zi* (purple), *lan* (blue), *jin* (golden), *yin* (silver). For comparison, the last right column includes results of modern Chinese color words that are mapped as linguistically similar to the target black and red terms, based on a randomly selected Wikipedia corpus in modern Chinese of 8 M token size.

Based on words that are selected in the similarity ranking, the model maps the candidate basic color terms in the hyperbolic coordinates for studying their typicality -similarity relationships. Importantly, the mapping shows that long after the early medieval periods (220 - 589) for which works in historical linguistics argue for within-category color semantic

competitions, basic-level color characters predominantly align on the contextuality dimension and possess similarly high degrees of typicality, showing evidence of persistent semantic competitions through more than 1000 years up to the early 20th century (marked by dark red dotted lines in subplots beginning from the Tang dynasty in Figure 3.3).

More specifically, for individual color categories, the mapping shows that *hong* and *zhu* were the two words of red ranked with highest semantic typicality, the pattern of which was persistent from the Tang dynasty (618 -907) to the end of the imperial era (1911) (Figure 3.3). The model also maps a cluster of green and green-blue basic terms (*lv*, *cui*, *bi*, *qing*), which may begin to form in the Nanbei period (420 -589) and stay relatively stable until the end of the imperial era (1911). This variability within the green color is not in Yao (1988)'s prediction.



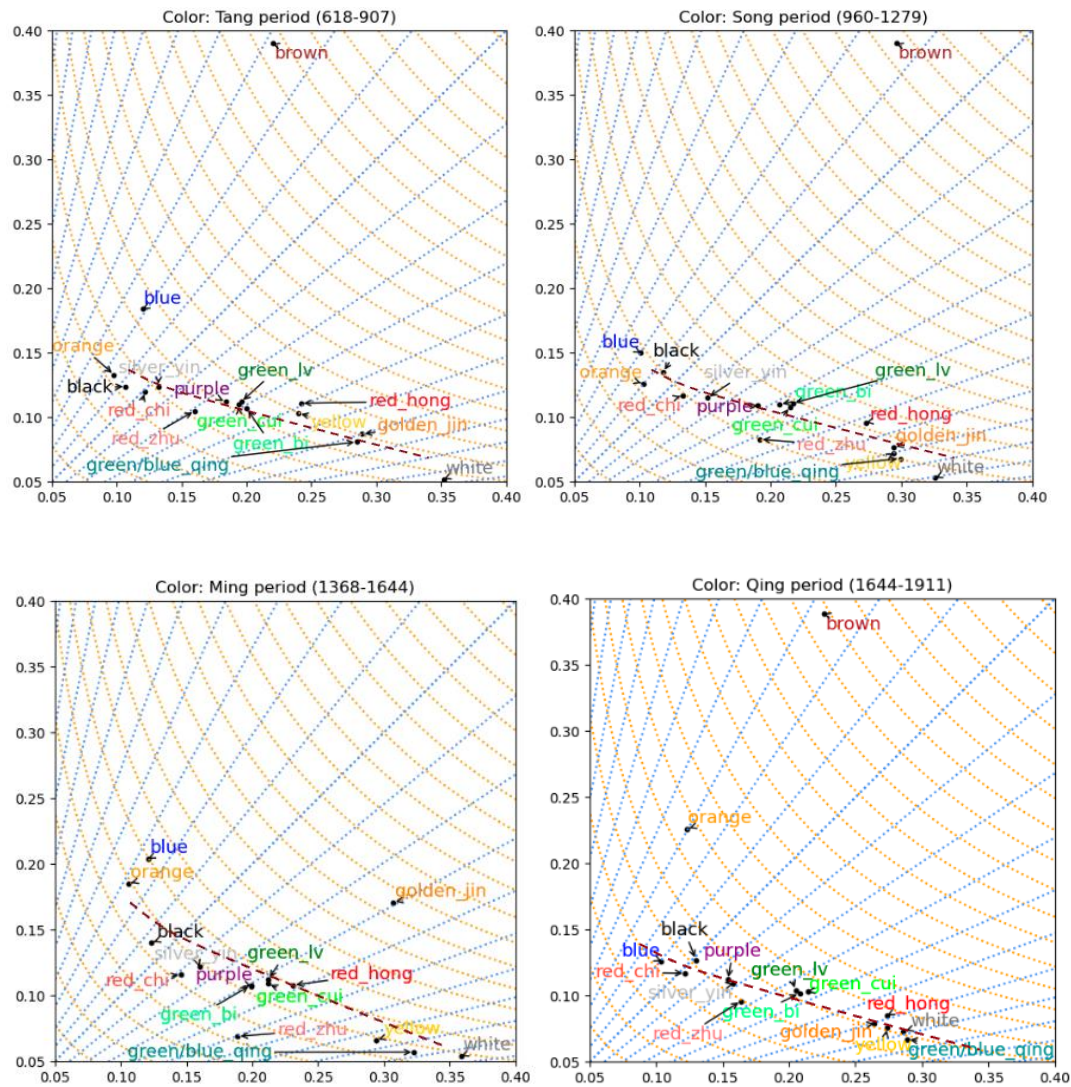


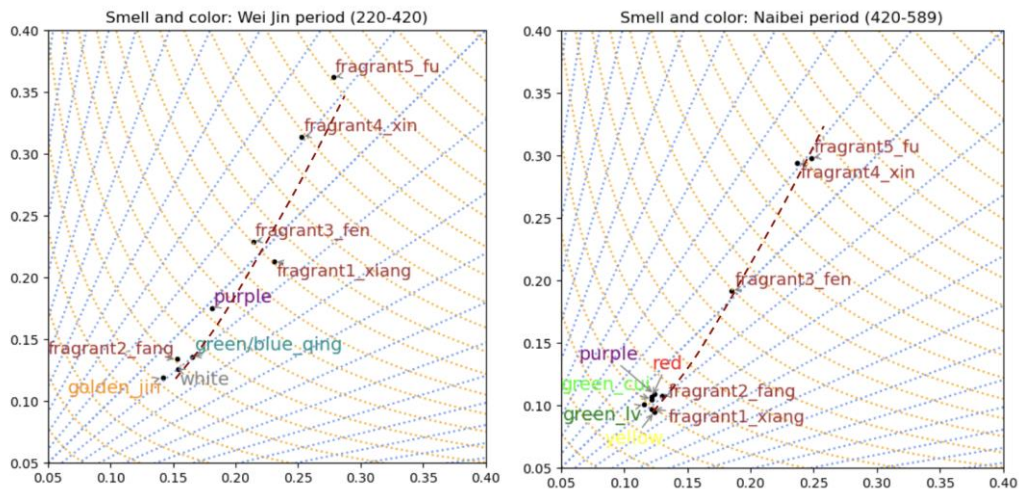
Figure 3.3. The typicality-contextuality mapping of candidate basic color terms in ancient Chinese across six main historical periods in ancient China. Beginning from the Tang dynasty (618 - 907), basic-level color characters predominantly align on the contextuality dimension and possess similarly high degrees of typicality (marked with dark red dotted lines).

Other specific observations on color domains besides black, red, and green include the follows:

- (i) The modeling results agree with Yao (1988) on the word purple (*zi*): both suggest that purple appeared early as a basic color term and was stably so throughout ancient Chinese history. My results further suggest that purple shows higher typicality than in both modern Chinese and English.
- (ii). The modeling results agree with Yao (1988) that the word blue (*lan*) appeared relatively late compared to most other basic color terms in modern Chinese, but it shows that *lan* may not be used as a stable basic color term at least in poetry even at the end of imperial history (1911), much later than Yao (1988) predicts.
- (iii). The modeling results suggest that golden (*jin*) and silver (*yin*) may be candidates for basic color terms at various periods in ancient China. Though these results are largely suggestive at the current stage of the study, these possibilities are consistent with the late appearances and unstable statuses of orange (*cheng* or *ju*) and gray (*hui*) as basic color terms in modern Chinese (Sun & Chen, 2018), which *jin* and *yin* may have competed with respectively.

For the domain of smell, the model selects five words of fragrance in ancient Chinese and creates the typicality- similarity mapping for them. A procedure similar to the selection of candidate basic words of color is performed for the language of smell, which points to five words of fragrance (*xiang*, *fang*, *fu*, *fen*, *xin*). The results demonstrate that there are two dominant basic terms (*xiang* and *fang*) for fragrance in ancient Chinese (*xiang* is the single basic term for fragrant smell in modern Chinese), while three other selected terms (*fu*, *fen*, *xin*) were gradually marginalized after the Tang dynasty (618 -907) (Figure 3.4). Interestingly, the typicality-

contextuality mapping shows that the most predominant words for fragrant smell semantically resemble multiple dominant basic color terms in ancient Chinese, which together align along the contextuality axis (marked in dark dotted lines in Figure 3.4). Less dominant smell terms, on the contrary, increasingly align along the typicality axis. This result suggests an intriguing possibility about the language of smell in ancient Chinese: though ancient Chinese has a relatively richer vocabulary for describing fragrant smell, such richness may not be a domain-specific property per se; in particular, the two most dominant terms for fragrant smell in ancient Chinese had increasingly aligned with multiple main basic color terms since the medieval period to the early 20th century.



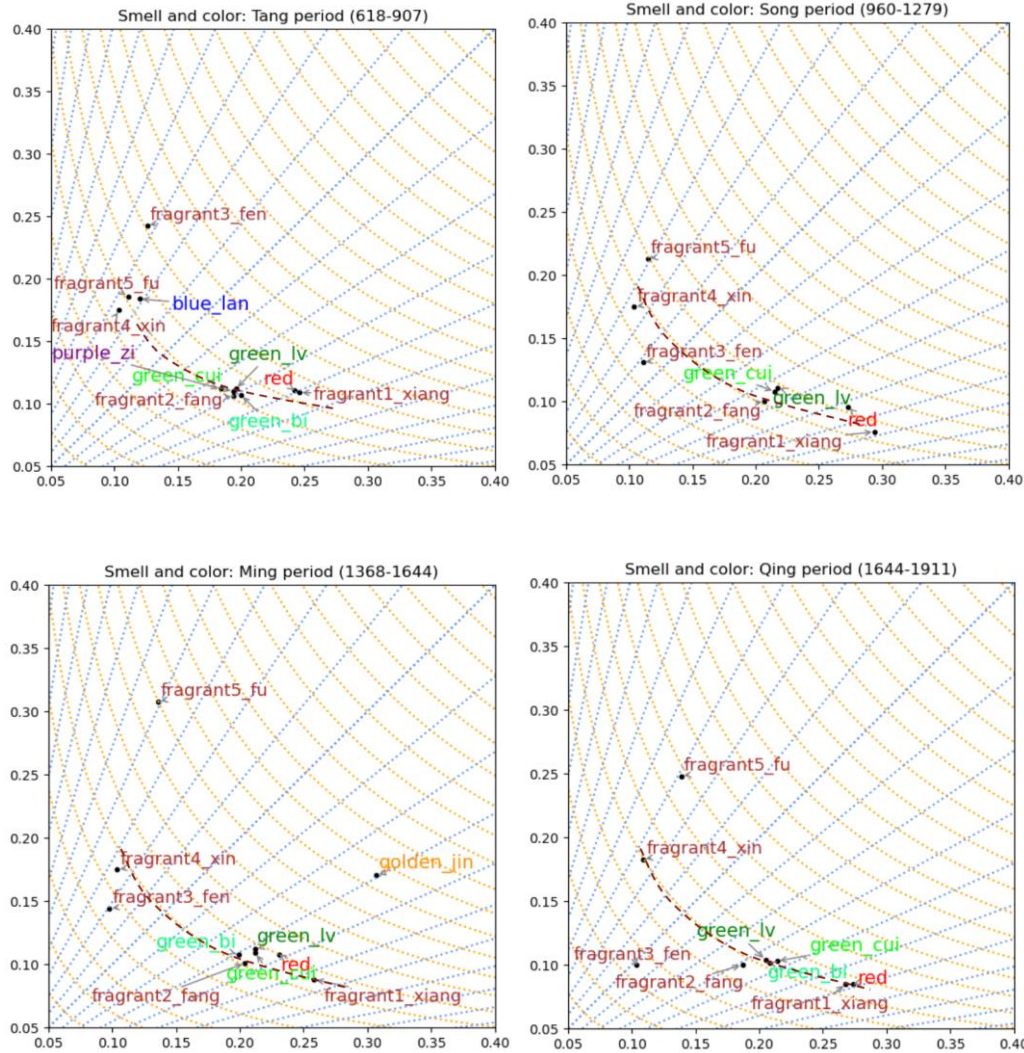


Figure 3.4. The typicality-contextuality mapping of five candidate basic terms (*xiang*, *fang*, *fen*, *fu*, *xin*; all shown in deep red fonts) for fragrance in ancient Chinese across six main historical periods in ancient China. Since the Tang dynasty, the two most dominant basic terms for fragrance (*xiang*, *fang*) predominantly aligned with color words than with other three increasingly marginalized terms for fragrance (*fen*, *fu*, *xin*) (marked with dark red dotted lines). Crucially, the dominance of *xiang* and *fang* as main basic-level terms for fragrance is better measured by high contextual dominant rather than high typicality. Additionally, it is also intriguing to observe that in all these periods, multiple basic color terms are consistently extracted and mapped to be semantically close to the smell terms.

In short, the typicality-contextuality model provides a semi-supervised procedure to select basic words for the semantic categories of color and smell, and geometrically map their relationships in terms of semantic similarity and typicality. Further, the model also gives predictions for possible cross-domain relationships between semantic categories in color and smell. Results show that within-category competitions in the color domain may be best captured along the contextuality dimension, whereas basic-level smell words are more stably distributed along the typicality-idiosyncrasy axis. Informed by the rich work on color and smell naming in cognitive science, psycholinguistics, and historical linguistics, the typicality-contextuality model offers a new quantitative framework for computationally addressing both domain-specific and cross-domain theories of semantic categories, as well as proposing new predictions or directions for empirical hypothesis and testing within and across the color and smell domains.

3.5. Results on shape-based classifiers

Results on the evolution of shape-based classifiers depict a similarly nuanced yet different kind of evolutionary picture. The similarity algorithm in the model extracts three groups of shape-based classifiers from the ancient Chinese poetry corpus: classifiers derived from one or two dimensional shapes, classifiers derived plant/flower-related shapes, and classifiers derived from cup/bowl related shapes (Table 3.3). Then the typicality-contextuality mapping traces the evolution of the three groups of classifiers in the poetry corpus from the early medieval era to the

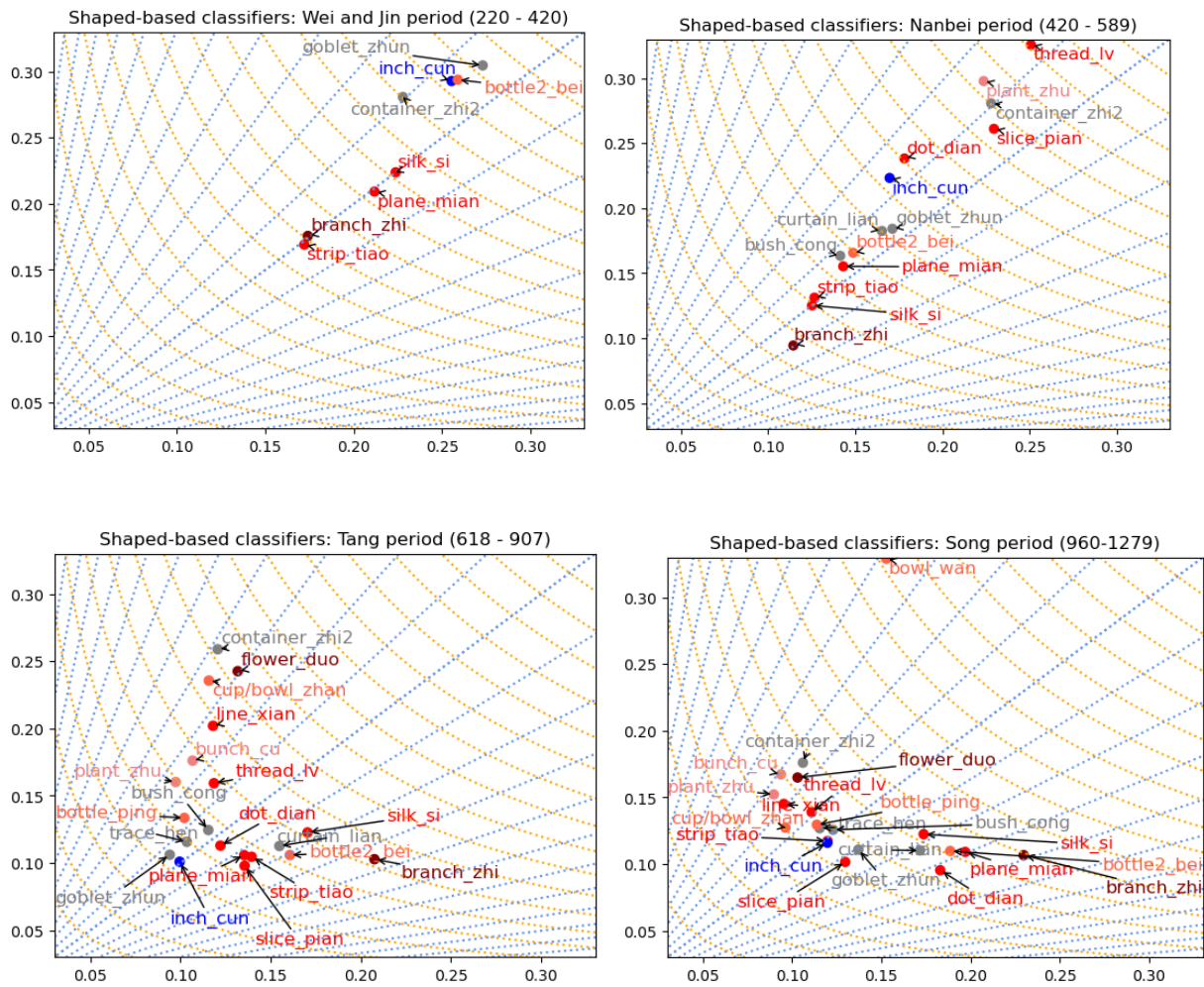
early 20th century. Methods and procedures are similar to the ones for examining the color and smell semantics represented in the last section (Section 3.4).

Shape nouns/classifiers derived from one-to-two dimensional shapes	tiao "strip," pian "slice," mian "plane," <u>dian</u> "dot," <u>xian</u> "line," <u>si</u> "silk," <u>zhi</u> "branch," <u>cun</u> "inch," <u>lv</u> "thread," <u>hen</u> "trace," <u>lian</u> "curtain"
Shape nouns/classifiers derived from plant/flower shapes	duo "flower," <u>zhu</u> "plant," <u>zhi</u> "branch," <u>cong</u> "bush," <u>cu</u> "bunch"
Shape nouns/classifiers derived from cup/bowl shapes	zhan "cup/bowl," duo "flower," <u>ping</u> "bottle," <u>wan</u> "bowl," <u>bei</u> "bottle2," <u>zhi2</u> "wine container," <u>zun</u> "goblet"

Table 3.3. Three groups of shape-based nouns and classifiers extracted by similarity algorithm of the model from the poetry corpus in ancient Chinese over 1500 years (up to early 20th century). Almost all shape-derived classifiers in ancient Chinese are also shape nouns. **Bold:** words that function as classifiers only (but not as shape nouns anymore) in standardized modern Chinese. Underline: shape nouns that are not necessarily classifiers in ancient Chinese, but function as both shape nouns and classifiers in standardized modern Chinese.

Table 3.3 presents the results of three categories of shape-based nouns and classifiers extracted by the similarity algorithm in the model from a 1500-year ancient Chinese poetry corpus. In the table, words in **bold** function solely as classifiers but not as shape nouns in standardized modern Chinese, which suggests that these terms may have shifted from nouns or nouns/classifiers in ancient Chinese to classifiers alone in modern Chinese, experiencing a substantial process of grammaticalization (Hopper & Traugott, 2003; Wu, 2004). Whereas words that are underlined have experienced a moderate degree of grammaticalization (evolving from nouns in ancient Chinese to nouns/classifiers in modern Chinese, or maintaining the dual

noun/classifier functions across ancient and modern Chinese). Further, in figures 3.5 and 3.6, words in gray are extracted by the similarity algorithm as semantically close to other shape nouns or classifiers in ancient Chinese, but they do not function as classifiers in modern Chinese. Additionally, words in dark red share cross-domain properties across more than one of the three shape domains, and a word in blue forms a unit of length derived from the corresponding shape-based noun/classifier in modern Chinese.



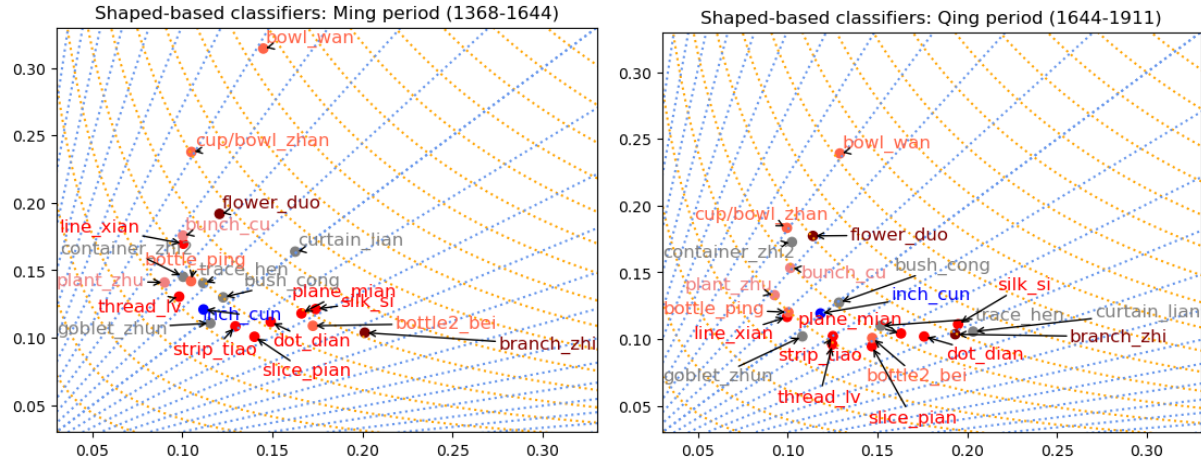


Figure 3.5. This figure provides a visual representation of the diachronic changes for three groups of shape words and shape-based classifiers in ancient Chinese over six historical periods: shape nouns/classifiers derived from one-to-two dimensional shapes (in red), shape nouns/classifiers derived from plant/flower shapes (in light coral red), shape nouns/classifiers derived from cup/bowl shapes (in tomato red). During the two early medieval times (220 - 420 and 220 - 589), all three groups of shape-based terms aligned on the typicality- idiosyncrasy axis (from bottom left to top right) with little differentiation. Differentiation increased starting from the Tang dynasty (618 - 907): the group related to one-to-two dimensional shapes increasingly shifted towards aligning along the contextuality axis, and differentiation among the three groups increased. Some exceptional cases are also color coded: *duo* ("flower") and *zhi* (branch) are extracted to belong to more than one group (in capital and dark red), *chi* ("inch") becomes a length unit in modern Chinese (in blue), and *zhi2* ("container"), *zun* ("goblet"), *hen* ("trace"), *cong* ("bush"), *lian* ("curtain") are not classifiers but still shape-related words in modern Chinese (in gray).

Then the typicality-contextuality mapping provides the following characterizations for the diachronic changes of shape words and shape-based classifiers:

- (i). In early medieval periods, three groups of shape-based words are observed to align on the typicality-idiosyncrasy axis. Moreover, words in the three groups are not in particular differentiated in the mapping. This pattern began to shift after the early medieval period during the Tang dynasty (618 - 907) (Figure 3.5, top and middle left).

- (ii). From the Song dynasty period (960 -1279) to the last imperial period, Qing dynasty (1644 -1911), the group related to one or two dimensional shapes increasingly aligned along the contextuality axis. Moreover, words of the three groups are increasingly, but not completely, differentiated in the mapping (Figure 3.5, middle right and bottom).
- (iii). In the *Wikipedia* data of standardized modern Chinese, the group of words related to one-to-two dimensional shapes are completely differentiated from the plant/flower-shaped and cup/bowl-shaped groups of words. While the first group of words predominantly align on the contextuality dimension, the other two groups of words predominantly sign on the typicality-idiosyncrasy axis. In addition, the words related to one-to-two dimensional shapes are predominantly contextually dominant (red in Figure 3.6), but the plant/flower-shaped and cup/bowl-shaped words are predominantly contextually dependent (light coral red and tomato red in Figure 3.6). Importantly, words that are classifiers derived from one-to-two dimensional shapes in modern Chinese (*tiao* "strip," *pian* "slice," *mian* "plane," *dian* "dot," *xian* "line," *si* "silk") are highly typical and contextually dominant, but they are distributed relatively widely along the contextuality axis (red in Figure 3.6). Words that are derived from one-to-two dimensional shapes but do not function as classifiers in modern Chinese (*hen* "trace," *lian* "curtain") are moderately or highly idiosyncratic as well as highly contextually dependent, but they are distributed relatively widely along the typicality- idiosyncrasy axis (gray in Figure 3.6).

(iv). Shape words that function as classifiers alone in modern Chinese tend to have high or moderate degrees of typicality, with more generic, classifiers derived from one or two dimensional shapes (*tiao* "strip," *pian* "slice," *mian* "plane") having high contextual dominancy (red and capital in Figure 3.6), and less generic, classifiers derived from plant/flower or cup/bowl shapes (*duo* "flower," *zhu* "plant") having high degrees of contextual dependency (dark red and capital and light coral red and capital in Figure 3.6)¹⁴.

Overall, the historical analysis on shape words and shape-related classifiers from early medieval to modern Chinese show gradually differentiated patterns of classifiers. The group of classifiers related to one-to-two dimensional shapes is identified to be increasingly more typical *and* distributed along the contextuality dimension. In contrast, classifiers that evolved predominantly from plant and utensil shapes are identified to be increasingly less typical *and* aligned within the typicality-idiosyncrasy dimension. Intriguingly, the results demonstrate that whether a shape-related term has a high degree typicality and contextual dominancy is not primarily correlated with whether it is exclusively a classifier but not a shape noun, but primarily dependent on whether it is derived from the group of words related to one-to-two dimensional shapes. This provides sets of new types of hypotheses and predictions for future computational and empirical research in cognitive science on classifiers.

¹⁴ Of all terms examined in this study, *cun* "inch" is the lone classifier derived from one-to-two dimensional shapes exhibiting high contextual dependency. However, in its role as a classifier in modern Chinese, *cun* primarily serves as a unit of length (corresponding to the English "inch"), which likely explains its exceptional status.

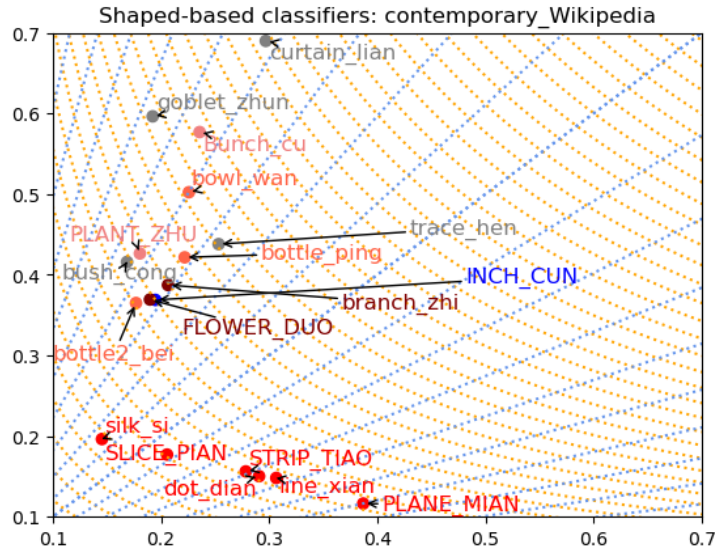


Figure 3.6. Classifiers and shape-words derived from ancient Chinese are mapped based on wikipedia corpus in standardized modern Chinese. In standardized modern Chinese, words related to one-to-two dimensional shapes (in red) are almost entirely distinct from plant/flower-shaped words (in light coral red) and cup/bowl-shaped words (in tomato red), with the former aligning predominantly on the contextuality dimension and the latter two on the typicality-idiosyncrasy axis. Further, classifiers derived from one-to-two dimensional shapes are highly typical and contextually dominant, and among most generic classifiers in modern Chinese, while those not functioning as classifiers are more idiosyncratic and contextually dependent (in gray). Lastly, words related to one-to-two dimensional shapes functioning solely as classifiers in modern Chinese tend to exhibit high typicality and contextual dominance (in RED CAPITAL), and plant/flower or cup/bowl-derived shape words functioning solely as classifiers are nonetheless exhibit lower typicality and more contextual dependency (in CAPITAL with non-red colors).

3.6. Discussion

One question this chapter does not explore is polysemy. Multiple words that are being tested in this study have more than one major meaning in one or more specific historical periods. Nonetheless, previous work in NLP shows that polysemy may not be a fundamental issue for

every embedding problem, while admitting the algorithm can be further improved by adding extra modules, once an embedding algorithm is demonstrated to perform above the task-specific baseline.

Another aspect that is not in particular addressed in the implementation sections is the problem of data size. Corpus in early medieval periods are comparatively small, and though the sizes of poetry corpuses in the later three periods are far beyond the minimum requirement for Model I to generate embedding with good quality (as shown in the convergence testing in section 2.4.3 in Chapter II), the whole poetry in these three periods are not used in this study in order to keep the modeling for later periods in the same scale with early periods (see Section 3.3 “Methods” for details). Nonetheless, robustness testing is performed using more than twice the amount of textual data from the poetry corpus of the Ming dynasty (token numbers = 10840439, compared to the size of the corpus used in actual simulations in this chapter (token numbers = 4459098). Results show that computation using more than the corpus of more 10 M word tokens from the same period does not change the overall results (Figure 3.7).

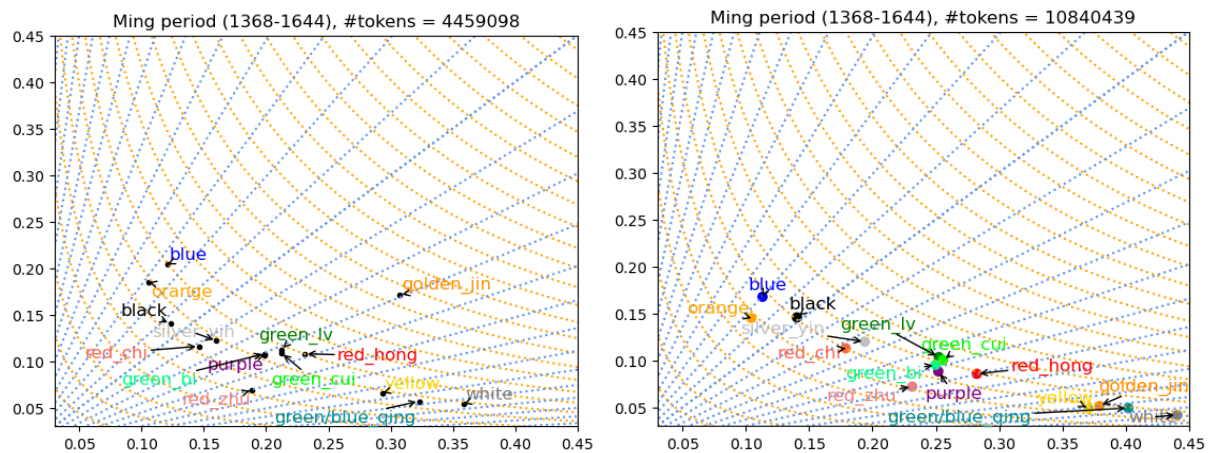


Figure 3.7. Simulations with more than twice the amount of word tokens for the poetry corpus in the Ming period (1368 - 1644) does not entail significant changes for the overall results.

Further, this study only uses the poetry corpus and leaves the problem of cross-genre variation unexplored. Although it is intriguing for future research to explore the cross-genre variability, by computationally addressing how much certain differentiations between the basic and non-basic semantic terms are robust or variable across genres and languages, choosing ancient poetry corpus has the following justifications. One practical justification is that ancient Chinese poetry is known to include rich usages of color and fragrance, hence in practice, results could be more robust than using non-poetry data, especially for corpus of small size. Focusing on ancient Chinese poetry data also has sociolinguistic and cognitive justifications.

Sociolinguistically, the written ancient Chinese have a long and relatively stable linguistic and cultural history for over 2000 years, and poetry in ancient China was not simply a written culture for entertainment among small social groups, but was a basic literary and oral genre in common education and ordinary social life (Owen, 1986). Cognitively, one assumption of the domain-specific approach to the problem of semantic categorization in cognitive science is that basic terms are relatively invariant across corpus and genres (but not across times). Indeed, results in this chapter show the robustness of such a formulation of basic semantic categories, as basic color terms are shown to be relatively stable across languages (Modern Chinese and English) as well as across genres (*Wikipedia* and poetry corpus).

Finally, it is also necessary to consider the embedding method applied in this study along with the deep neural network model, which has become the most dominant type of model in NLP

during the past decade. The reasons that deep neural network models are not used in this study is that these models are crudely cognitive driven at most, and these models usually require billions of word tokens for training, while only being able to provide relatively "thin" descriptions of many linguistic properties. The goal of the typicality-contextuality model is to provide cognitively and socioculturally driven model that can produce "thicker" descriptions of semantic properties (Geertz, 1973), and to achieve such a goal through training corpus in the size that is reasonable from the perspective of human language learning and development (i.e., millions of word tokens).

3.7. Chapter summary

This chapter III applies the typicality-contextuality model (Model I) to probe into the intriguing semantics of color, smell, and shape-based classifiers in ancient Chinese history over more than 1500 years of poetry corpus history. The exploration of competitions among multiple basic color domains in ancient Chinese via the typicality-contextuality mapping yield fascinating insights. Results show that unlike the basic-level color semantics in modern English and Chinese, which largely align along the typicality-idiosyncrasy axis, the competitions within basic-level color semantics in ancient Chinese primarily occur along the contextuality axis, which is orthogonal to the typicality-idiosyncrasy axis in the hyperbolic coordinates. Further investigations also lead to the identification of five basic terms for pleasant smells. Results reveal a stable overlap between these smell terms and several main color words throughout ancient

Chinese history, as well as suggesting a remarkable dependence of the language of smell on the color domain in a significant portion of ancient Chinese history. Finally, simulations on shape-based classifiers also shed light on the classical question in Chinese historical linguistics about their historical evolution. Results from the typicality-contextuality mapping cluster ancient Chinese classifiers into three groups. One group of classifiers, primarily stemming from words one-to-two dimensional shapes, is identified to be increasingly typical *and* distributed along the contextuality dimension. In contrast, classifiers that evolved predominantly from plant and utensil shapes are identified to be increasingly idiosyncratic *and* aligned along the typicality-idiosyncrasy dimension. Together, these findings not only enrich our understanding of ancient Chinese linguistics, but also demonstrate the effectiveness of the typicality-contextuality model and the domain-oriented semantic analysis.

Chapter IV: A singularity-synchronicity model of dynamic relationships between metacognition and semantic change (Model II)

4.1. Chapter Introduction

The goal of this chapter is to develop a cognition-driven dynamical systems model that is capable of causally categorizing patterns of diachronic changes of basic-level semantic categories. The next chapter (Chapter V) will apply this model to corpus data in English extrapolated from Google Ngram, a large corpus of digitized text (Michel et al., 2010). Together, these two chapters will provide meaningful cognitive insights into the dynamism and evolution of semantic categories of color, kinship, and smell over time, which are in principle also generalized to other semantic domains.

A fundamental hypothesis underpinning the model proposed in this chapter is that the metacognition (Yeung & Summerfield, 2012; Bruno et al., 2023; Fleming & Daw, 2017) of semantic change is both a response to and a driving force behind diachronic semantic transformations. Two pivotal measures are employed to measure and relate language change with metacognition: "singularity" (as "expressiveness") and "synchronicity" (as "conventionality"), which constitute the two basic variables of the dynamical-systems model. The model derives a set of dynamic modes relating language change to metacognitive sensitivity to language change (Section 4.5). In Chapter V, the empirical measures of singularity and synchronicity will be derived, based on which the metacognitive model can be empirically

applied to investigate the interplay between cognition and empirical language changes of multiple semantic domains observed from the Google Books corpus data.

Empirically, the metacognitive dynamical-systems approach proposed in this chapter intends to address an important puzzle in applying cognition-driven principles to modeling the dynamics of language change: under certain cognitive hypothesis, any natural language would move towards a single static equilibrium over a sufficiently long period of time, but in reality, no language seems to do so (e.g, Detges & Waltereit, 2008). In the literature, there is a lack of approach that can both endogenously and realistically incorporate multiple equilibria and multiple dynamic scenarios for modeling language change (Niyogi, 2006; Gell-Mann & Ruhlen, 2011; Nowak & Krakauer, 1999). An immediate empirical goal of the metacognitive model is to develop such a computational methodology. One building block of the two basic measures in the model is grounded in the domain and ensemble-driven language variation (Montemurro and Zanette, 2016). Domain and ensemble-driven temporal patterns, such as oscillation and synchronization, are ubiquitous phenomena among physical, biological, and ecological phenomena, which are, in particular, well modeled by the dynamical systems approach in multiple disciplines (Freeman 2008; Kay, 2014; Strogatz, 2000). The fact that diachronic changes in natural language also show oscillatory and coherence patterns further indicates the plausibility of using measures based upon ensemble and domain-oriented temporal patterns to relate the micro-level cognitive mechanism with the macro-level socio-historical phenomena such as language change.

4.2. Background

4.2.1. *The dynamical systems approach to language change*

In the field of mathematical and computational sciences, pioneering work on how micro-level cognitive and cultural mechanisms drive macro-level language change has used Bayesian statistics (Griffiths & Kalish, 2007), game theory (Clark & Kimbrough, 2017), and dynamical systems (Niyogi, 2006; Gell-Mann & Ruhlen, 2011; Nowak & Krakauer, 1999). Notably, in *The Computational Nature of Language and its Evolution* (2006), computer scientist Niyogi develops a comprehensive framework based on the dynamical systems approach to model how *a priori* cognitive mechanism in language acquisition and learning, applied to different parameter ranges, may lead to varied types of phonological and syntactic changes.

In the last decade, the statistical and distributional learning approach has emerged as a predominant framework for modeling human natural language in computational linguistics and NLP. Although primarily developed as a computational toolkit to discover and model linguistic patterns and regularities in corpus data, recent research has begun to explore its implications for addressing cognitive or socio-cultural questions on natural language. Psycholinguists use algorithms such as word embedding to investigate the cognitive implications of NLP algorithms, applying or testing them in cognitive and psycholinguistic experiments (Ettinger et al., 2016). Meanwhile, corpus-driven word-embedding algorithms are also used to quantitatively map large-scale socio-cultural patterns, such as changes in gender and racial stereotypes, as well as how social classes and professions interact in society and history (Garg et al., 2018; Kozlowski,

2019). However, it is unclear how to causally relate the statistical and distributional approach to psycholinguistic processing with macro-scale language variability in sociocultural spacetime.

A non-reductionist and integrative methodology that relates different levels of natural and social phenomena is common in many disciplines (Anderson, 1972; Strogatz 2000). However, for various conceptual, historical, or technical reasons, such a methodology is applied relatively less in today's cognitive science and many subfields of linguistics (Lizardo et al. 2020; Goldsmith, 2019). Following the general spirit of the works of Labov (2006), which suggest that language change intersects with both cognitive and sociocultural factors, this chapter presents a dynamical model focusing on how metacognition interacts with ensemble diachronic changes of basic-level semantic categories. Built on the premise that metacognition of semantic change both reacts to and steers the actual semantic change, the model captures the mutual dynamic interplay between metacognition and cognitive knowledge of semantic categorization.

4.2.2. Synchronicity and singularity in semantic categorization

One justification for the modeling approach developed in this chapter comes from Montemurro and Zanette (2016), who arguably provided the first empirical evidence of domain-oriented synchronization and oscillatory patterns in diachronic linguistic changes through computational corpus analysis. They observed that over the last 200 years of language corpus data across multiple Indo-European languages, rankings of single-word frequencies within the same semantic categories exhibit periodic patterns of synchronization to varying degrees (many of which follow a period of 14 to 20 years; see Figure 4.1). However, Montemurro and Zanette

(2016) do not specify the micro-level mechanisms that may support the diachronic linguistic oscillatory and coherent patterns observed in the corpus. In my work, I consider the dynamical relationship between temporal synchronicity and singularity as a fundamental micro-level psychological mechanism that may generate patterns of diachronic changes of language.

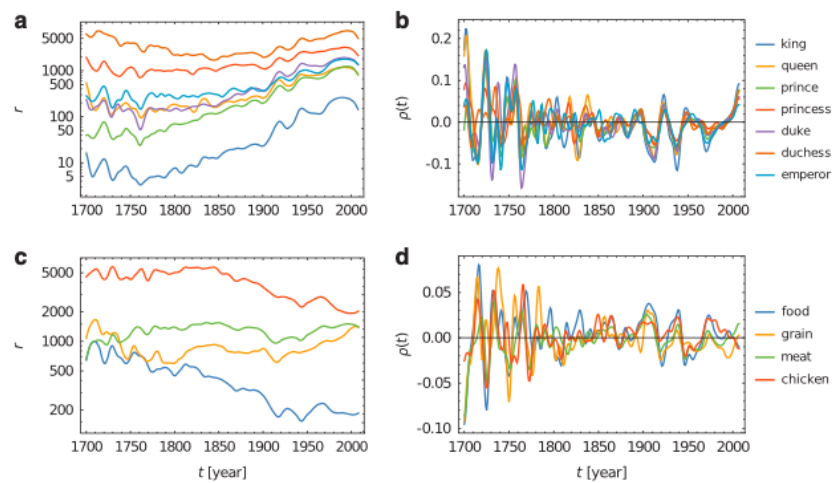


Figure 4.1. Montemurro and Zanette (2016) investigate the diachronic evolution of word frequency ranks in English from 1700 to 2008. (a, c) Variations of the logarithm of the rank over three centuries for two groups of related words, each belonging to the same semantic category. (b, d) Yearly differences in the logarithm of the rank of the same words, reflecting synchronized oscillatory variations in word prevalence. Corpus: Google Books; adapted from Montemurro and Zanette (2016).

Conceptually, formulating the measures of synchronicity and singularity in questions regarding semantic categories is also driven by the theoretical development in sociolinguistics and linguistic anthropology in the past 30 years. Theories in these two highly related fields point out that a major socio-psychological pressure in language communication originates from the

incongruity between what sociolinguists call the language community and the speech community. In Silverstein's (2014) formulation, the language community is "a primary reference group" of language ideology in society. The members of the language community are oriented to speak, using the Swiss linguist Saussure's classic term, *langue*, treating it as the ideal commutative norm. The speech community, by contrast, involves "the interpretability of verbal behavior" in relation to expectations of "appropriateness-to and effectiveness-in dynamic real-time contexts" in real communication with *parole* (speech), which always deviates to a certain degree from the ideal language norm of *langue* (Silverstein, 2014). In other words, the language community creates the norm as a desired conception of homogeneity and coherence in speaking and writing, which indexes the contextual variability of usages and projects and emanates the ideal language norm in the sociocultural spacetime. This creates a socially evaluative metric (the "standard") for members of a society. However, these members also belong to each of their speech communities, which often live not in the center of the ideal socio-political spacetime, and frequently face expectations and anxieties created through their cognizance of the ideal language norm. Hence, humans as sociocultural animals always need to orient themselves through simultaneously judging others' communicative behaviors and adjusting their own, according to each's interpretation of the ideal language norm (Silverstein, 2014).

Language norms, as described in the literature of linguistic anthropology and sociolinguistics, have two important characteristics that are crucial to this study. The first is that the language norm does not entail communicative consequences on any single linguistic unit (such as words or sentences), but often on collective aspects of language (such as genres and domains). Second, language norms create expectations and anxieties on an average speaker

within the language community: an individual can choose to speak according to the linguistic convention they cognize to facilitate conventionality that the language norm recognizes (such as saying a car is "red" similar to the way saying a plant is "green"), or to violate the linguistic norm to some degree to facilitate expressiveness that encourages idiosyncratic expressions (such as saying a car is "warmly red" in contrast to the way saying a plant is "green"¹⁵). Importantly, as work in linguistic anthropology emphasizes, evaluating such possibilities of conventionality and expressiveness is always reflective and metacognitive (Silverstein, 1976; 1993; 2022; Gal & Irvine 2020).

4.3. A metacognitive framework of domain-oriented “conventionality” (synchronicity) and “expressiveness” (singularity)

Labov (2016) arguably provides one of the most systematic frameworks for differentiating the psychological and sociocultural factors behind language change. However, recent research in social psychology, sociolinguistics, and linguistic anthropology rejects a sharp division between the psychological and sociocultural aspects of understanding language change. In this chapter, I consider metacognition as crucial for developing an effective socio-cognitive

¹⁵ The examples here indicate that for an average English speaker, the intuition that "saying a car is 'warmly red' " entails high expressiveness is not only about the relatively rare occurrence of the expression "warmly red" itself, but also related to the fact that "saying a flower is 'warmly green' " is also rare in contemporary ordinary English. In other words, evaluating expressions related to "red" according to language norms is not only based on the co-occurrences of "red" with its contextual words, but also based on the co-occurrences of "green" and other color words with their contextual words, i.e., how "red," "green," and other basic-level color words collectively "synchronize" with one another in cognition.

methodology for investigating the interplay between semantic change and cognition. Researchers in psychology consider metacognition as a deliberate and reflective awareness of the sociocultural environment. However, there is a lack of work that examines how theories of metacognition in cognitive psychology (Yeung & Summerfield, 2012; Bruno et al., 2023; Fleming & Daw, 2017) can be related to theories in sociolinguistics and linguistic anthropology, which treat metalinguistics as crucial for generating or mediating language variations and patterns over time and across socio-cultural spaces (Silverstein, 1993; Irvine & Gal, 2020). This paper is one of the first to propose a formal methodology to build such a connection.

Specifically, I propose four building assumptions for the domain-oriented metacognitive model:

- (a) (Non-metacognitive) Semantic knowledge is mediated by the first-order tradeoff between the language user's cognition of expressiveness (singularity) and conventionality (synchronicity) of semantic knowledge¹⁶.
- (b) (Metacognitive) The average language speaker's metacognitive awareness facilitates the following cognitive dynamics: that reflecting on expressiveness may influence conventionality, and reflecting on conventionality may reciprocally influence expressiveness.
- (c) (Metacognitive) The average language speaker's metacognitive awareness also

¹⁶ For instance, both extremely high and low expressiveness/singularity, as "nonsense" and "cliché", would be thought to convey low conventionality.

facilitates the following cognitive dynamics: that reflecting on expressiveness may influence expressiveness itself, and reflecting on conventionality may also influence conventionality itself.

- (d) (Metacognition of language variation and change) The average language speaker's metacognitive awareness is developed and driven by cognitively encoding synchronic and diachronic variability of domain and ensemble-driven linguistic patterns that they encounter in their cognitive and social environments.

Built upon these four assumptions, two additional derivative assumptions are followed:

- (e) (Reflective preference for (a)) The speaker also has a metacognitive awareness of the expressiveness-conventionality tradeoff described in (a), a self-feedback cognitive process of the first-order cognitive tradeoff.
- (f) (Reflective aversion to (a)) The speaker also has a metacognitive awareness that the metacognitive processes described in (b) and (c) may violate the synchronicity-singularity tradeoff described in (a), which reflectively make the first-order tradeoff unpreferred.

The following section will formalize these six assumptions into six formal components of the dynamical-systems model. In the model, assumptions (a), (b), and (c) will be incorporated as linear terms, and assumptions (e) and (f) will be formalized as a quadratic term. Assumption (d) will be integrated in the form of first-order time derivatives of singularity and asynchronicity, which necessitates solutions through a system of two differential equations.

4.4. The dynamical-systems model of metacognition

- 4.4.1. The model

Incorporating the four basic assumptions and two derived assumptions above as formal components, this section develops a formal metacognitive model using the dynamical-systems approach. In the next chapter (Chapter V), specific corpus-based empirical measures for synchronicity and singularity will be developed, and the model will be applied to multiple domain-specific empirical cases.

Below is the formal description of the model. For an average speaker A of generation k who learns to understand and speak the natural language L , let V be singularity of using words in a specific semantic domain l of L from generation $k-1$, and let P be asynchronicity¹⁷ of using words in the same semantic category of L from generation $k-1$. The speaker A 's cognitive process for learning and using a word or expression w_l from domain l is driven by all the cognitive and metacognitive mechanisms characterized by assumptions from (a) to (e), which can be grouped by the three following directions of optimization. On the one hand, for the sake of expressiveness, speaker A tends to utilize high singularity and asynchronicity for using w_l from

¹⁷ Through this chapter and the rest of the dissertation, asynchronicity is used in an exchangeable way with synchronicity, and it is measured by one minus synchronicity. In most cases I use the word "synchronicity" only for the sake of simplicity. The two basic variables in the model, however, are singularity and asynchronicity, in which cases "synchronicity" and "asynchronicity" are not exchangeable and will explain the rationale behind in the rest of this section.

semantic domain l . On the other hand, speaker A is also aware of the need and pressure of social communication, which tends to synchronize the speaker's semantic knowledge with the socio-cultural linguistic norm by decreasing singularity and asynchronicity. Finally, speaker A also cognizes the tradeoff between conventionality and expressiveness at both cognitive and metacognitive levels (Table 4.1).

For the average speaker A in generation k , who learns and uses a word or expression w_l based on their diachronic language input that form their semantic knowledge about w_l and the semantic domain l that w_l belongs to¹⁸, I propose a dynamical-systems model to capture all three sets of cognitive mechanism characterized above, formed by the following set of two differential equations of singularity as V and unconventionality as P (Equations 4.1 and 4.2):

$$\dot{V} = V(a - V)(V - 1) - bP + \epsilon_V \quad (4.1)$$

$$\dot{P} = cV - dP + \epsilon_P \quad (4.2)$$

In these equations, the two variables $P, V \in R$ ¹⁹, the parameters $a, b, c, d \in R^+$ and $\epsilon_V, \epsilon_P \in R$.

Each of parameters a, b, c, d respectively measures the metacognitive mechanism between each

¹⁸ Canonically, I assume that speaker A encounters the diachronic semantic input from generation $k-l$ throughout their developmental period, but human language learning is not just influenced by direct interactions with parents and individuals from the parent generation, as much dynamical modeling work on language change assumes (e.g., Niyogi, 2008). It also includes socialization, schooling, education, from which a child encounters language inputs from other generations as well as their own generation, and learns to recognize and differentiate synchronic and diachronic characteristics of language inputs from different linguistic domains and categories (i.e., genres). These observations provide an additional socio-cognitive justification for using corpora like the Google Books corpus across decades and hundreds of years to apply this model empirically, as shown in Chapter V and part of Chapter VI.

¹⁹ Popularity and singularity in this dissertation are not defined in a purely information-oriented way. Therefore, I do not measure them in the range of $[0, 1]$. In psychological experiments, researchers use

corresponding term with the changes of diachronic synchronicity singularity cognized by speaker A's from the language input in their cognitive and social environment, which are characterized by first-order derivatives of each V and P ; and ϵ_V , ϵ_P represent cognitive factors that drive V and P in the non-metacognitive setting (i.e, when all metacognitive mechanisms characterized by the two equations are absent). This two-dimensional dynamical system²⁰ characterizes how the metacognitive feedback and self feedback on expressiveness and conventionality, the first-order cognitive tradeoff between expressiveness and conventionality, and the metacognitive feedback on the first-order tradeoff together constrain the expected changes of expressiveness and conventionality over time.

	"Pro-expressiveness" attitude	"Pro-conventionality" attitude	"Ambivalent" attitude
\dot{V} Singularity (expressiveness) changing over time (Assumption d)	V self-reinforcing preference of singularity (Assumption c)	$-bP$ negative feedback of increasing asynchronicity on singularity (Assumption b)	$(a - V)(V - I)$ second-order reflective tradeoff: impulsive and instrumental attitudes towards singularity (Assumptions e and f)

rating methods among others to operationalize cognitive variables that are not operationalized by probability values between 0 and 1. Hence, P and V can be any real number values in the model.

²⁰ Formally, this dynamical-systems model resembles the Fitzhugh-Nagumo model in computational neuroscience for modeling single-neuron behaviors, which is used to study varied resting, excitatory, and oscillatory patterns in biological neural activities (FitzHugh, 1961; Nagumo et al., 1962). See section 5.5.4 for further discussion on this connection.

\dot{P}	cV	$-dP$	
<p><u>A</u>synchronicity (<u>un</u>conventional ity) changing over time</p> <p>(Assumption d)</p>	<p>positive feedback of increasing singularity on asynchronicity</p> <p>(Assumptions a and b)</p>	<p>self-reinforcing aversion to asynchronicity</p> <p>(Assumption c)</p>	

Table 4.1. Each cell represents one specific type of metacognitive feedback to either singularity or synchronicity in the speaker's metacognitive awareness of generating a word or expression in communication. In this table, the metacognitive components are organized according to their motivational valence: "pro-novelty," "pro-conventionality," and "ambivalent". Cognitive assumptions from which each formal component is derived is identified in the parentheses.

Table 4.1 summarizes the cognitive characteristics and motivations for each formal component of the model. For the process of speaker A 's acquisition and usage of semantic knowledge of words in semantic domain l , I assume that the cognitive drive towards social conventionality means that increasing asynchronicity P will provide negative feedback on singularity V (term $-bP$ in Equation 4.1). I also assume that speaker A possesses a meta-awareness of social conventionality, in which the speaker's reflective knowledge of such need and pressure of social communication tends to drive them in two directions. The first direction is towards a metacognitively, self-reinforcing aversion to asynchronicity (term $-dP$ in Equation 4.2), and the second direction is towards a medium level of singularity (term $(a - V)(V - l)$ in Equation 4.1); because both extremely low ("cliche") or high singularity ("nonsense") in linguistic expression can be metacognitively thought to be either preferable or not preferable for

communication. Cognitively, this quadratic term is also further justified by literature in cognitive psychology demonstrates that generating a "more singular" and novel expression would require a higher cognitive load and a more deliberate level of attention from the speaker, which entails deeper reflective processing of the sociocultural pressure of conventionality (Xie & Salvendy, 2000; Ginns & Leppink, 2019). Further, I also assume that speaker A has an opposite tendency to generate novel semantic expressions from semantic domain l to promote expressiveness. This tendency is formulated by a first-order cognitive tradeoff mechanism between synchronicity and singularity (term cV in Equation 4.2), and by a reflective metacognitive mechanism that can be characterized as self-reinforcing aversion to asynchronicity.

- 4.4.2. *Empirical motivations for the dynamics of the model*

Much of the empirical motivations for the above formulations originates from the general behavioral science literature on cognitive processing and cognitive decision-making, as I consider making linguistic choices regarding semantic knowledge also belongs to questions of perceptual processing and cognitive decision-making. The term of singularity aversion ($-bP$ in Equation 4.1) is motivated by the literature on uncertainty aversion (e.g., Kahneman & Tversky, 1979; Ghirardato & Marinacci, 2001; Kelsey & Le Roux, 2015). The term of asynchronicity aversion ($-dP$ in Equation 4.2) uses the measure of synchronicity of a single-word or multiword expression (which will be formally defined in Section 5.2 in next chapter). The reflective mechanism that mediates medium-level semantic singularity, as characterized by term ($a -$

$V)(V - I)$ in Equation (4.1), is empirically motivated by the literature on "Go/NoGo" decision-making. A number of studies in decision-making science demonstrates that the Go/NoGo decision-making is more likely to be impulsive than other decision-making scenarios (such as two-alternative choice; e.g., Frederick et al. 2011). Additionally, research also shows that making a Go/NoGo decision can be staged: with the more impulsive, pro-novelty process coming first, followed by the more instrumental, pro-conventionality like tendency (Berditchevskaia et al. 2016). Here, the specific behavior-related assumption behind the quadratic term is that the impulsive attitude tends to lead to linguistic decisions with high semantic singularity, and the instrumental attitude tends to lead to decisions with low singularity, and in the non-metacognitive scenario, they may balance with one another.

4.5. Five metacognitive dynamical modes derived from the model

Equations (4.1) and (4.2) form a two-dimensional dynamical system, which can be described by a dynamical systems map of the two variables V and P . Each nullcline represents when the derivative of one variable is equal to zero (i.e., when equation (4.1) or (4.2) is equal to zero), one variable is a function of the other. The dynamics between variables V and P are captured by how they are attracted or repelled by one or multiple intersections (nodes) of the V and P nullclines in the phase plane. Dynamically, there are three basic types of nodes: a stable node, an unstable node, and a spiral. A stable node (also called an "attractor") forms an equilibrium point towards which both V and P dynamically move; an unstable node forms a

point from which both V and P move away; a spiral (also called a "focus") is a point where the variables either move toward or move away from in a spiral or oscillatory fashion (a stable spiral or an unstable spiral) (Figure 4.2).

Different outcomes of the dynamical system can be determined by the Jacobian matrix of the linearized system formed by Equations (4.1) and (4.2):

$$\begin{pmatrix} -3V^2 + 2(a+1)V - a & -b \\ c & -d \end{pmatrix} \quad (4.3)$$

Stability of equilibria is then determined by the eigenvalues of the Jacobian matrix. Let $A = -3V^2 + 2(a+1)V - a$. Equilibrium stability can be found by the characteristic equation:

$$\det \begin{pmatrix} A - \lambda & -b \\ c & -d - \lambda \end{pmatrix} = 0 \quad (4.4)$$

which has two eigenvalues as solutions:

$$\lambda_1 = \frac{\tau + \sqrt{\tau^2 - 4\Delta}}{2} \quad \text{and} \quad \lambda_2 = \frac{\tau - \sqrt{\tau^2 - 4\Delta}}{2},$$

where $\tau = A - d$ and $\Delta = -Ad + bc$. (4.5)

A potential equilibrium point is determined by the intersection(s) of the V and P nullclines. If $\tau < 0$, the solutions are either real negative eigenvalues (when $\tau^2 - 4\Delta > 0$) or complex eigenvalues that have a negative real part (when $\tau^2 - 4\Delta < 0$). In the first case, the system

directly converges to a stable node as the equilibrium; in the second case, the system forms a stable spiral to which the system is attracted in an oscillatory manner. On the other hand, if $\tau > 0$, there will be an unstable spiral or node in the system (Figure 4.2).

I now analyze how this dynamical system model generates different types of metacognitive trajectories for the speaker's decision on articulation of semantic knowledge. For analytic clarity, I assume that parameters a, b are fixed and let $\epsilon_V = 0$, and varying c, d, ϵ_P alone will generate the whole spectrum of dynamics, as fixing c, d and varying a, b will also generate the same typology of dynamics as characterized in the analysis.

Analytically, three main types of cognitive dynamics are characterized by the model. In the first type of dynamics, the speaker is more attentive to the first-order tradeoff between synchronicity and singularity than the self-reinforcing aversion to asynchronicity (i.e., $\frac{c}{d}$ is sufficiently large), and the speaker's default perception of asynchronicity change is either sufficiently high or low (ϵ_P is high or low). When the ratio between c and d is sufficiently large, dynamically there is only one potential equilibrium (namely, one intersection of the V and P nullclines). Depending on whether ϵ_P is sufficiently high or low, there are two sub-scenarios of dynamics, which I respectively term "student," the speaker who is willing to create new linguistic norms, and "professor," the speaker who tends to maintain old linguistic norms:

Proposition 1 ("student")²¹: If $\frac{c}{d}$ is sufficiently large and the speaker's default perception of asynchronicity change in time is sufficiently high (i.e., ϵ_P

²¹ Chapter V will provide empirical examples for applying these formal propositions and illustrate how these formal propositions can inform cognitive and language modeling in actual cases. In particular, section 5.1 provides the transition from theoretical dynamical analysis characterized in this section to

is high), then there is one stable node representing the equilibrium of low singularity. The speaker is willing to create new linguistic norms, if (1.1): proper initial conditions lead them to high singularity and asynchronicity first, and (1.2): the asynchronicity of the final equilibrium is low enough.

Proposition 2 ("professor"): If $\frac{c}{d}$ is sufficiently large and the speaker's default perception of asynchronicity change in time is sufficiently low (i.e., ϵ_P is low), then there is one stable node representing the equilibrium of high singularity, so the speaker is likely to maintain the old linguistic norms.

In the second type of dynamics, the ratio between c and d is still sufficiently large, but the speaker's default perception of asynchronicity change is moderate, a stable or unstable spiral will emerge, the scenario of which I call "politician," the speaker who is not committed to either make new linguistic norms or stay with old norms:

Proposition 3 ("politician"): If $\frac{c}{d}$ is sufficiently large, and the speaker's default perception of asynchronicity change is moderate (i.e., ϵ_P is moderate), then there is a spiral of moderate risk and gain uncertainties. In this dynamics, the speaker is mostly a "sophist" who is not easily committed or uncommitted to create new linguistic norms, unless (3.1): proper initial conditions lead them to high gain and

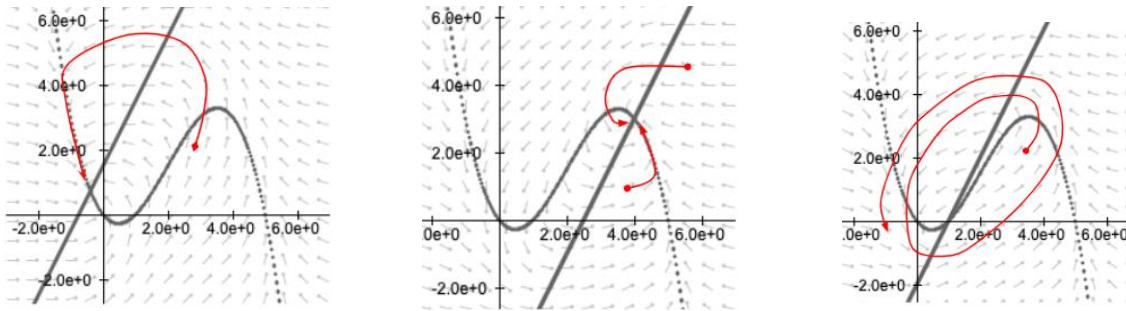
more empirically grounded dynamical analysis that directly informs studies in three empirical sections on basic-level color, kinship, and smell semantics in Chapter V (Sections 5.3 - 5.5).

risk uncertainties at first, (3.2): the final equilibrium is a stable spiral, and (3.3): the speaker has high tolerance for prolonged fluctuations of uncertainties.

In the third type of dynamics, the speaker is much less attentive to asynchronicity than to the first-order risk-and-gain tradeoff (i.e., $\frac{c}{d}$ is sufficiently small), then unless ϵ_P is extremely high or low, there are two stable nodes in the dynamics. Nonetheless, different initial conditions will directly lead the speaker to either equilibrium only, the scenario I call "citizen," the speaker who weighs the norms of different novel expressions differently):

Proposition 4 ("citizen"): If $\frac{c}{d}$ is sufficiently small, then there are two stable nodes (unless ϵ_P is extremely high or low²²), one of which is an equilibrium of low singularity, and the other is an equilibrium of high singularity. Different initial conditions would always lead the speaker to monotonically go to either equilibrium, so the speaker is unlikely to be willing to create new linguistic norms.

²² In either case, the dynamics will follow either Proposition 1 or 2.



Pst. 1 ("student")

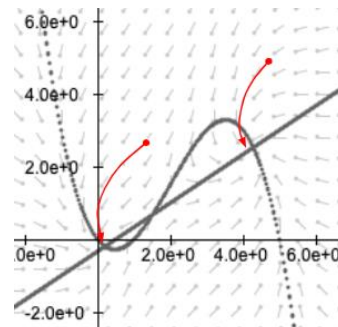
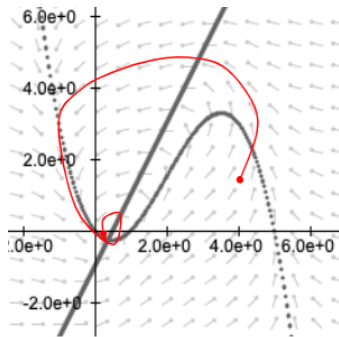
Pst. 2 ("professor")

Pst. 3 ("politician")

$c = 4, d = 2, \epsilon_P = 3$: stable node spiral

$c = 4, d = 2, \epsilon_P = -10$: stable node

$c = 4, d = 2, \epsilon_P = -4$: unstable



Pst. 3 ("student" if under conditions (3.1) to (3.3))

Pst. 4 ("citizen")

$c = 4, d = 2, \epsilon_P = -2$: stable spiral

$c = 4, d = 6, \epsilon_P = -2$: two stable nodes

Figure 4.2. The dynamical-systems maps of five metacognitive scenarios are simulated and depicted to illustrate Propositions (Pst.) 1 through 4 (for this specific visualization, $a = 5, b = 4,$ and $\epsilon_V = 0$). The x axis maps singularity and the y axis maps asynchronicity, both perceived by the speaker. As stated by Proposition 5, only two sub-scenarios in propositions 1 and 4 yield bittersweet trajectories that would most likely lead a speaker to articulate a novel and impactful expression. In the first scenario, a stable node with low asynchronicity and low singularity emerges (top row, left); in the second scenario, a stable spiral emerges (bottom row, left). Notice that asynchronicity and singularity are measured numerically here such as in rating tasks in experimental psychology, in which "zero" means that the speaker is psychologically neutral to synchronicity or singularity of a word or expression.

According to the above four propositions, psychological dynamic trajectories that follow the main hypothesis only emerge in the "student" and "politician" scenarios with proper initial conditions. I further determine two "new norm maker" scenarios in the following proposition, in which the speaker is most likely to articulate a novel expression:

Proposition 5: ("new norm marker"): Given $\frac{c}{d}$ is sufficiently large, if the initial singularity ranges from moderate to high and the initial asynchronicity ranges from low to moderate, then the speaker is likely to create a new linguistic norm: first, a fastly converging trajectory towards the eventual post-reform equilibrium of lower singularity and moderate asynchronicity, as characterized by proposition 1 under conditions (1.1) and (1.2); second, a slowly converging, oscillatory trajectory towards the eventual equilibrium of moderate singularity and lower asynchronicity, as characterized by proposition 3 under conditions (3.1) to (3.3).

In the first case in Proposition 5 (equivalent to Proposition 1 under conditions (1.1) and (1.2); e.g., Figure 4.2, top left), the equilibrium forms a stable node, which is more stable and resilient to perturbations than a stable spiral is, so the speaker expects a one-shot trajectory, in which both high asynchronicity and singularity are cognized, but they are then perceived to decrease and be stabilized to an equilibrium with both low asynchronicity and singularity. In the second case in proposition 5 (equivalent to Proposition 1 under conditions (3.1), (3.2), and (3.3); e.g., Figure 4.2, bottom left), the equilibrium forms a stable spiral, in which both high

asynchronicity and singularity are cognized, but then they are gradually stabilized to the equilibrium, the process of which is damped and oscillatory. The second dynamics is expected to take more time, experience more fluctuations, and be in smaller parameter ranges than in the first case.

4.6. Chapter discussion and summary

The core objective of this chapter has been to formulate a (meta)cognition-driven dynamical systems model, the purpose of which is to dynamically and causally categorize patterns of diachronic changes in basic-level semantic categories and how they interact with cognition. A fundamental hypothesis underpinning this model is that the metacognition of semantic change is both a response to and a driving force behind actual semantic transformations. To measure and establish a relationship between language change and metacognition, singularity and asynchronicity as two basic measures of semantic expressiveness and conventionality are employed, forming two basic variables of the model.

Formally, the model is developed via a dynamical systems approach, the broad goal of which is to tackle an essential conundrum in modeling language variation and change in current literature: most modeling work on language change simulates language change within contexts of single-equilibrium or single dynamic scenarios, but empirically, over an extensive timeframe

no language moves towards equilibrium. This model aims to bridge this gap by establishing a computational methodology that is capable of endogenously incorporating multiple equilibria and dynamic scenarios in order to model empirically observed language change.

Incorporating metacognition as the empirical foundation of the model is influenced by the contemporary literature of sociolinguistics and linguistic anthropology, in which metacognition is a core concept for understanding language variation and change. But in current literature from the metacognitive perspective, there is little work developing quantitative measures or computational models for metacognition or metacognition-driven linguistic phenomena. This model also addresses such a notable shortfall in the existing literature. Though for specific empirical purpose of this thesis, the measures of singularity and asynchronicity are assumed as the two basic cognitive dimensions of the metacognitive model, in principle, the canonical model itself can be extended to incorporate metacognitive mechanisms based on any other alternative cognitive dimensions and can be applied to questions other than those on semantic categories²³.

Building upon the theoretical and formal groundwork laid out in this chapter, Chapter V will apply this model to historical corpus data in American English. In combination, these two chapters strive to deliver both theoretical and empirical insights into the dynamic relationship between cognition and evolution of semantic categories over time. In particular, Chapter V will

²³ In a working paper, I use the same canonical model of metacognition to investigate decision-making scenarios in state-lead political and economic reform, with uncertainty in risk assessment and uncertainty in expected utility as two basic cognitive dimensions. Though these two works address entirely different empirical questions and are based on distinct groups of empirical literature, formally they are both developed from the same canonical dynamical-systems model of metacognition. See Ji (working paper a).

focus on the semantic domains of color, kinship, and smell, and one section in Chapter VI will further extend the framework to other semantic domains.

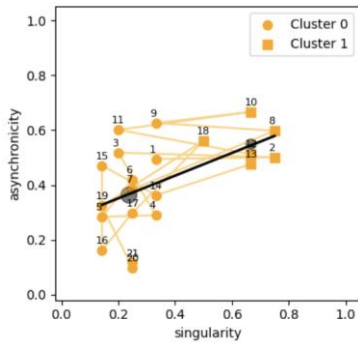
Chapter V. Applying Model II to investigate how metacognition is relevant to semantic changes in color, kinship, and smell domains in written American English

5.1. Five empirical scenarios of meta-level and non-meta-level cognitive dynamics

Using the procedures developed in the method section (Section 5.2) in this chapter, I will transform diachronic Google Books data of year-based word frequency percentage changes into proper numerical formulas (Equations 5.1 and 5.2). Based on such a transformation, I will then demonstrate that simple linear regression and k -means clustering procedures can identify empirical scenarios from diachronic corpus data according to the metacognitive dynamical model developed in the last chapter, and differentiate metacognitive scenarios from non-metacognitive scenarios. In the empirical sections of this chapter (and in Section 6.5 of Chapter VI), I will demonstrate the effectiveness of employing this computational procedure to apply the metacognitive model developed in Chapter V to actual diachronic corpus data.

Categorically, the above computational procedure generates an empirical typology consisting of five types of cognitive dynamics. Among the five types, three are metacognitive scenarios (scenarios 5.1a, 5.1d, and 5.1e), one is weakly metacognitive (scenario 5.1b), and one is non-metacognitive (scenario 5.1c). The three metacognitive scenarios can be captured by at least one type of metacognitive dynamics that are theoretically driven in Section 4.5 in the last chapter (Propositions 1 to 4). The scenario of weak metacognition means that the language speaker is cognitively neutral to at least one cognitive variable (asynchronicity or singularity, or

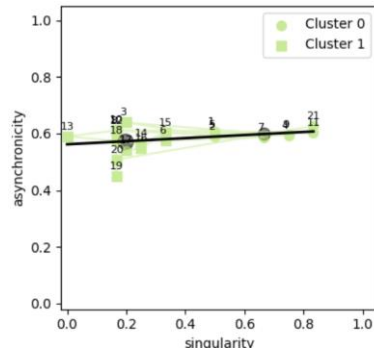
both). The non-metacognitive scenario in the simulation always entails the first-order cognitive tradeoff between synchronicity and singularity. This scenario means that though the speaker is not "cognitively neutral" to how synchronicity and singularity affects their cognitive-decision on articulation a word or a language expression w , asynchronicity and singularity are cognitively evaluated as in an "either-or" relationship, with little self or mutual metacognitive feedback that is described in the metacognitive model (Equations 4.1 and 4.2 in Section 4.4).



(a): "citizen"

metacognition with 2 attractors

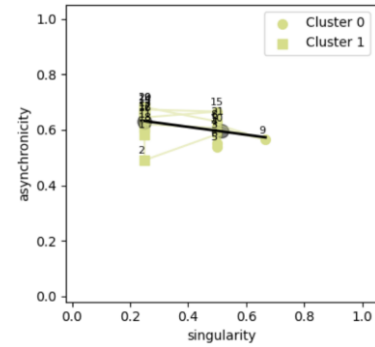
(+ *coeff.*, *strong alignment*)



(b): "child"

weak metacognition

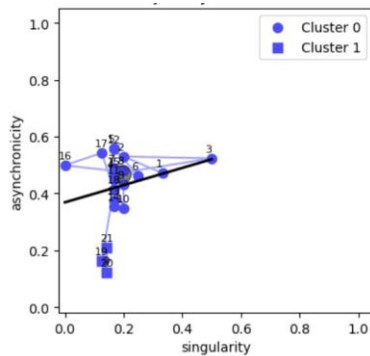
(~ 0 *coeff.*, *strong alignment*)



(c): "scientist"

non-metacognitive tradeoff

(- *coeff.*, *strong alignment*)

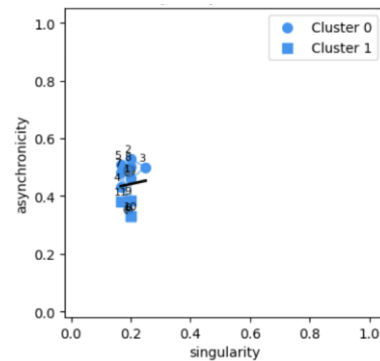


(d): "politician," "student," or "shifting around"

metacognitive with semi-stable attractors

or, shifting btw multiple metacog. modes

(+ *coeff.*, *weak alignment*)



(e): "professor"

metacognition with one stable attractor

(2 clustering centers collapsing into 1, *weak alignment*)

Figure 5.1. Dynamical scenarios 5.1a ("citizen"), 5.1b ("child"), 5.1c ("scientist"), 5.1d ("politician," "student" or "shifting around"), and 5.1e ("professor") identified from corpus data by the singularity and asynchronicity, the empirical measures of which will be developed in section 5.2. These data-driven scenarios are compared with the theoretical scenarios derived from the dynamical-systems-based analysis of the metacognitive model, which are developed in section 4.4 and 4.5. Scenarios 5.1a, d, and e are the corresponding empirical cases of the five metacognitive modes theoretically characterized in Section 4.5. Scenario 5.1b indicates weak metacognition and scenario 5.1c non-metacognitive, first-order tradeoff between singularity and asynchronicity. In each representation, the x axis = singularity and the y axis = asynchronicity. The number on each dot corresponds to time windows.

The empirical typology are generated with linear regression and k-means clustering procedures performed on the epoch-based diachronic data using Python, in which the epoch is determined to be 20 years and its justification is discussed in Section 5.2, where empirical measures of synchronicity and singularity based on diachronic corpus data are derived. The goal of linear regression is to discover the fitted line that approximate the linear nullcline predicted in the metacognitive dynamical-system models (the P nullcline derived from Equation (4.2) when the equation is equal to zero, which is visually shown as the linear nullcline in all subplots in Figure 4.2). When the regression coefficient is positive, the corresponding empirical scenario is identified to be metacognitive (as the 2-D dynamical model requires the P nullcline to have a positive slope). The goal of the k-means clustering procedure²⁴ is to identify cluster centers that approximate the two attractors that are characterized by the dynamical mode described in Proposition 4 in Section 4.5 of Chapter IV (which is visually shown in the subplot at the bottom

²⁴ And the dynamical-systems analysis developed in Section 4.5 predicts no more than two attractors, there is no need to perform the k-means clustering procedure here for $k > 2$. Notice that this k-means procedure is different from the time-series k-means procedure which is performed on ensemble patterns of diachronic corpus data, which is introduced in Section 5.2.

right of Figure 4.2), which corresponds to the empirical scenario 5.1a. If the two cluster centers generated from data are too close, the empirical scenario may approximate other metacognitive modes with one equilibrium, which corresponds to the scenario 5.1e. A strong alignment between a fitted line coefficient and two independent cluster centers indicates that the empirical scenarios approximate a stable metacognitive dynamical mode that is theoretically derived from the dynamical model in Section 4.5 (when the coefficient is positive: scenario 5.1a) or a stable non-metacognitive mode (when the coefficient is near zero or negative: scenario 5.1b and 5.1c). A weak alignment would indicate an unstable shift between multiple metacognitive scenarios (scenario 5.1d), or a metacognitive scenario if two cluster centers appear to degenerate into one (scenario 5.1e).

Specific descriptions of the empirical typology of metacognitive and cognitive dynamics are as follows:

Scenario 5.1a ("citizen": metacognition with two equilibria; Figure 5.1a): When linear regression generates a fitted curve with positive slope and the two cluster centers approximately align along the fitted curve, this approximates the dynamic scenario of Proposition 4 ("citizen"; Figure 4.2). The fitted curve approximates the linear nullcline (P nullcline), and the centers of the two cluster terms approximate the two attractors theoretically described in the dynamic mode of Proposition 4 ("citizen"), where one equilibrium is of lower singularity and lower asynchronicity, and the other equilibrium is of higher singularity and higher asynchronicity.

Scenario 5.1b ("child": weak metacognition; Figure 5.1b). When linear regression generates a fitted curve with slope close to zero and the two cluster centers approximately align along the fitted curve, this indicates that metacognitive feedback on asynchronicity may *not* be significant ($\frac{c}{d}$ in Equation 4.2 is close to zero), because the asynchronicity level does not change much diachronically.

Scenario 5.1c ("scientist": non-metacognitive tradeoff; Figure 5.1c). When linear regression generates a fitted line with negative slope and the two cluster centers approximately align along the fitted line, this indicates that asynchronicity and singularity are cognitively evaluated as in a first-order trade-off relationship, with little self or mutual metacognitive feedback.

Scenario 5.1d ("student," or "politician", or "shifting around": semi-stable or unstable metacognition/shifting between multiple metacognitive modes; Figure 5.1d). When linear regression generates a fitted curve with positive slope, but the two cluster centers distinctly do not align along the fitted curve, this indicates that either at least one major shift between dynamic scenarios characterized in the metacognitive model (Figure 4.2) may occur diachronically, and more than one mode of metacognitive dynamics is required to capture it. This scenario is cognitively involved with dynamical modes characterized by Proposition 1("student"), Proposition 3 ("politician"), or multiple propositions of metacognitive dynamics characterized in Section 4.5.

Scenario 5.1e ("professor": metacognition with one equilibrium; Figure 5.1e). When the data is better clustered around one center than two centers, usually they possess small variance on the singularity dimension and small to moderate variance on the asynchronicity dimension. It may indicate three possibilities: the empirical dynamics can be captured by one theoretical scenario of a single attractor (Proposition 2), the metacognitive feedback on singularity in particular may not be significant, or the empirical dynamics cannot be best captured by the metacognitive model²⁵.

Out of these five empirically derived scenarios, scenarios 5.1a, 5.1d, and 5.1e are the corresponding empirical cases of the five metacognitive scenarios theoretically derived in Section 4.5 in Chapter VI. In contrast, scenario 5.1b indicates weak metacognition ("child"), and scenario 5.1c ("scientist") indicates non-metacognitive, first-order tradeoff between singularity and asynchronicity.

The empirical scope of applications in this chapter concentrates on semantic domains previously studied in the first model introduced in Chapter II: color, kinship, and smell. The aim is to probe relatively short-term semantic changes in modern English of the last 200 years, rather than long-term semantic changes such as what have been examined for ancient Chinese in Chapters III. The rationale behind this bifurcation, choosing Chinese for the study of long-term changes and English for short-term changes, is primarily practical. Chinese boasts a long and

²⁵ The last case can be generated by the fact that a linguistic expression may be highly semantically stable within a substantial period of time.

continuous history of recorded written language change, making it an ideal candidate for studying long-term shifts. However, its data possesses relatively lower temporal resolution. Conversely, while the modern English corpus records a shorter history of continuous change, its data offers a relatively higher temporal resolution, making it well-suited for investigating short-term fluctuations.

5.2. Deriving the corpus-based empirical measures of singularity and synchronicity

For the actual implementation of the model, I derive corpus-based empirical measures of diachronic (a)synchronicity and singularity in the following manner. Firstly, for each target semantic domain under study, such as "color" or "kinship," I obtain candidate words of the basic-level categories of this semantic domain via the similarity mapping as described in Chapter II. Then, I obtain the single-word frequency percentage of each candidate word in the whole corpus of a fixed historical window. The procedure is repeated using fixed temporal steps that cover the entire diachronic range included in the corpus. The corpus used for most empirical studies in Chapters V and VI will be the Google Books corpus of American English from 1800 to 2020. One study in Section 5.3 in Chapter V will also employ the Google Books corpus of British English from 1800 to 2020.

Once n words are selected as the candidate basic-level words for a specific semantic category, the raw data of the single-word frequency percentage of each word in the whole target corpus in every year from 1800 to 2020 is downloaded from the Google Ngram Viewer (Michel

et al., 2010) and coded by years. For each group of diachronic data, the diachronic window size is set to be m years, where m is an even number, and the temporal step is set to be $m/2$ years. Then a Euclidean time series k -means clustering procedure is performed. Then, the pairwise between-cluster and within-cluster Euclidean distances between each pair of words are calculated.

The word that has the highest single-word frequency percentage w_{Ref} is identified as the reference word for the clustering within each period ΔT . For word w that is not w_{Ref} , the diachronic synchronicity measure $p(w)$ within ΔT is defined in the following manner:

$$p^{\Delta T}(w) = \begin{cases} 1/\sqrt{1 + d^{\Delta T}(w, w_{Ref})} & \text{if } w \in C_{Ref} \\ 1/\sqrt{1 + d^{\Delta T}(C_w, C_{Ref})} & \text{if } w \notin C_{Ref} \end{cases} \quad (5.1)$$

where C_{Ref} is the cluster of which $w_{Ref} \in C_{Ref}$, C_w is the cluster of which $w \in C_w$ if $w \notin C_{Ref}$, and d is the Euclidean distance between clusters.

The diachronic singularity measure $v(w)$ within ΔT for word w is defined in the following manner:

$$v^{\Delta T}(w) = \begin{cases} (n_m^{\Delta T} - n_l^{\Delta T}) / (1 + n_m^{\Delta T} - n_l^{\Delta T}) & \text{if } w \in C_l \text{ and } n_m^{\Delta T} \neq n_l^{\Delta T} \\ 1 / (1 + n_m^{\Delta T} - n_l^{\Delta T}) & \text{if } w \in C_m \text{ and } n_m^{\Delta T} \neq n_l^{\Delta T} \\ 1/2 & \text{if } n_m^{\Delta T} = n_l^{\Delta T} \end{cases} \quad (5.2)$$

where $n_m^{\Delta T}$ is the number of words in the cluster that contains the largest number of words among all clusters, identified as C_m for period ΔT . $n_l^{\Delta T}$ is the number of words in the cluster that contains the least number of words among all clusters, identified as C_l for period ΔT .

The synchronicity measure $p(w)$ and singularity measure $v(w)$ within ΔT for word w will form the two basic measures for implementing the dynamical model with empirical corpus data. In terms of parameter settings, k is set to equal 3 for all time-series k -means clustering procedures²⁶. In practice, simulations generally perform with a high quality when k is equal to 2, 3, and 4. But in the procedures described in Equations (5.1) and (5.2), perhaps because both within-cluster differences and between-cluster differences are incorporated into computation, empirical testing shows little discrepancy when k is chosen among any of the three numbers. For the diachronic window size m and temporal step $m/2$, in all current implementations, m is set to be 20 years and the temporal step is therefore 10 years. One of the first set computational studies investigating linguistic ensemble regularity in diachronic language change, as mentioned in Section 4.2.2 of Chapter IV, discovers synchronic regularity with a period of 14 to 20 years (Montemurro and Zanette, 2016). In some cases, ensemble regularity in language change can also appear visually recognizable from Google Ngram Viewer, a visualization tool of language change based on the Google Books corpus since the year 1800 (Figure 5.2).

The study mentioned above does not specify the mechanisms behind regularity emerging from language change at the ensemble level. It is noticeable, nonetheless, that 14 to 20 years are approximately equivalent to the time that an average human speaker develops full semantic and pragmatic knowledge for communication (Hart & Risley, 1995; Schneidman et al., 2013). As pragmatic and social knowledge that may affect cognitive knowledge of semantic categories is

²⁶ Note that this time-series k -means procedure is different from the non-time-series k -means procedure that is performed for approximating the attractors of the dynamical systems predicted by Model II, which is discussed in Section 5.1.

within the scope of the modeling in this chapter, the diachronic time window is chosen around the upper bound of human language development that also includes language socialization. Hence, in the studies of this chapter, it is assumed that modeling based on periods of around 20 years or its multiplications may be optimal for capturing generation-based regularity from diachronic corpus data.

All implementations and visualizations are conducted in Python.

5.3. Implementation of Model II on basic-level color semantics in American and British English from 1800 to 2020

Applying the empirical procedures and measures developed in Section 5.2, the metacognitive model (Model II) is first implemented examining basic-level color words ("black," "white," "red," "yellow," "green," "blue," "orange," "brown," "purple," "pink," and "gray") in both American and British English from 1800 to 2020. Based on a comparative modeling study on color semantics across these two English languages, I will show that the metacognitive model is capable of differentiating metacognitively-mediated semantic and pragmatic changes, which are often related to sociocultural context and dynamics, from non-metacognitive ones. For basic-level color words in American English, "black," "white," and "brown," which are closely related to the sociocultural history of race (Emirbayer & Desmond, 2015), are shown to metacognitively differ from other words in the 20th century but not in the 19th century. In comparison, for basic-level color words in British English, the metacognitive model differentiates "black," "white,"

"red," and "yellow" from other color words, and shows cognitive dynamics different from the group of "black," "white," and "brown" in American English. The collective semantic pattern emerging in British English color words in the 20th century is better correlated with domain-specific typicality than sociocultural context, whereas for the case of American English color words it is the other way around.

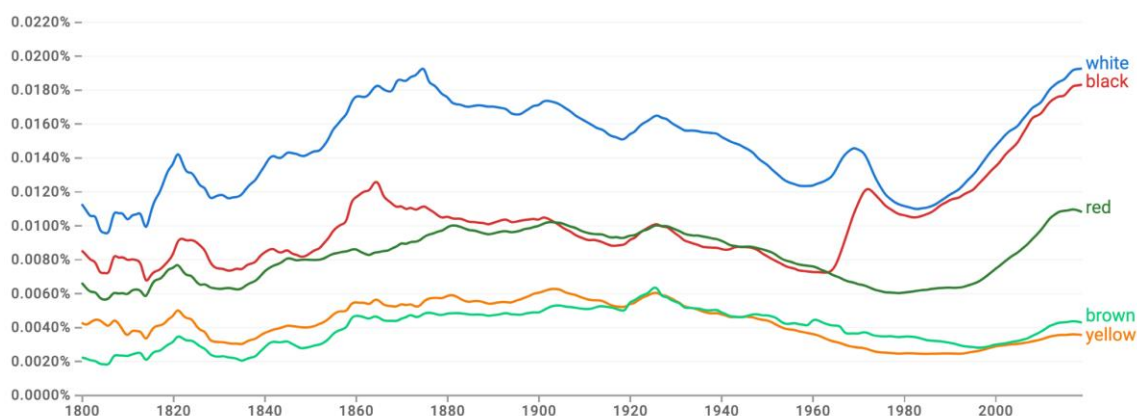


Figure 5.2. Diachronic evolution of the single-word frequencies of "white," "black," "red," "yellow," and "brown" from 1800 to 2020. Data are extracted from the whole Google Books corpus in American English (Smoothing = 3). The y axis = the percentage of single-word frequency in the whole Google Books corpus in American English, and the x axis = year.²⁷ The empirical analysis developed in this chapter assumes that the language speaker's metacognitive awareness is sensitive to the diachronic pattern of the domain-specific ensemble activity of semantic knowledge, which I measure by how words are diachronically synchronized within their collective semantic domain. Cognitive consequences of semantic synchronization in the diachronic dimension, however, cannot be simply captured by any measure of first-order synchronization alone, as shown in the changes of word frequency levels in this figure. This chapter provides a set of empirical examples demonstrating that metacognitive mechanisms reciprocally interact with domain-specific semantic synchronization in the diachronic dimension, which is important for modeling the relationship between semantics and cognition.

²⁷ <https://books.google.com/ngrams/>

Figure 5.3 displays a metacognitive dynamic mapping of color semantics in American English from 1820 to 2020, divided into two periods of one hundred years (1820 to 1920, and 1920 to 2020). Each dot represents data from a 20-year window size with a 10-year step, extracted from the Google Books corpus and measured by the singularity and asynchronicity scores developed in Section 5.2. In this test, "black" is set to be the reference word for measuring synchronicity degrees of all other color words, and linear regression and k-means clustering procedures (in which $k=2$) are performed on the data (recall Section 5.1). Data from the period from 1820 to 1920 are plotted in light colors and data from the period from 1920 to 2020 are plotted in corresponding normal colors. Table 5.1 shows the results that are geometrically represented in Figure 5.3, including the regression coefficient, size of each cluster centers, level of alignment between the regression line and cluster centers for each color word. Numeral values for positions of cluster centers and alignments between cluster centers and regression lines are not shown in the table but are represented in Figure 5.3.

The right two columns in Table 5.1 summarize the predicted dynamical scenarios for basic-level color semantics across the two periods. Results indicate that in reference to the term "black," almost all basic-level color categories in 19th-century American English can be captured by a single scenario of weakly or non-metacognitive (scenario 5.1b), as characterized by the near-zero regression coefficients (within the range of ± 0.1) and strong alignments between regression lines and cluster centers. But in the 20th and 21st-century American English, "red," "yellow," "green," "purple," and "pink," in reference to the term "black," all turned from the

weakly or non-metacognitive scenario to the unstable or shifting metacognitive scenario (scenario 5.1d); "orange" and "gray" turned from the weakly or non-metacognitive scenario to a stable metacognitive scenario (scenario 5.1a); and "blue" shifted from a stable metacognitive scenario (scenario 5.1a or e) in the period of 1820 to 1920 to the unstable metacognitive scenario which characterizes the majority of color words from 1920 to 2020²⁸. Noticeably, only "white" and "brown" stayed in the weakly or non-metacognitive scenario (scenarios 5.1b and 5.1c) for both periods, in reference to "black".

²⁸ While the observation that "blue" is the only term that is in a metacognitive scenario in the 19th century is interesting and needs further explanation, this does not affect the central argument here, which is "white" and "brown" are two terms that have been co-evolving with "black," more than any other basic-level color terms.

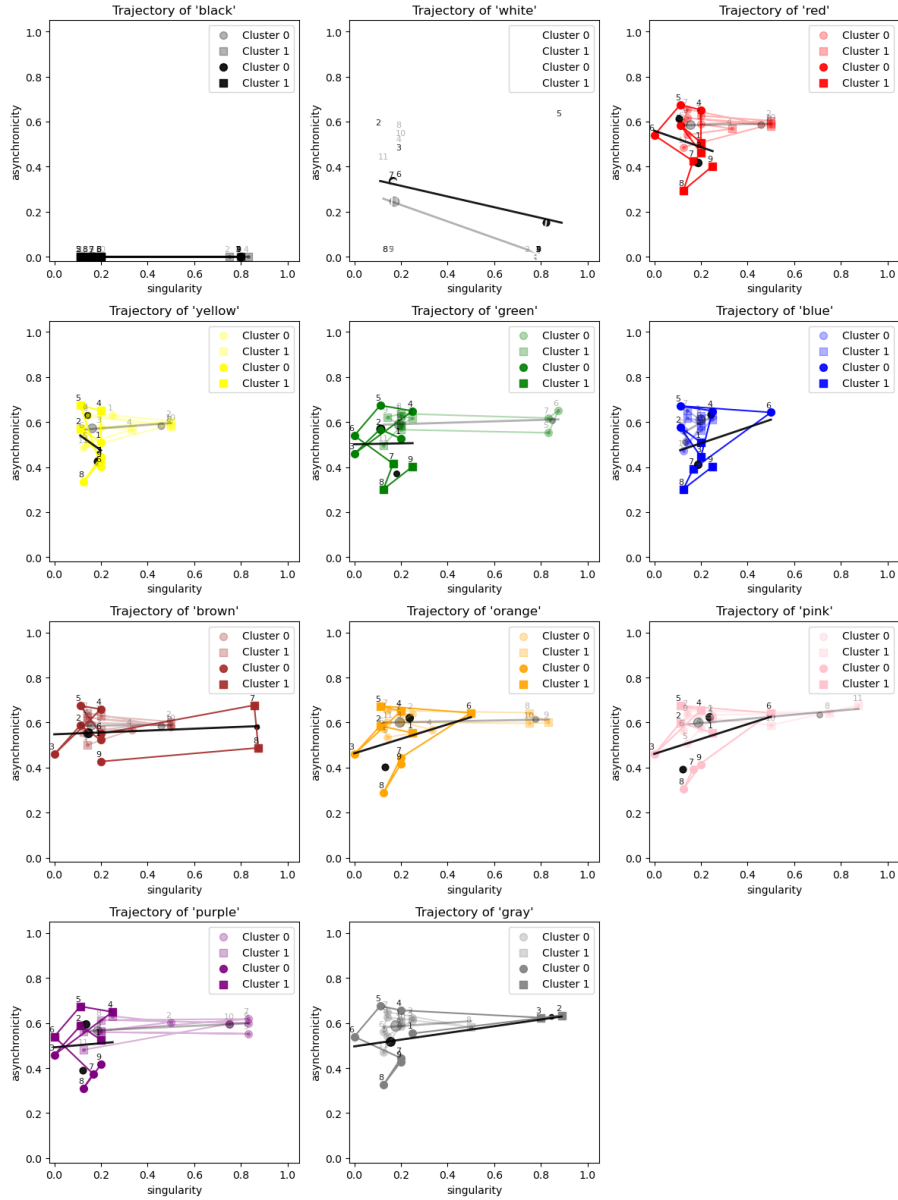


Figure 5.3. Diachronic trajectories of basic-level color words in **American English** mapped in the singularity-asynchronicity space from 1820 to 1920 (marked by corresponding light colors) and from 1920 to 2020 (marked by corresponding normal colors), with a 20-year window size, a 10-year step, and "black" as the reference term for synchronicity (hence its asynchronicity score = 0). Almost all basic-level color categories in 19th-century American English can be captured by a single scenario of weakly or non-metacognitive (scenario 5.1b: "child"), as characterized by near-zero regression coefficients (within the range of ± 0.1). But in the period from 1920 to 2020, "brown" and "white" are the only two terms that are characterized as weakly or non-metacognitive (with scenarios 5.1b and c, respectively) in reference to "black". They are

both differentiated from all other non-racially sensitive color terms in American English in the 20th century, which are dominantly shifting between multiple metacognitive scenarios (5.1d), or stably metacognitive (scenario 5.1 or 5.1e). See Section 5.1 for details for the classificatory methodology. Corresponding numeral and dynamical results are listed Table 5.1 (**Figure 5.3**).

Together, these results suggest two observations. First, from a metacognitive point of view, "brown," "black," and "white" have been synchronically co-evolving since the early 19th century, more than with other basic-level color words especially after the 20th century. Second, these results also have a direct cognitive implications: intuitively, this means that for an average American English speaker in the 20th and 21st-century (who was exposed to the language environment similar to the American English written corpus from the same period), when the term "black" is used, terms "brown" and "white" would appear less "reflective" in cognition than other basic-level color terms. In other way around, when "red," "yellow," "green," "purple," "pink," "blue," "orange," or "gray" is used in communication, the speaker would similarly reflect how much any of this term is different from "black," "brown," and "white". On the other hand, for the period of 1820 to 1920, almost all basic-level color words in this period show weak metacognition (near-zero coefficients in scenario 5.1b), and the word "white" shows a non-metacognitive tradeoff between asynchronicity and singularity (negative coefficient in scenario 5.1c), suggesting that socioculturally-driven metacognitive feedback on basic-level color semantics in American English during that period may not be strong overall²⁹.

²⁹ This does not mean that socio-cultural factors concerning race in American English were not important in the 19th and early 20th century (Wu, 2003; Emirbayer & Desmond, 2015). What the analysis indicates is that individuals during that period may not be strongly metacognitively aware of these factors when color words are used in communication (i.e., cognitive *reflection* on the socio-cultural connotations of color words is weak). Non-metacognitive tradeoff between asynchronicity and singularity can still play a cognitive role, as shown here in the dynamical scenario of "white" (when "black" is the reference word).

American English Ref. word: "black"	Regression Coeff.		Size of the majority cluster		Size of the minority cluster		Alignment between fitted lines and cluster centers		Predicted dynamic scenarios	
	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020
black	0	0	6	6	3	3	reference	reference	reference	reference
white	-0.363	<u>-0.159</u>	7	6	2	3	strong	strong	5.1c	<u>5.1c</u>
red	0.021	-0.388	6	5	3	4	strong	weak	5.1b	5.1d
yellow	0.078	-0.825	6	6	3	3	strong	weak	5.1b	5.1d
green	0.031	-0.004	7	5	2	4	strong	weak	5.1b	5.1d
blue	0.584	0.349	7	5	2	4	moderate	weak	5.1a or e	5.1d
orange	0.019	0.316	7	7	2	2	strong	moderate to strong	5.1b	5.1a
pink	0.093	0.324	7	5	2	4	strong	weak	5.1b	5.1d
purple	0.048	0.067	5	5	4	4	strong	weak	5.1b	5.1d or 5.1e
brown	0.007	<u>0.036</u>	7	7	2	2	strong	strong	5.1b	<u>5.1b</u>

The result of non-meta-level tradeoff for "white" (scenario 5.1c) with reference to "black" indicates that the cognitive contrast between these two words is captured by the model from the corpus in this period, no matter whether the contrast is primarily due to the opposite perceptual or socio-symbolic differences of these two words, or both. But these differences are poorly reflected in the semantic cognition.

gray	0.073	0.150	8	7	1	2	strong	strong	5.1b	5.1a
Mean (μ)	0.059	-0.014					overall: strong	overall: weak	overall: 5.1b	overall: 5.1d
SD (σ)	0.034	0.142								

Table 5.1. Dynamical analysis of basic-level color terms in **American English** for the period from 1820 to 1920 and from 1920 to 2020, with the reference word for synchronicity as "**black**." A bold or bold underlined term shows dynamics that differentiates itself from the semantic ensemble dynamics in the entire color domain. Dynamical trajectories plots based on these results are shown in Figure 5.3 and interpretations of these results are shown in the captioning of Figure 5.3. See Section 5.1 for details for the classificatory methodology.

As in 20th-century sociocultural history of America, "black," "brown," and "white" all developed strong connotations about race (Emirbayer & Desmond, 2015), one may further infer that what the metacognitive modeling captures from the domain-specific diachronic semantic synchronization patterns entails these basic-level color words' connotations related to American sociocultural history of race. Noticeably, no first-order measures of diachronic synchronization, such as frequency-based synchronicity or singularity scores alone, can capture all of these connotations³⁰.

Now let us turn to a comparative analysis on the basic-level color semantics in British English. Based on the Google Books corpus in British English, I perform a similar analysis on basic-level color words from 1820 to 1920 and from 1920 to 2020. The analysis indicates,

³⁰ Though diachronic synchronization between "black" and "white" after the 1960s can be even visually identified from changes of their frequencies in during this period (Figure 5.2), which co-occurred with the rise of civil movements during that time, it is difficult to examine the development of equally important socio-cultural connotation of "brown" by these first-order measures.

however, the coefficients and alignment patterns of basic-level color semantics in British English are better correlated with semantic typicality (Figure 5.4; Table 5.2). Most color words in both periods are characterized by the weakly metacognitive or non-metacognitive scenarios (scenarios 5.1b or 5.1c), which indicate that in reference to "black," no strong metacognitive awareness is involved. On the other hand, the modeling identifies possible shifting between metacognitive and non-metacognitive scenarios for "white," "red," "yellow," which are among the most typical basic-level color words identified in multiple empirical studies as well as my typicality-contextuality mapping in Chapter 2 (Section 2.5).

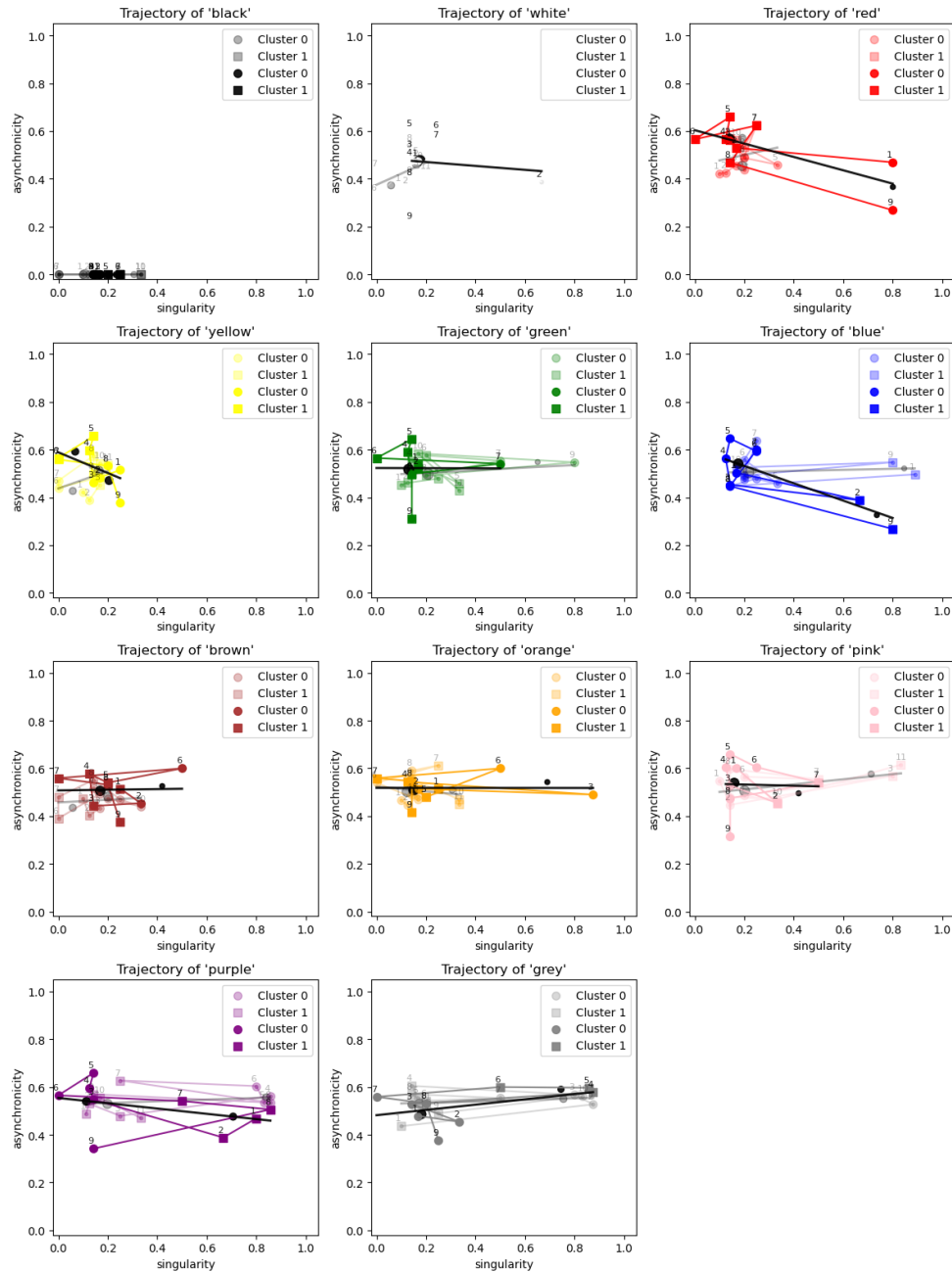


Figure 5.4. Diachronic trajectories of basic-level color words in **British English** mapped in the singularity-asynchronicity space from 1820 and 2020 and from 1920 to 2020, with a 20-year window size, a 10-year step, and "**black**" as the reference term for synchronicity. "White," "red," and "yellow" show a moderate level of dynamic heterogeneity across the two periods. All other terms show both near-zero coefficients and moderate to strong alignments between regression lines and cluster centers, suggesting they are stably non-metacognitive. In contrast to American

English since the 20th century, these patterns suggest that at the metacognitive level, basic-level color semantics in British English since the 19th century are predominantly domain-dependent rather than culture-dependent (**Figure 5.4**).

This observation certainly does not suggest that color semantics in British English is not influenced by sociocultural factors. Historians have long demonstrated that Britain has an equally rich socio-cultural history of race as America does (Said, 1979; Hawksley, 2013; Flauder, 2015). Beyond specific color words' symbolic meaning in the socio-historical contexts of race and colonization, "purple" has also been a unique and important color term in both religious and political histories of Britain³¹; synchronicity and singularity scores for "purple" since the 1800s (using the method proposed in Section 5.1) also show the exceptional status of "purple" compared to other color terms in British English (data not shown here). The empirical purpose of the metacognitive model, however, is to identify what socio-historical factors matter metacognitively and how to distinguish them from less metacognitively significant ones. Modeling work in this section shows that sociocultural contexts in 19th century America and in Britain over the last 200 years may not be metacognitively relevant to color semantics in the two English languages, whereas the sociohistorical factors relating color to race can be relevant to color semantics for the average American English speaker in the 20th and 21st century³².

³¹ For instance, in the Victorian era in the 19th century, purple was linked with royalty and nobility due to the rarity and expense of purple dye (Flanders, 2015). In the early 20th century, purple was adopted by the Women's Social and Political Union as a symbol of dignity during their campaigns for women's suffrage (Hawksley, 2013).

³² One may delve further into the puzzle of "brown" in color naming studies in cognitive science, which were initiated by cognitive scientists based in America. For example, Berlin and Kay's classical study on color (1969) considers "brown" to occupy an independent "evolutionary" stage immediately following "green" and "blue". Interestingly, their empirical studies were conducted in the 1960s and 1970s, during

British English Ref. word: "black"	Regression Coeff.		Size of the majority cluster		Size of the minority cluster		Alignment between fitted lines and cluster centers		Predicted dynamic scenarios	
	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020	1820 - 1920	1920 - 2020
black	0	0	6	6	3	3	reference	reference	reference	reference
white	<u>0.466</u>	<u>-0.081</u>	6	8	3	1	moderate to weak	<u>weak</u>	<u>5.1a or e</u>	<u>5.1d</u>
red	<u>0.230</u>	<u>-0.404</u>	6	8	3	1	<u>weak</u>	<u>strong</u>	<u>5.1e</u>	<u>5.1c</u>
yellow	<u>0.363</u>	<u>-0.508</u>	6	5	3	4	moderate to weak	moderate	<u>5.1e</u>	<u>5.1c or e</u>
green	0.054	-0.002	8	8	1	1	strong	strong	5.1b	5.1b
blue	0.022	-0.375	8	6	1	3	strong	strong	5.1b	5.1c

which period when the immigration of Hispanic communities to the United States witnessed a surge. Additionally, many of their studies were conducted in the San Francisco Bay area in California, where the Hispanic communities were growing fast (Berlin and Kay, 1969; Miller & Quigley, 1990; García, 2015). This increase of Hispanic immigration to California was argued to be driven by a variety of factors, including changes in social and economic conditions in Latin American countries, and several key pieces of policy and legislation enacted during this period affected immigration patterns, such as the end of the Bracero Program in 1964 which strongly discriminated workers from Mexico, and the passing of the Hart-Celler Act in 1965, which abolished the national-origins quota system that had previously given preference to immigrants from Western Europe (Ortiz & Telles, 2012; García, 2015; Ngai, 2014). Further inquiry is needed to investigate how much the above sociocultural dynamics was correlated with the emergence of "brown" as a salient color term in American English in the second half of the 20th century, and perhaps in turn conditioned the term "brown" to become metacognitively sensitive to the sociocultural context.

orange	-0.100	-0.001	7	5	2	4	strong	moderate	5.1b	5.1b
pink	0.098	0.005	6	7	3	2	strong	strong	5.1b	5.1b
purple	0.035	-0.113	6	6	3	3	strong	strong	5.1b	5.1b
brown	0.057	0.013	6	7	3	2	strong	moderate	5.1b	5.1b
grey	0.030	0.104	8	7	1	2	strong	strong	5.1b	5.1b
Mean (μ)	0.099	-0.136					overall: strong	overall: strong	overall: 5.1b	overall: 5.1b
SD (σ)	0.019	0.062								

Table 5.2. Dynamical analysis of basic-level color terms in **British English** for the period from 1820 to 1920 and from 1920 to 2020, with the reference word for synchronicity as "black". An underlined term shows dynamics that differentiates itself from the semantic ensemble dynamics in the entire color domain, which appears domain-specific rather than culture-dependent in British English. Dynamical trajectories plots based on these results are shown in Figure 5.4 and interpretations of these results are shown in the captioning of Figure 5.4.

Broadly, the analysis in this section proposes a sophisticated picture of what metacognitive dynamical analysis can achieve for investigating domain-specific semantic categories, by connecting cognition with both linguistically endogenous semantic properties and exogenous sociocultural contexts that shape semantics and cognition. The study indicates that the influence of synchronic and diachronic sociocultural context on color naming semantics can be encoded in higher-level cognition (i.e., at the metacognitive level) and form metacognitive feedback to semantic cognition. Further, the modeling also shows how dynamical shifts between metacognitive and non-meta-level cognition are related to cognition and basic-level color

categories³³. Together, this inquiry indicates the possibility that metacognitive dynamics and the dynamical shifts between metacognitive and non-metacognitive modes of semantic cognition reciprocally interact with the diachronic change of color semantics in culture and society. One both drives and is driven by the other.

5.4. Implementation of Model II on basic-level kinship semantics in American English from 1800 to 2020

I apply here the metacognitive model to kinship semantics in American English. The similar procedures of linear regression and k -means clustering are performed. Intriguingly, dynamic patterns of the basic-level kinship semantics emerging from the analysis are different from those of the basic-level color semantics. For the 200-year period from 1800 to 2020 and when "father" is set to be the reference word for synchronicity, results are as follows:

³³ For basic-level color semantics, the empirical studies in Chapter III show that in ancient Chinese, semantic competitions occurred and persisted for several main color domains ("red," "green/blue," "black"). Modeling work in Chapter III shows that competitions of color semantics in ancient Chinese are aligned in a different cognitive dimension (the contextuality dimension) from color semantics in modern English (which predominantly align along the typicality-idiosyncrasy dimension). But in principle, metacognitive mechanisms can also be applied to develop a dynamic account for semantic competition in the color domains in ancient Chinese, which were arguably much more pervasive than basic-level color semantics in English in the last two hundred years (Section 3.4, Chapter III). The main practical difficulty is data sparsity and temporal uncertainty of much of the corpus data before the early modern era.

- (i). The terms "father," "mother," "brother," and "sister" vary in a relatively small range of singularity. However, clustering centers do not linearly align along the simulated linear nullclines (Table 5.3)³⁴. This is best characterized by the empirical scenario 5.1d ("politician," or "student": Figure 5.1d), which indicates shifts between more than one metacognitive mode predicted by the dynamical-systems model in Section 4.5. It suggests that all four words may have experienced a major shift (or shifts) between different metacognitive modes over the past 200 years.
- (ii). The terms "grandfather," "grandmother," "husband," and "wife" can be captured by the metacognitive scenario 5.1a ("citizen"), similar to many non-race related basic-level color words in American English: for all four terms, two cluster centers in each case align along linear nullclines determined by linear regressions, and linear nullclines for four words are all positive (Table 5.3). The "citizen" mode in the metacognitive model alone can characterize this empirical scenario (Figure 5.1a and Figure 4.2 Pst.4).
- (iii). "Son" and "daughter" are shown to be the group in between: for both terms, two cluster centers in each case align along linear nullclines determined by the regression coefficients, but linear nullclines for both terms are near zero (Table 5.3). Hence, they can be best captured by the empirical scenario of weak metacognition (5.1b: "child").

³⁴ Here, the result on the alignment of clustering centers and the regression lines for the term "father" is generated using "mother" as the reference word, the simulation of which is not shown here.

American English Ref. Word: "father" Period: 1800 - 2020	Regression Coeffs.	Size of the majority cluster	Size of the minority cluster	Alignment between fitted lines and cluster centers	Predicted dynamic scenarios
father	0	20	1	reference	reference
mother	0.303	18	3	weak	5.1d
brother	0.651	15	6	weak	5.1d
sister	0.331	18	3	weak	5.1d
son	- 0.094	17	4	strong	5.1b
daughter	0.001	16	5	strong	5.1b
husband	0.123	11	10	strong	5.1a
wife	0.133	13	8	strong	5.1a
grandmother	0.084	16	5	strong	5.1a
grandfather	0.283	18	3	strong	5.1a
Mean (μ)	0.235				
SD (σ)	0.192				

Table 5.3. Dynamical scenarios of basic-level kinship terms in American English for the period from 1820 to 2020. Reference word = "father". The kinship terms are classified into three dynamical scenarios. 5.1a ("citizen") = metacognition with two equilibria; 5.1b ("child") = weak

metacognition on synchronicity. 5.1d ("politician" or "student") = shifting between multiple metacognitive modes.

From these observations over the entire periods of the 19th and 20th centuries, I further hypothesize that these observed kinship patterns are related to the change of the social history of the American family over the past 200 years, in particular marked by shifts in parent-child and sibling relationships in response to societal and economic changes. As work in sociology and anthropology of family show, the United States in the 1800s was dominated by agricultural families where children were economic assets and sibling relationships were practical and supportive (Coontz, 1993; Ruggles, 1994). The Industrial Revolution transitioned more middle-class families to urban areas, with increased sibling interactions due to shared experiences. From the early to mid-20th century, the nuclear family emerged with defined parental roles, and a renewed focus on shaping children's futures (Ruggles, 1994; Volo & Volo, 2007; Chambers & Gracia, 2022). Sibling bonds continued to strengthen during this period through shared activities and experiences. In the second half of the 20th century, family structures diversified due to increasing divorce rates, single-parent families, and the legalization of same-sex marriages, further altering parent-child and sibling dynamics. Hence, since the early-mid 20th century, parent-child relationships and sibling dynamics have significantly evolved (Volo & Volo, 2007; Chambers & Gracia, 2022).

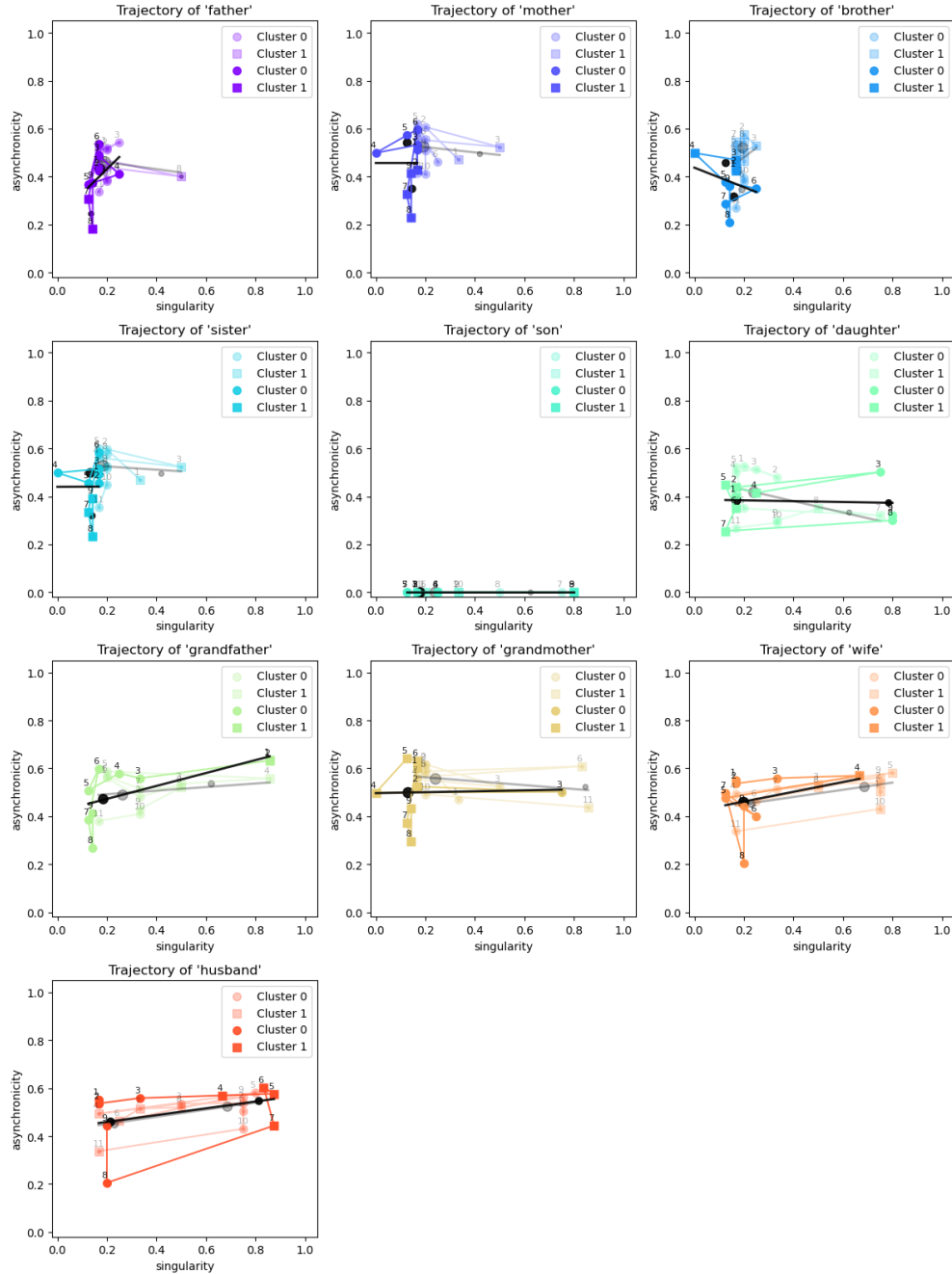


Figure 5.5. Diachronic trajectories of basic-level kinship words in American English mapped in the singularity- asynchronicity space from 1820 to 1920 and from 1920 to 2020, with a 20-year window size and a 10-year step. "Son" is chosen to be the reference word for both periods. Normal colors are chosen to represent data for the period from 1920 to 2020, and Corresponding light colors represent data for the period from 1820 to 1920. Procedures to generate these empirical dynamic scenarios are the same with the case of basic-level color semantics.

Corresponding numeral and dynamical results are listed in the "1920-2020" and "1820 -1920" columns in Table 5.4. I further propose that these patterns can be accounted for by how each group of kinship terms are correlated with their connotations with the socio-cultural evolution of parent-child relationships and sibling dynamics represented in the American English corpus (**Figure 5.5**).

To computationally examine this possibility, I conduct further analysis dividing the diachronic data into two periods of one hundred years. Figure 5.5 and Table 5.4 present results from simulations on the basic-level kinship terms in American English in the periods from 1820 to 1920 and from 1920 to 2020. In particular, an alternative term "son" is chosen as the reference word for enhancing possibilities of capturing diachronic variations in both parent-child and sibling dynamics. Results can be summarized as follows:

- (i). From the period of 1820 - 1920 to the period of 1920 -2020, the dynamics for "father," "mother," "sister" changed from the scenario of weak metacognition (scenario 5.1c: "scientist") to the scenario of one or multiple metacognitive modes (scenario 5.1e "professor" or 5.1d "politician," or "student"), in reference to "son" (Figure 5.5; underlined and bold items in Table 5.4). "Brother" was in the mode of shifting metacognitive dynamics since the 1820 -1920 period (scenario 5.1d), but regression coefficients greatly changed across these two periods (0.780 for the period from 1820 to 1920, and - 0.449 for the period from 1920 to 2020, as shown in Table 5.4), suggesting that the semantic shifting of "brother" in reference to "son" across these two periods were not consistent. Together, these results suggest that in reference to the term "son," all four words either experience unstable metacognitive modes, or shift from the weakly metacognitive mode (5.1c) to one or

multiple shifting metacognitive modes (5.1e or 5.1d) over the two historical periods.

- (ii). In particular, "father" and "brother" are found to enter into unstable shifts in metacognitive dynamics in the period from 1920 to 2020 (scenarios 5.1d: "politician," or "student"). For the term "father," such an instability is highlighted by both large changes of regression coefficients (from 0.780 to - 0.449) and shifts in the alignment between fitted lines and cluster centers (from strong to weak alignments) (Table 5.4). This indicates that the paternal parent-child relationship and fraternal sibling relationship changed tremendously from the 19th to 20th century in American family history, as argued in the family studies literature (Coontz 1993; Ruggles 1994; Volo & Volo 2007; Chambers & Gracia 2022)³⁵.
- (iii). The terms "husband," "wife," and "grandfather" do not experience dynamical shifts across these two periods in reference to "son". "Husband" and "wife" can be captured by scenario 5.1a ("citizen") for both periods, and "grandfather" can be captured by scenario 5.1b ("child") for both periods. This suggests that semantic relations between these three terms and "son" are stable. "Grandmother" shows a mild shift from the weak metacognition scenario (scenario 5.1b) to the metacognition mode with two equilibria (scenario 5.1a). The discrepancy between "grandmother" and "grandfather" may capture the rise of grandmother's role in the

³⁵ Additional simulations with "daughter" as the reference word do not generate the equally large shifts for "mother" and "sister," further highlighting that the basic-level family semantic shifts are not gender symmetrical.

American family since the depression in the 1930s, as demonstrated in the sociology and human development literature (Dunifon & Bajracharya, 2012; Meyer, 2014).

American English	Regression Coeff.		Size of the majority cluster		Size of the minority cluster		Alignment between fitted lines and cluster centers		Predicted dynamic scenarios	
	1820 -1920	1920 - 2020	1820 -1920	1920 -2020	1820 -1920	1920 - 2020	1820 -1920	1920 - 2020	1820 -1920	1920 - 2020
Ref. word: "son"										
father	<u>-0.143</u>	<u>0.994</u>	8	6	1	3	strong	weak	<u>5.1c</u>	<u>5.1d</u>
mother	-0.108	-0.095	8	5	1	4	strong	weak	5.1c	5.1e
brother	<u>0.780</u>	<u>-0.449</u>	7	5	2	4	weak	weak	5.1d	<u>5.1d</u>
sister	-0.067	-0.086	8	5	1	4	strong	weak	5.1c	5.1e
son	0	0	8	7	1	2	reference	reference	reference	reference
daughter	-0.238	-0.013	8	7	1	2	strong	strong	5.1c	5.1b
husband	0.149	0.188	6	5	3	4	strong	strong	5.1a	5.1a
wife	0.149	0.244	6	7	3	2	strong	strong	5.1a	5.1a
grandmother	<u>-0.078</u>	<u>0.289</u>	7	7	2	2	strong	strong	<u>5.1b</u>	<u>5.1a</u>
grandfather	0.074	0.029	6	7	3	2	strong	strong	5.1b	5.1b

Mean (μ)	0.063	0.134		overall strong			
SD (σ)	0.1015	0.209					

Table 5.4. Dynamical analysis of basic-level kinship terms in American English for the period from 1820 to 1920 and from 1920 to 2020. Reference word = "son". A bold or bold underlined term shows the dynamics that differentiates itself from the semantic ensemble dynamics in the entire kinship domain. In reference to "son", "father" and "brother" are exceptional for both being in the shifting metacognitive scenario (scenario 5.1d: "politician" or "student") during the period of 1920 to 2020, and "mother" and "sister" are exceptional for both being in transiting from a non-metacognitive scenario (scenario 5.1c: "scientist") to a metacognitive scenario (5.1e: "professor") across the 19th and 20th centuries. Dynamical trajectories plots for both periods are shown in Figure 5.5. See Section 5.1 for details for the classificatory methodology.

Overall, results from the two periods indicate two types of kinship semantic dynamics in reference to "son". Terms related to parent-child and sibling relationships, such as "father," "mother," "brother," and "sister," entered into a more unstable mode of metacognitive dynamics after 1920 in reference to "son". Other terms, by contrast, experienced relatively little or no dynamical changes in relation to "son". Together, this study suggests a new computational methodology studying shifts in complex social relationships like family relationships, while connecting semantic meaning and social cognition with social changes.

Noticeably, relating these results and observations on English kinship semantics generated by Model II with the results on kinship terms produced by Model I in Chapter II (described in section 2.3) can also shed light on the discrepancy in semantic typicality judgment on basic-level kinship terms in different cognitive experiments (as summarized in Table 2.3 in Chapter II). The typicality-contextuality mapping generated by Model I shows two groups of semantic relations, one of which can be captured as {"brother": "sister" :: "son" : "daughter"} and the other as {"father" : "mother" :: "grandfather" : "grandmother"} (Figure 2.2). On the other

hand, the metacognitive model (Model II) applied in this chapter determines that "father," "mother," "brother," "sister" are semantically closest because they share similar dynamic history in the sociocultural context, and such a context, as the model argues, is encoded in the semantic categorization of kinship in natural language. Empirically, both types of grouping can be found in empirical results summarized in Table 2.3, which suggest the possibility that groups of participants may use different cognitive representations, one driven by weighted co-occurrence patterns and the other driven by metacognition of sociocultural connotations, to make semantic grouping and typicality judgments. Applying both models to the semantic domain of kinship in English points towards an intriguing general research direction for computationally and empirically exploring the possibility of competing cognitive representations for certain semantic categories driven by complex socio-pragmatic relations in society and culture.

5.5. Implementation of Model II on basic-level smell semantics in American English from 1800 to 2020

The last empirical application in this chapter focuses on the domain of smell terms. In this group of tests, I selected six adjectives describing "good smell" ("fragrant," "aromatic," "perfumed," "savory," "scented," "redolent"). Simulations are also performed on the corresponding nouns of these six adjectives.

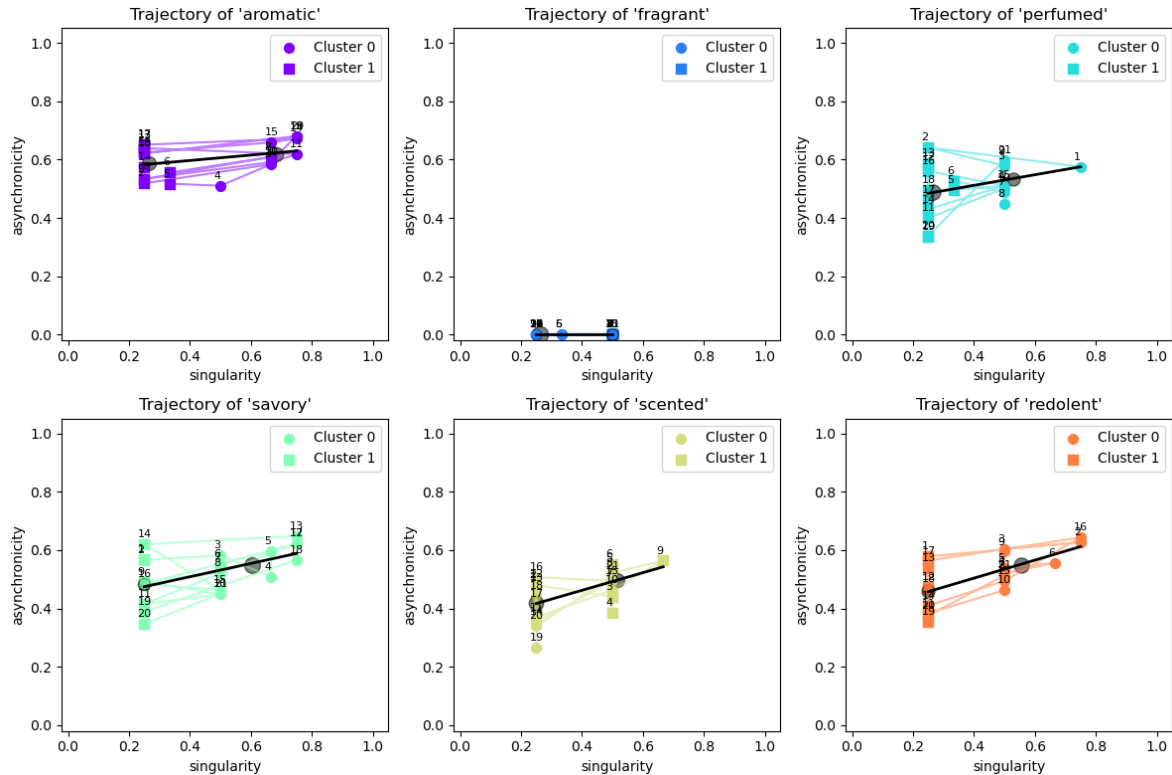


Figure 5.6. Diachronic trajectories of basic-level smell semantics mapped in the singularity-asynchronicity space from 1800 to 2020 in American English, with a 20-year window size and a 10-year step. Both smell adjectives and their corresponding nouns are simulated (only visualization for the former is shown here). For smell adjectives, "fragrant" is chosen to be the reference word the asynchronicity level of which is always set to be zero. Procedures to generate these empirical dynamic scenarios are the same with the case of basic-level color and kinship semantics. Results here indicate that basic-level smell terms (with positive valence) in American English corpus may have been consistently in the stable metacognitive mode with two equilibria (5.1a: "citizen") and overly stabilized at least since 1800.

Figure 5.6 shows the results for six smell adjectives. When "fragrant" is the reference word for synchronicity, the other five English adjectives for smell show patterns similar to those for the basic-level color words in English: a typical "citizen" scenario emerges for all these terms, where cluster centers align along the simulated linear nullclines and the slopes of all nullclines are shown to be positive. This indicates that perhaps similar to most basic-level color

terms, cognitive representations of "good smell" semantics at the word level in English have been relatively stable for the last 200 years. In addition, these smell adjectives together have less variability in both synchronicity and singularity (Standard deviations (SDs) for the slope of fitted lines and the sizes of two clusters for all smell adjectives are 0.077, 0.748, 0.748 respectively; data from Table 5.5) than basic-level color and kinship semantics during the same historical time (SDs for the slope of fitted lines and the sizes of two clusters for all color terms are 0.176, 2.000, 2.017 respectively, and for all kinship terms 0.192, 2.511, 2.501 respectively). This indicates that from the dynamical perspective, the overall semantic and cognitive stability of basic-level smell semantics are even stabler and less metacognitively varied than basic-level color semantics in English. Along with the results in Chapter III that basic-level semantics of good smell in ancient Chinese were much more stable since the early medieval period than before that³⁶, it is possible that smell semantics are overly stable in certain language-related sociocultural histories and may be related to the paucity of smell language observed in many languages and societies (Majid, 2021).

³⁶ Work in Chinese historical linguistics, nonetheless, indicates that such stability of smell semantics may not be the case for periods earlier than the medieval periods (Tong, 1984). But as the data size for the corpus in early Chinese periods is extremely small, novel quantitative techniques are needed for computationally addressing this possibility.

American English Ref word: Fragrant Period: 1800- 2020	Regression Coeff.	Size of the majority cluster	Size of the minority cluster	Alignment between fitted lines and cluster centers	Predicted dynamic scenarios
fragrant	0	14	7	strong	5.1a
aromatic	0.093	11	10	strong	5.1a
perfumed	0.181	12	9	strong	5.1a
savory	0.227	13	8	strong	5.1a
scented	0.302	11	10	strong	5.1a
redolent	0.309	12	9	strong	5.1a
Mean (μ)	0.222	11.8	9.2	overall: strong	overall: 5.1a
SD (σ)	0.077	0.748	0.748		

Table 5.5. Dynamical scenarios of basic-level smell terms in American English for the period from 1800 to 2020. Reference word = "fragrant". All the smell terms are classified into one stable metacognitive scenario 5.1a.

5.6. Discussion

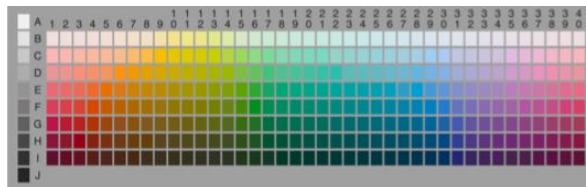
This section further discusses how the modeling work represented in this chapter can be contextualized in contrast to two major yet distinct cognitive dynamic approaches to language variation and change in the literature: the domain-oriented communicative approach and other dynamical-systems approaches relating cognition with language change. Based on the theoretical results in Chapter IV and empirical results in Chapter V, the advantages of the metacognition-based dynamical systems model compared with these approaches are also discussed. As the metacognitive model is formally inspired by one type of single-neuron dynamical modeling in computational neuroscience, I also further review the potential conceptual and formal connections between dynamical approaches to neural systems and to sociocultural phenomena.

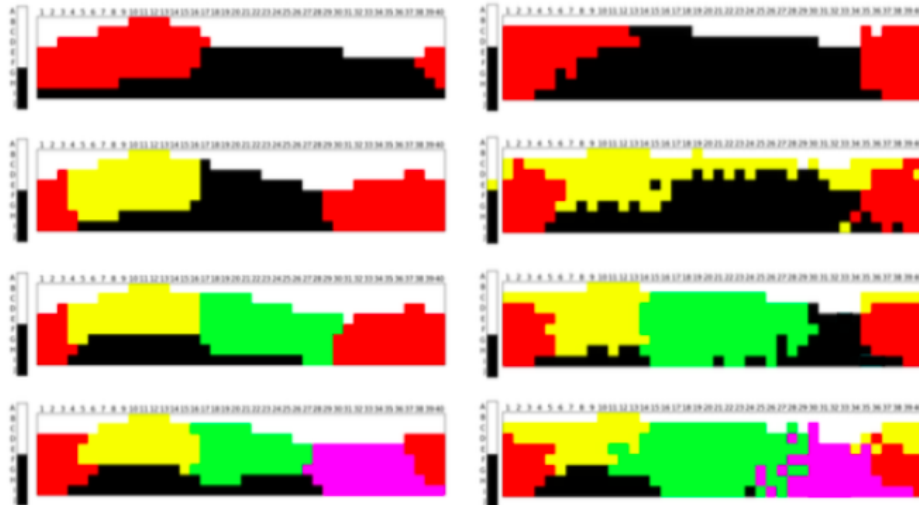
5.6.1. The communicative efficiency framework and the evolution of basic-level color terms

As highlighted earlier, in the cognitive science literature a significant framework that seeks to interconnect the diachronic and synchronic aspects of language and perception originates from the study of basic color semantics across cultures (Berlin & Kay, 1969). This framework has been further refined into a computational model by Regier et al. (2009; 2015), providing it with a stronger cognitive-functional basis. According to this framework, basic color words are considered "near-optimal" in a cognitive sense for designating and communicating the

perceptual experience of color. From this perspective, language bolsters communicative efficiency by maintaining a tradeoff balance between accuracy and complexity, enabling language users to name and convey perceptual experiences to others (Figure 5.7). This hypothesis has undergone rigorous testing through behavioral and computational studies employing color words from 110 languages, and extensions are also made to investigate kinship and object shape naming patterns across many languages (Zaslavsky et al. 2019). Recently, the evolutionary prediction of color terms, as initially developed by Berlin and Kay (1969), has been computationally probed by applying the theory of information bottleneck to the question (Zaslavsky et al. 2018; Figure 3.2). This study asserts its support for the evolutionary prediction based on prior research on color naming, and advocates for the extension of this framework to other linguistic and cognitive domains for more comprehensive evaluation.

A.





B.

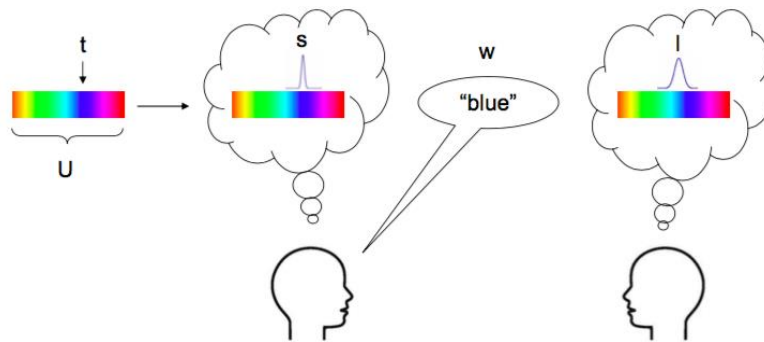


Figure 5.7. A: Theoretical optima of computationally partitioning the human perceptual color space (top) for $n=3,4,5,6$ categories (left, top to bottom), compared with color naming systems (right, top to bottom) of Ejagham (Bantoid, Nigeria/Cameroon), Culina (Arawan, Peru/Brazil), Iduna (Austronesian, Papua New Guinea), and Buglere (Chibchan, Panama). B: A scenario illustrating informative communication. The speaker mentally represents this target object as a probability distribution s over U , centered at t . The speaker then attempts to convey this mental representation to a listener, by producing a word w —here, the word "blue". The listener wishes to understand the speaker's intended message, and to that end attempts to mentally reconstruct the speaker's representation s from the word the speaker has just used to describe it. The listener's mental reconstruction l is also a probability distribution over U , and is intended to approximate the speaker's mental representation. The category named by the word "blue"—which will in most cases be broader than s (Adapted from Regier & Kay, 2015).

In particular, this group of cognitive-computational studies on color naming presumes a specific functional need for language with respect to perception, and it develops predictions on the potential trajectory of language change across cultural history based on such a functional assumption at the cognitive level. However, both theoretically and empirically, these studies only indirectly engage with language change: the synchronic observations on color terms across cultures are translated into diachronic predictions on how these terms might evolve over time. Furthermore, a segment of critics contend that the functional conception of these studies is too narrow to study language meaning. For these critics, these studies emphasize too much the reference and denotational aspects of language meaning (Sanders, 1992; Lucy, 1997).

The metacognitive model proposes an alternative framework. While not directly addressing the functional foundation and computational assumption of the above framework, my semantic model is built upon a different functional hypothesis (metacognition) and a different computational framework (dynamical systems). Empirically, my model establishes a reciprocally dynamical connection between the synchronic/ cognitive and diachronic/historical dimensions regarding basic-level semantic categories. Further, though the goal of the modeling is to evaluate and test domain-specific hypotheses, the methodology itself is not confined to a particular domain. These features are intended to address the theoretical and empirical challenges previously noted in the traditional color-naming research paradigm. Through the direct engagement of synchronic

and diachronic perspectives, and the flexibility of a non-domain-limited methodology, the model brings a fresh perspective to tackle enduring questions in the study of domain-oriented semantic categories.

5.6.2. Other dynamical systems approaches relating cognition with language change

As previously mentioned, using a dynamical-systems approach to study language change is not new, but the majority of such work simulates language change mostly within contexts of a single-equilibrium or single dynamics scenario (Section 4.6). Applying methods of dynamical systems and iteration maps, Niyogi (2006) models diverse processes and paths of phonological or grammatical changes based on a number of computational hypotheses regarding the linguistic rule-learning processes across generations. Each of these language learning algorithms determines a specific computational-cognitive procedure to acquire a specific linguistic rule out of two possible choices (later in the book he extends the model to multiple choices). Some of the cognitive-computational learning algorithms include what he terms "memoryless learner," "cue-based learner," and "batch learner":

- Memoryless learner: updates the decision regarding which rule it learns based on every input from the speaker of the last generation;

- Cue-based learner: assumes a minimum threshold of the number of any type of inputs that the algorithm receives for deciding on which rule it would acquire;
- Batch learner: determines the rule based on the proportion of one kind of input that suggests a linguistic rule over other kinds of inputs suggesting an alternative linguistic rule.

Examples of learning a specific linguistic rule include whether the language speaker tends to put a sentence-final particle before or after a verb, or to pronounce the monophthong version of a word or not. Niyogi (2006)'s modeling shows that the three assumptions could yield different dynamics and outcomes of how a specific linguistic rule is learned or not in the population.

One example of empirically successful dynamical modeling for morphological and syntactic competitions focuses on the S-shape curve type of change (Niyogi and Berwick, 1997; Yang, 2002; Niyogi, 2006), where one morphological pattern or syntactic rule steadily replaces the other over a period of time. Empirically, S-shape curves are observed in many case studies on morphological and syntactic change in multiple languages (Santorini, 1993; Taylor, 1999; Haerberli, 2002). The dynamical-systems approaches developed to model this type of change usually start with the change of the probability of the speaker learning a linguistic pattern over generations, and thus in this type of modeling, language changes at the macro-historical level are directly driven by micro-level cognitive mechanisms of language learning. Figure 5.8 shows an empirically observed historical change from 1400 to 1950 of how the position of inflectional head (INFL) in Yiddish changed from predominantly INF-final to an INFL medial one (Santorini

1993), which Niyogi and Berwick (1997), Yang (2002), and Niyogi (2006) successfully develops dynamical models to account for.

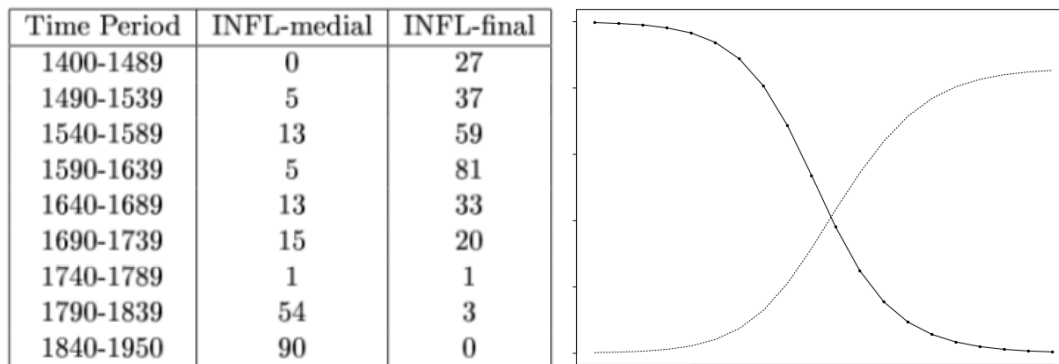


Figure 5.8. Left (Adapted from Niyogi, 2006): changes in the position of inflectional head (INFL) in the Yiddish sentence from 1400 to 1950, during which Yiddish changed from a predominantly INFL-final language to a NFL-medial one (Santorini, 1993). Similar shifts are also observed in the history of Greek and English (Taylor, 1999; Haerberli, 2002). Right: (Adapted from Niyogi and Berwick, 1997): based on various cognitive language learning algorithms, a cognitive-driven dynamical-system model developed in Niyogi and Berwick (1997) is able to simulate such a type of change, by which the plot is generated. Percentage of a population using two alternative expressions L1 and L2 is measured on the *y*-axis, and the population’s evolving over generations is measured on the *x*-axis. Yang (2002) develops an alternative dynamical model to generate similar patterns of S-shaped curves for syntactical competition.

Nonetheless, modeling on semantic and pragmatic changes have been far less explored in the literature of language change, and one reason could lie in in the fact unlike most cases in phonology, morphology, and syntax, determining proper or definite dimensions and spaces for semantic or pragmatic properties is much more difficult and often implausible. Researchers in this field have widely disagreed on how to practically determine these dimensions or spaces for computational modeling of semantics or semantic change (Yang, 2002).

5.6.3. *Advantages of the metacognitive model over other dynamical and non-dynamical approaches*

In the color naming research of Regier et al. (2015) and Zaslavsky et al. (2018), the meaning space is not rule- or symbol-based as in many existing works of dynamical modeling, but is mapped to categorical perceptual similarity in a specific perceptual domain. Both lines of research face difficulties regarding determining definite semantic spaces for computation beyond the limited cases they study. The metacognitive model, on the other hand, does not assume an *a priori* definition of semantic space; rather, it generates domain-specific semantic categories directly from corpus data, only with minor interference from the researcher.

In addition, the communicative efficiency and information bottleneck approaches to color naming evolution does not incorporate linguistic competition (Regier et al., 2015 and Zaslavsky et al., 2018). While Niyogi and Berwick (1997) and Niyogi (2006)'s modeling studies focus on competition, his approach simulates language change mostly towards one single-equilibrium or within a single dynamic scenario. This approach may be effective for modeling a small range of cases like competitions between two rules of inflectional head (INFL) (Figure 5.7), but it cannot model many other long-term competitions, such as competitions that occur in multiple dimensions (such as basic-level color and smell semantics in ancient Chinese, as shown in Chapter III), competitions that are also driven by ensemble factors (such as color, kinship, and smell semantics in the last 200 years in English), or competitions that shift between different dynamic scenarios (such as scenario 5.1d that appears in multiple examples in the domains of

color and kinship in American English). As also shown in the empirical examples in this chapter, the metacognitive model also provides richer characterizations for non-competitive scenarios than both approaches, able to differentiate cases according to different types and degrees of metacognition (Figure 5.1) and different levels of stability (e.g., potential overstability in the basic-level smell semantics in modern English).

The metacognitive model also develops a more direct link between cognitive mechanisms and patterns of language variation and change than the other types of models discussed here. The model not only directly engages with the historical corpus data, but also proposes a theoretical framework on how cognition models semantic meaning via patterns of language change. According to the empirical definitions of singularity and synchronicity (Section 5.2), both basic variables are measured based on the assumption that exposures to language change in sociocultural history where the speaker is situated matter for developing cognitive knowledge of basic-level semantic categories.

Lastly, the framework that the metacognitive model develops also takes an integrative approach to semantic and pragmatic aspects of language cognition and changes. In the current NLP literature, diachronic modeling has become an attractive field relating statistical and distributional language modeling with socio-cultural phenomena (Hamilton et al., 2016; Garg et al., 2018; Kozłowski et al., 2018; recall Section 3.2.1 on more detailed review on this literature). It is not obvious, however, how the current statistical and distributional modeling framework is capable of differentiating or integrating cognitive and social factors that interact with diachronic change, and how it accounts for language variation and change causally and dynamically. These

difficulties have been fundamental for sociolinguistics since the 1950s (Labov 1994). On other hand, pragmatic modeling approaches, notably those encompassing social, cultural, and historical characteristics of texts and contexts, have also begun to become a new research focus in NLP (Wagner et al., 2015; Tanweer et al., 2021; Nguyen et al., 2021). The metacognitive modeling approach incorporates the cognitive and social aspects with questions on language change, and engages with the current literature both on language variation and change and on social pragmatics in sociolinguistics and linguistic anthropology (Silverstein, 1993; 2020; Eckert, 2018). From the perspective of the contemporary literature of sociolinguistics and linguistic anthropology, metapragmatics (Silverstein, 1976; 1993) is recognized as a fundamental link between large-scale language variation and change and the micro-level interactional mechanism. Developing a formal dynamical model based on metacognition helps connect questions on language variations and social pragmatics, which are still largely independently investigated in NLP.

5.6.4. Conceptual and formal inspirations from dynamical-systems-based models of single neuron and neural synchronization

Lastly, it is noteworthy to delve into a more in-depth discussion of the formal resemblance between the metacognitive dynamical-systems model and the Fitzhugh-Nagumo model, a classical model in computational neuroscience for modeling single-neuron behaviors (noted in footnote 15 in Section 4.4, Chapter IV). The Fitzhugh-Nagumo model is extensively

employed in the exploration of varied resting, excitatory, and oscillatory patterns in biological neural activities (FitzHugh, 1961; Nagumo et al., 1962). This section discusses possible conceptual and computational links binding the dynamical-systems modeling in the biological and cognitive-/socio-linguistic realms.

Neuroscientists take complex and dynamical views on relationships between individual neural spiking, circuit mechanism, and ensemble oscillation and synchronization patterns, and assume sophisticated computational and biological interactions between physiology and psychology (e.g., Usrey & Reid, 1999; Kay, 2014). Though a number of natural and social science studies in the late 20th century began to apply similar dynamical views to investigate social phenomena (Wallerstein, 1998; Prigogine, 2000), the success of large-data modeling in particular in the last decade has contributed to the dominance of statistical modeling methods in social sciences, which has marginalized dynamical modeling approaches in many social science disciplines. One major intention of developing and applying the dynamical metacognitive model in parallel with a statistical-learning-based model in this dissertation is to explore a common ground from which to investigate social phenomena through both dynamical and large-scale statistical points of view. On the other hand, many existing studies that use a dynamical-systems approach to such language variation and change are influenced by works in evolutionary biology and ecology (Nowak & Krakauer, 1999; Niyogi, 2006; Mufwene, 2008). Dynamical-systems approaches from computational neuroscience, nonetheless, provide frameworks and formal tools that are not most often used in evolutionary biology and ecology, but they may suggest alternative powerful dynamical-systems modeling methods. In particular, neurologically-driven dynamical models are intrinsically suitable for modeling metacognition and metacognition-

driven language and sociocultural phenomena, as both are phenomena that illustrate what neuroscientist and philosopher Walter Freeman names "circular causality" (Freeman, 2001; 2008; 2013).

In short, applications of Model II in this chapter further demonstrates that it is fruitful to draw connections between dynamical systems approaches in both biological sciences and sociocultural phenomena. Developing a language change model from a single-neuron model opens the door for further borrowing dynamical models of ensemble activities, such as oscillatory and synchronization phenomena to model ensemble dynamics in language and society.

5.7. Chapter summary

In this chapter, I derived five scenarios of metacognitive dynamics related to word-level diachronic language change. Then based on the American English corpus over the last 200 years from Google Ngram Viewer, the dynamical-systems model developed in Chapter IV was implemented to examine three specific semantic domains: color, kinship, and smell. The results provide computational understanding of how metacognition may be sensitive to social history, as well as how it may reciprocally shape basic-level semantic shifts within these three domains. For color semantics, the simulation shows that the metacognitive model can identify metacognitively sensitive sociohistorical-dependent properties of basic-level color words. In the domain of basic-level kinship semantics, three dynamic scenarios

are found to capture how diachronic changes of sociocultural connotations of kinship terms interact with the categorical representations of basic-level kinship terms in cognition. Within the domain of smell semantics, the dynamical analysis proposes a prediction regarding the potential overstability of basic-level smell semantics in the sociocultural histories related to American English. I then further discussed the advantages of the metacognition-based dynamical systems model in comparison to other dynamics-driven approaches, as well as the potential conceptual and formal connections between dynamical approaches to neural systems and to sociocultural phenomena. Together, studies in this chapter offer a new domain-informed cognitive-computational framework for future empirical work on modeling semantic categories, bridging cognitive dynamics and linguistic variation and change together.

Chapter VI. Extending the Two Modeling Approaches to Other Semantic Domains

6.1. Chapter introduction

This chapter will focus on extending the modeling and analytic approaches from previous chapters to three additional groups of semantic or task domains, about which substantial empirical work in the cognitive science literature is still lacking.

The first group focuses on utilizing model I to examine plural suffixes in English. In contrast to languages like Chinese and Japanese, which possess rich vocabulary of classifiers but sparse vocabulary of plural markers, many Indo-European languages such as English show opposite patterns, for which plural markers are often richly grammaticalized but classifiers are sparse (Aikhenvald, 2000). In linguistic relativity research, such cross-language structural discrepancy generates profound research interests in empirical studies. In particular, cognitive psychologists demonstrate that a language that possesses a rich vocabulary of singular and plural markers can make the native speaker of such a language more sensitive to the numerical dimension of the perceptual world after a certain developmental age. In contrast, a language that possesses a rich vocabulary of classifiers (which are often shape-based or shape-derived) can make its native speakers more sensitive to the shape dimension of the perceptual world (Lucy and Gaskin, 2001). Chapter III contains a study using Model I to examine the generation and development of classifiers in ancient Chinese. In this chapter, one proposed extension is to use Model I to examine how plural markers can be cognitively mapped based on the typicality-

contextuality mapping. The domains of body parts, time, and length will be investigated, and hypotheses on how plural suffixes in these semantic domains are cognitively significant for the American English speaker will be proposed. This group of studies also intends to extend the domain-specific analysis to multimodal cases, where plural markers and the domains of body parts or of time and length interact.

The second group of studies in this chapter will harness Model I to investigate pragmatic knowledge within a strong semantic and knowledge-based context, specifically focusing on place names. Place names are one of the classical semantic and knowledge domains used to evaluate and benchmark models in the contemporary distributional linguistics literature (Mikolov et al. 2013). While classical word embedding algorithms such as *word2vec* are highly effective at producing interpretable representations of world-related knowledge, alternative mapping of such knowledge has been relatively marginalized. Chapter II contains a preliminary example of applying the typicality-contextuality model to map relations between city/capital names, which shows that even semantic knowledge as simple as place names can be modeled through alternative yet effective mapping strategies, with representations and evaluative benchmarks different from representations in vector-based algorithms (Figure 2.7 left). The second group of studies in this chapter further proposes that semantic knowledge generated from the corpus that is strongly oriented towards "objective" semantic knowledge can be pragmatically decoded. The embedding of country names is produced and modeled by Model I from *Wikipedia* data in English to illustrate this point.

The third group of studies uses Model II to determine how much different semantic domains or semantic categories are metacognitively sensitive, based on corpus-based modeling of language change in society and culture. A number of results in Chapter V demonstrated the possibility that pragmatic and social cognition are relevant to modeling semantic categories within the domain of color, kinship, and smell languages. This group of studies will further apply this methodology to domains that are considered to be highly pragmatically and metacognitively saturated. These domains often form important research areas in other social science disciplines such as political science, sociolinguistics and linguistic anthropology, and sociology. Specific domains under study in this section include a number of basic-level terms regarding political attitudes, religions, and professions.

Overall, the intent of this chapter is to present the power and potential of these two innovative models in understanding and interpreting a myriad of semantic domains other than color, kinship, and smell. The exploration within this chapter hopes to generate further investigations and applications in domain-oriented studies of semantic categories in future computational and empirical studies in cognitive science.

6.2. Methods

The method to select specific words of a certain semantic domain is similar to methods used in previous chapters for the domain of color, smell, kinship, and shape. Empirically, domains that are selected to study in this chapter are also demonstrated to be important in

psychological research on typicality and similarity knowledge categories (Uyeda & Mandler, 1980; Banks et al., 2022). For each domain, initial steps of word choice are in reference to words that have been chosen in psychological studies of semantic categories (primarily with Uyeda & Mandler, 1980 and Banks et al., 2022). The similarity algorithm of Model I is then used to generate words that are categorically related to words that are initially manually chosen. Words that are chosen by both the algorithm and the empirical studies in either Uyeda & Mandler (1980) or Banks et al. (2022) are immediately included. Words that are chosen by the algorithm but do not appear in either study are judged and filtered according to the author's knowledge of the corresponding semantic category.

All studies in this chapter will be performed on the same randomly chosen *Wikipedia* corpus in English of 8 M tokens.

6.3. Plural suffixes of basic-level body parts, time, and length words (by Model I)

Empirical research in language and thought demonstrates that plural suffixes and shape-based classifiers may be in a semantical and cognitive trade-off relationship across languages (Lucy and Gaskin, 2001). Most East Asian languages, including but not limited to Chinese, Japanese, Korean, Vietnamese, Malay, Burmese, Thai, and Hmong, as well as numerous South Asian languages such as Bengali and Munda, feature shape-based classifiers prominently. Classifier systems are also observed in Australian Aboriginal languages, specifically in Yidiny and Murrinhpatha, and in Mesoamerican languages, including Classic Maya and its

contemporary derivatives like Cuchumatán Mayan and Otomanguean (Aikhenvald, 2000; Hopkins, 2012). On the other hand, plural suffixes are not widely observed in these languages. In contrast, most European languages usually develop a rich system of plural markers but in general lack a well-developed classifier-system. In cognitive science, it is observed that such a cross-language discrepancy has an important cognitive consequence. For example, Lucy & Gaskin (2001) find that children's classification preferences for objects at a certain developmental stage start to align with their language group, depending on whether the language has a specific type of classifier system or not.

Chapter III includes a study employing Model I to investigate the evolution of three main groups of shape-based classifiers in ancient Chinese. This part further provides two preliminary studies using Model I to examine how plural markers in contemporary English can be mapped based on the two cognitive dimensions of typicality and contextuality. The first study in section 6.3.1 focuses on the domain of basic-level human body part terms, and the second study in section 6.3.2 centers on the domain of basic-level time and length words.

- 6.3.1. Mapping singular and plural forms of basic body-part terms in contemporary English (in comparison with contemporary Chinese)

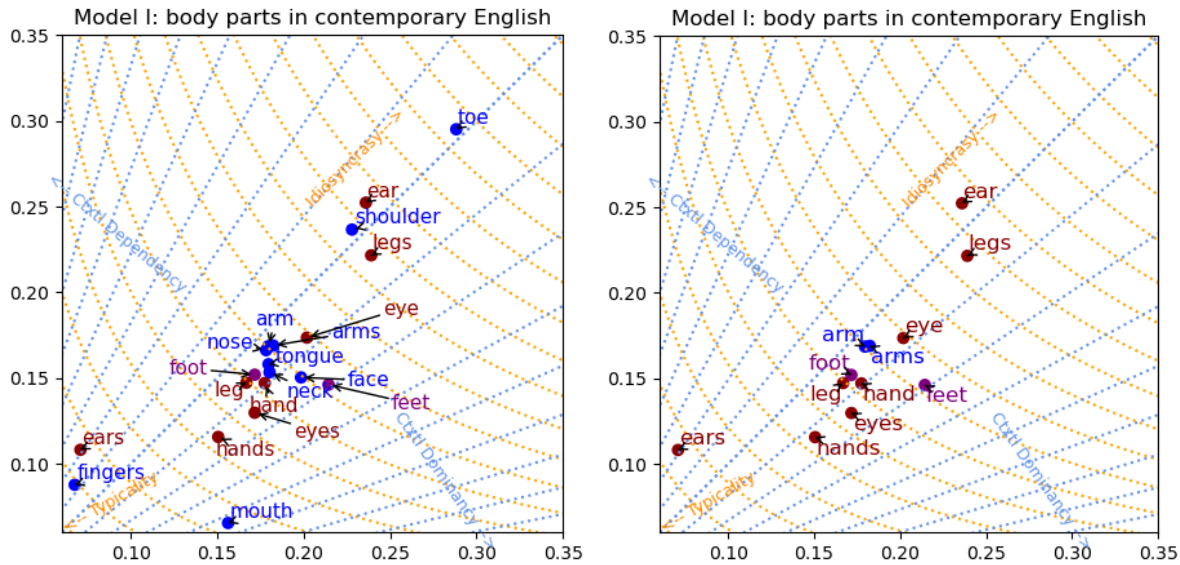


Figure 6.1. Typicality-contextuality mapping of basic-level body-part terms in English as identified by the similarity algorithm. Left: In general, the terms with both singular and plural forms (depicted in dark red or purple) exhibit more densely distributional patterns on the typicality axis than those with either one form being extracted or undifferentiated forms (blue). Right: The distribution of singular and plural forms of body-part nouns in English on the typicality axis is speculated to be approximately ordered by the proportion of the body part's homunculus somatosensory size over its physical size (see Figure 6.2). For example, plurals of body parts with large homunculus-to-physical size ratios, such as "ears," "hands," and "eyes," exhibit greater linguistic typicality than their singular counterparts. Conversely, for "legs" with a smaller homunculus-to-physical size ratio, the plural form is less typical than its singular form "leg". Data: an 8 M tokens of *Wikipedia* corpus in English.

Figure 6.1 shows how the basic-level terms in English extracted by the similarity algorithm are mapped in the typicality-contextuality representation. Figure 6.1 (left) shows the typicality-contextuality mapping for all the extracted terms, and Figure 6.1 (right) shows the same results only for terms for which both singular and plural forms are extracted. Three main observations are drawn from the mapping. First, terms of which both singular and plural forms (in dark red or purple) are extracted are more densely distributed on the typicality axis than terms

of which either singular or plural forms are extracted (blue). Second, except for "foot" and "feet"³⁷, terms for which both singular and plural forms are extracted are predominantly aligned along the typicality-idiosyncrasy axis (Figure 6.1 right). Third, except for "foot" and "feet," the distribution of singular and plural forms of English body-part words on the typicality/idiosyncrasy axis shows a correlation with the homunculus model of the human somatosensory system (Figure 6.2 right). Intriguingly, the singular-plural forms distribution shows an ordering related to the proportion between the homunculus size and the physical size of the body part that a term denotes. Plurals of body parts that have a disproportionately large representation in the homunculus model compared to their physical size, such as "ears," "hands," or "eyes," are linguistically more typical than their corresponding singular forms ("ear," "hand," or "eye"). However, for a body part like leg, which has a small representation in the homunculus relative to legs' physical size, the plural form "legs" is mapped as linguistically less typical than its singular form "leg". "Arm" and "arms" form a case in between, as there's no significant differentiation in the mapping between the plural "arms" and its singular counterpart "arm."

³⁷ The exceptional status of "foot" and "feet" is obviously due to the fact that they are also length units in modern English.

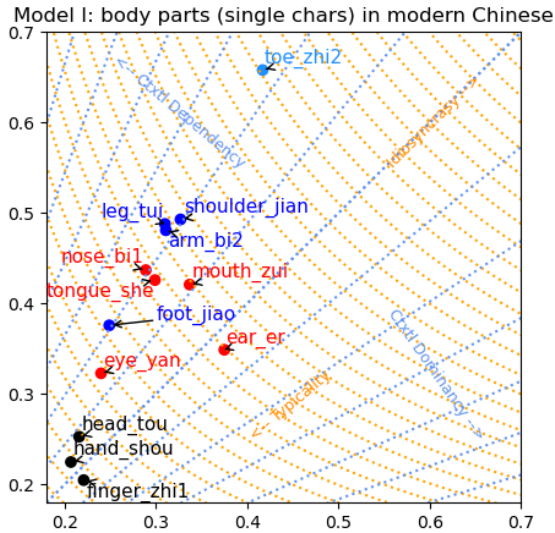


Figure 6.2. Left: Typicality-contextuality mapping shows that the semantic mapping of single-character body part terms in standardized contemporary Chinese is influenced by the homunculus size and physical size combined, with homunculus size being more weighted, instead of by the proportion between the two sizes as in cases of singular and plural forms of body-part terms in English. The most typical group (marked in black) represents areas with the largest homunculus and physical sizes. The second most typical group (marked in red) corresponds to areas with a large homunculus size but small physical size. Less typical are parts with smaller homunculus but larger physical size (marked in blue), and the least typical group (marked in dodger blue) comprises areas with small sizes in both dimensions. With the exception of the term "foot," all other body-part terms are mapped in this specific somatosensory order. Right: a 3-D representation of the somatosensory homunculus model³⁸. The cortical homunculus represents the area of the human brain's somatosensory cortex dedicated to sensory functions. Data: an 8 M tokens of *Wikipedia* corpus in standardized Chinese.

In comparison, single-character body part terms in standardized contemporary Chinese, which lack plural markers, are also mapped in the typicality-contextuality representation. Like their English counterparts, basic-level body-part terms in standardized contemporary Chinese

³⁸ Created by Sharon Price-James and adapted from <https://www.britannica.com/science/homunculus-biology>.

primarily align along the typicality/idiosyncrasy direction. However, unlike English, the distribution of Chinese body-part terms along this axis appears to be influenced by the body part's homunculus size and its physical size combined, with the homunculus size carrying more weight (Figure 6.2, left), instead of the proportion between these two sizes as in the cases of singular and plural forms of body-part terms in English. Body parts with large homunculus and physical sizes such as *tou* "head," *shou* "hand," and *zhi1* "finger" tend to be linguistically most typical (marked in black, Figure 6.2, left). Those with a large homunculus size but small physical size, like *yan* "eye," *bi2* "nose," *er* "ear," *zui* "mouth," and *she* "tongue," are usually ranked as the second most typical group (marked in red, Figure 6.2, left). Conversely, parts with a smaller homunculus size but larger physical size like *jian* "shoulder," *bi2* "arm," and *tui* "leg" are generally less typical (marked in blue, Figure 6.2, left). Body parts with small homunculus and physical sizes such as *zhi2* "toe" are typically ranked as the least typical (marked in dodger blue, Figure 6.2, left). With the exception of the term *jiao* "foot," which is mapped as the second most typical group instead of the third where it should theoretically fall, all other terms adhere to this specific somatosensation-related logic.

In short, basic-level body-part terms in English which have both often-used singular and plural forms are more densely distributed on the typicality axis. Further, they show a correlation with the homunculus model of the somatosensory system, where the distribution of singular and plural forms relates to the proportion between the homunculus size and the physical size of the body part. In standardized contemporary Chinese, though basic-level body-part terms also primarily align along the typicality/idiosyncrasy direction, they are influenced by both the homunculus size and physical size combined, with the homunculus size carrying more weight.

The overall mapping of body-part terms in both languages adheres to a somatosensation-related logic, at the same time the mapping is mediated by whether the basic-level body-part terms have plural markers or not.

- 6.3.2. Mapping singular and plural forms of basic terms of time and length in contemporary English

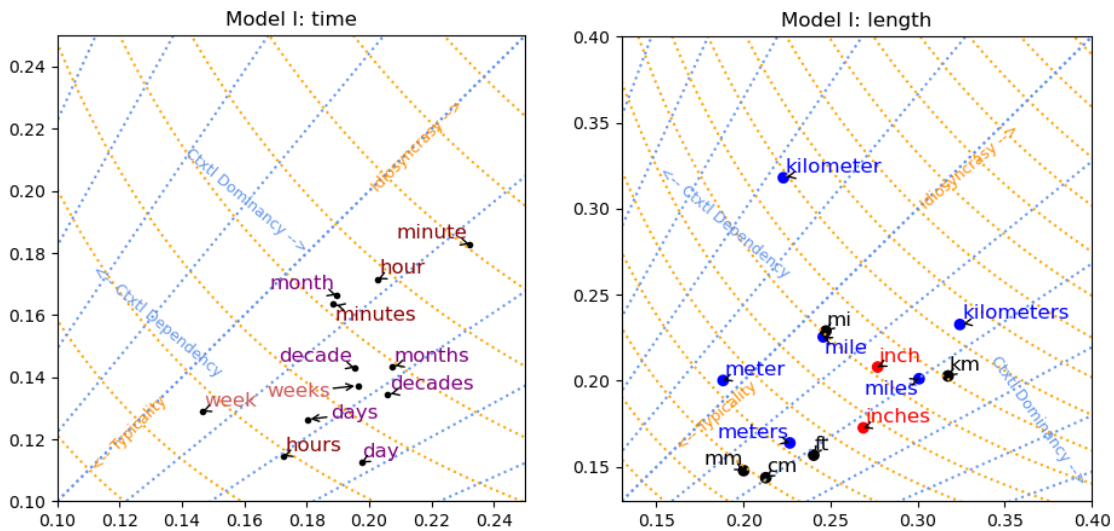


Figure 6.3. Mapping singular and plural forms of basic-level time and length terms in English reveals distinct perception-related ordering. Two patterns emerge: one aligns on the typicality axis with similar contextuality degrees, the other aligns on the contextuality axis with similar typicality degrees. Abbreviated length terms which do not possess singular and plural forms ("mm," "cm," etc.) align primarily on the typicality axis, similar to other semantic categories in English that are studied in this dissertation. The two plots are scaled differently to achieve the optimal visual resolution.

The mapping of singular and plural forms within the domain of basic-level time and length terms in English reveals a perception-related ordering different from that seen in the domain of body parts. Figure 6.3 shows English time and length terms both the singular and plural forms of which are extracted by the similarity algorithm of Model I from a *Wikipedia* corpus of 8 M word tokens. Two major ordering patterns emerge. The first pattern, which is the characteristic alignment pattern for the domains in English that have been investigated in this dissertation, predominantly aligns on the typicality dimension with similar degrees of contextuality. The second pattern, which is the majority alignment pattern for English countable time and length nouns examined in this section, predominantly aligns on the contextuality dimension with similar degrees of typicality. Among all countable time and length terms being extracted by the algorithm, those that follow the former pattern include "minutes" and "minute," "hours" and "hour"; and those that follow the latter pattern include "days" and "day," "decades" and "decade," "months" and "month," "meters" and "meter," "miles" and "mile," "kilometers" and "kilometer".

These binary patterns suggest that in English, plural suffixes might contribute to creating heterogeneous representations of time and length scales associated with countable nouns, underscoring their potential cognitive implications. In our examples, the plural forms "hours" and "minutes" exhibit similar contextuality to their corresponding singular forms ("hour" and "minute"), but display greater degrees of typicality. On the other hand, plural forms like "decades," "months," "meters," "miles," and "kilometers" exhibit similar typicality to their

singular counterparts ("decade," "month," "meter," "mile," "kilometer") but show a higher degree of contextual dominance³⁹.

In comparison, the algorithm also extracts related length terms that do not possess the singular and plural differentiation, including abbreviated length terms like "mm," "cm," "ft," "mi," and "km". Results show that they predominantly align along the typicality axis and are highly similar on the contextuality dimension, which follows the characteristic alignment pattern of the basic-level semantic categories in English that have been examined in previous sections.

In short, while basic-level semantic categories in English that have been studied in this dissertation follow a predominant alignment pattern that contains wider distributions on the hierarchical (typicality) dimension but narrower distributions on the similarity (contextuality) dimension, the alternative alignment pattern can emerge when plural markers are present in certain domains. This alternative pattern allows a wider distribution on the similarity (contextuality) dimension but a narrow distribution of the hierarchical (typicality) dimension. Recalling that the group of most typical shape-based classifiers in medieval and modern Chinese (classifiers derived from one-or-two dimensional shapes; Section 3.5, Chapter III) are also widely distributed on the similarity dimension but not on the typicality dimension, studies in this subsection offer new theoretical possibilities and empirical predictions for further investigating

³⁹ While "day" and "days" also display similar typicality yet differentiated degrees of contextual dominance, "day" exhibits higher contextual dominance than "days," a pattern which is inverse to the previously mentioned examples. This contrast invites further investigation to discern its cognitive implications.

the heterogeneous cross-linguistic distributions of classifiers and plural markers that have been investigated in both cognitive science and linguistics.

6.4. Pragmatic knowledge of country-naming (by Model I)

As mentioned in Section 2.7 in Chapter II, place naming forms one of the most classical benchmarks for embedding algorithms in the NLP literature (e.g., Mikolov et al., 2013; Pennington et al., 2014). Results in Section 2.7 also show that the typicality-contextuality mapping can model both the hierarchical and similarity relationships for capital and country names.

This section moves further by showing that the typicality-contextuality mapping that Model I generates also has both strong semantic *and* pragmatic implications for modeling place-naming. Figure 6.4 shows that the typicality-contextuality mapping of 66 country names in English (based on a *Wikipedia* corpus of 8 M token size) not only follows the characteristic pattern of basic-level semantic-category distributions in English (i.e., aligning on the typicality dimension rather than contextuality dimension), but also shows high sensitivity to these nation-states' comparative power and international influence as well as geopolitical context centered around the USA.

Applying the hierarchical clustering algorithm on the hyperbolic distances between these country names shows four groups of country names in English (Figure 6.4). The first two groups

are highly correlated with the country's comparative power and international influence, measured by a survey study ranking 60 most influential countries conducted in 2021 (CEOWorld, 2021): among the group of countries with overall highest typicality scoring in the mapping (marked in red), 17 out of 19 countries are ranked in the 60 most influential countries, and among the group of countries with overall lower typicality scoring (marked in dodger blue and light blue), only 10 out of 32 countries are ranked in the 60 most influential countries.

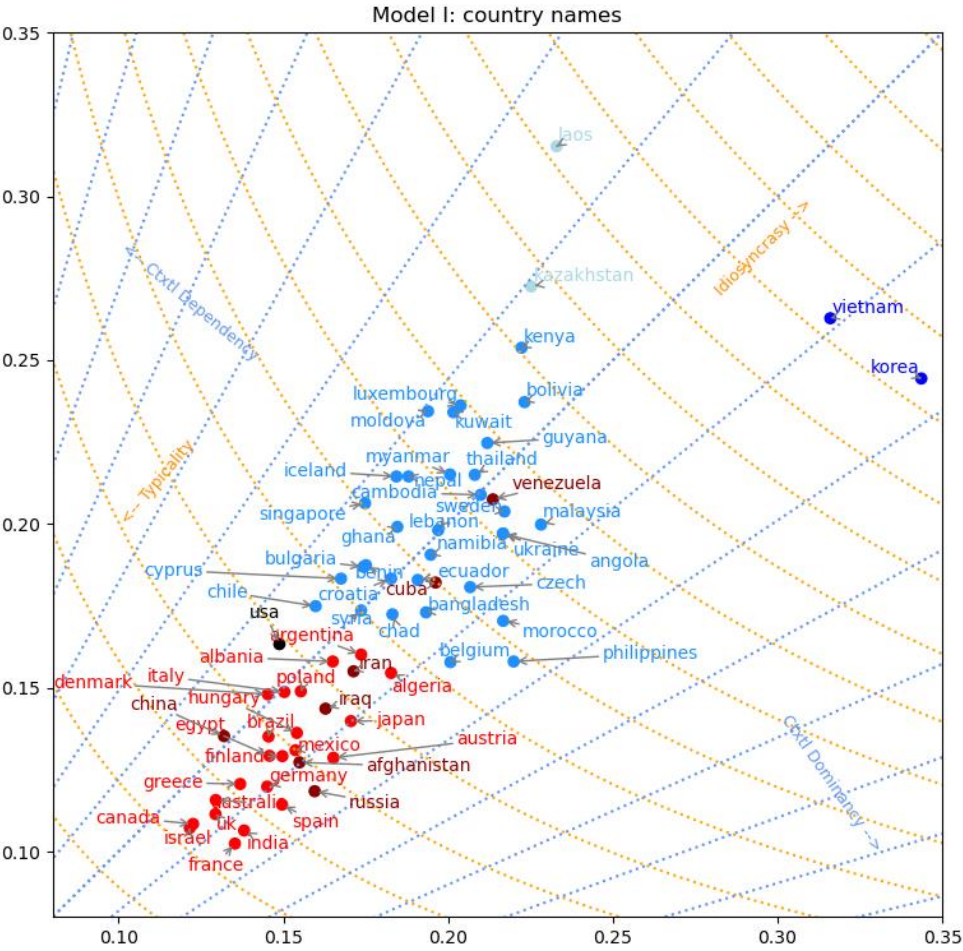


Figure 6.4. The typicality-contextuality mapping of 66 country names in English, based on a *Wikipedia* corpus, exhibits both semantic and pragmatic characteristics of nation state names. Hierarchical clustering reveals four groups of country names (red, blue, dodger blue, light blue) the results of which are arguably sensitive to comparative power, international influence, and geopolitical context centered around the USA. Nonetheless, countries that are most crucial to the contemporary geopolitical competition of the USA (Russia, Afghanistan, etc.) do not completely align with the results of hierarchical clustering (which are manually marked in dark red).

The second two groups consist of countries that are most crucial to the contemporary geopolitical competition of the USA. Russia, Afghanistan, Iraq, China, Russia, Cuba, and Venezuela are not fully computationally distinguishable in the mapping (which are manually marked in dark red in Figure 6.4), which is best accounted for by their pragmatic importance for contemporary US international politics. In contrast, Vietnam and Korea are mapped as the two most "idiosyncratic" (the least "typical") yet among the most "contextually dominant" country names (marked in blue). These two countries do not possess the most crucial direct influence on today's international politics, but they are arguably among the most important non-superpower countries that are relevant to the US's national historical memory since World War Two. In particular, Korea is ranked with both the highest degrees of idiosyncrasy and contextual dominance (= lowest degrees of typicality and contextual dependency) among all countries being mapped, highlighting its unique position in the linguistic-cognitive representation of country names in corpus-based English⁴⁰.

⁴⁰ For the case of Korea, the overlap between the historical memory of the Korean War in the 1950s and the contemporary geopolitical position of North Korea in the US international politics makes Korea a complex case to model. Further, for better distinguishing a geopolitical ally/partner and an opponent, it would be more helpful to consider North Korea and South Korea separately, but for grasping their geopolitical significance as a whole, mapping them together in a single term can be more productive. The purpose of this study, however, is not to exhaust all these empirical possibilities, but only to show the

In short, mapping country names using typicality-contextuality models demonstrates the model's sensitivity to multiple types of corpus-driven pragmatic knowledge in the place-name domain, showing the model's capacity of modeling both semantic and pragmatic domain-knowledge in a multidimensional way.

6.5. Metacognitive awareness for semantic categories of political attitudes, religions, and professions (by Models I and II)

This section shows three studies on modeling metacognitive awareness for semantic categories of political attitudes, religions, and professions. As mentioned in the discussion section in Chapter V (section 5.5), the meta-level, self-reflective aspect of language use is fundamental in the theories of contemporary linguistic anthropology and sociolinguistics, under the concepts of metasemantics and metapragmatics (Silverstein, 1976; 1993; Urban, 2006). Metalinguistics is the dominant analytic methodology for today's sociological and anthropological investigations of language-related social, cultural, religious, and political phenomena (Ji, in review). This section further explores the potential of using the metacognitive model developed in this dissertation to examine questions related to semantic categories in a number of domains that today's sociolinguists, anthropologists, and political scientists are most interested in. This section also shows examples in these areas that illustrate how the two models

methodological potential of this modeling approach to investigate these questions. Actual explorations of these questions are left to future work.

can be together employed in addressing the same research questions relating semantics, social cognition, and sociocultural phenomena.

All simulations conducted by Model II in this section are based on the Google Books corpus from 1800 to 2020, and the procedures are the same as in studies in Chapter V using Model II to investigate semantic knowledge and cognition in the domains of color, kinship, and smell.

6.5.1. Political attitudes

Based on a number of empirical studies of English speakers' typicality judgments that are previously used in this dissertation (Uyeda & Mandler, 1980; Banks et al., 2022), the two most typical words for political attitudes in English are chosen ("democracy" and "socialism"). Considering them as target words, Model I is then used to generate other terms that are close to the two terms based on a randomly chosen *Wikipedia* Corpus of 8 M word tokens. In this study, more adjectives related to political systems are generated than nouns. Hence these adjectives are used for the mapping in this study.

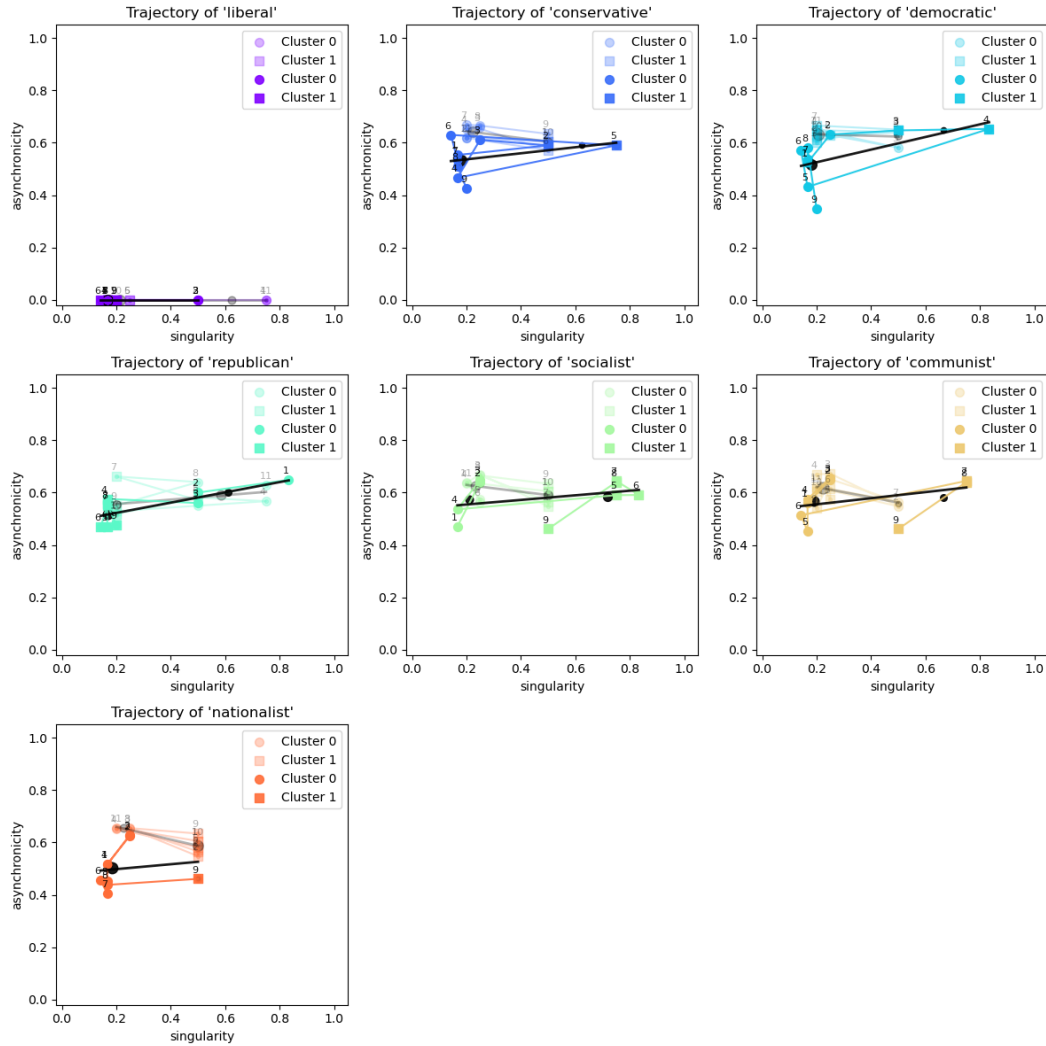


Figure 6.5. Model II's simulation results depict the evolution of political attitude adjectives from 1820 to 1920 (light colors) and from 1920 to 2020 (normal colors) in American English (with "liberal" being the reference term for synchronicity). The regression coefficients for these adjectives reveal cognitive patterns over time. In the 19th century, the terms exhibited non-reflective usage, except for "republican" and "democratic" which had weakly reflective usage. From 1920 to 2020, all political adjectives showed slightly to moderately positive coefficients, indicating overall increasing levels of metacognitive awareness of political ideologies, which are not specific to any political attitude.

The terms related to political attitudes that are selected by the algorithm include "democratic," "socialist," "liberal," "conservative," "republican," "communist," and "nationalist". Given historically, the adjective form "liberal" appears much more prevalently than "democratic." Therefore, "liberal" is selected as the reference word for the simulation by Model II in this study. The analysis follows the same procedure as the dynamical analysis on color, kinship, and smell domains based on Google Books corpus from 1800 to 2020 in Chapter V.

Figure 6.5 shows the results of Model II's simulation on the seven adjectives of political attitudes across two periods: from 1820 to 1920 and from 1920 to 2020 (respectively shown in light colors and normal colors in the figure). The results can be summarized as follows:

- (i). From 1820 to 1920, the regression coefficients for all political adjectives were either negative ("conservative," "socialist," "communist," "nationalist") or negative yet near zero ("republican" and "democratic") in reference to "liberal". This means that the former group can be characterized with the non-meta-level cognitive tradeoff scenario (scenario 5.1c: "scientist") and the latter group can be characterized with the weak metacognition scenario (scenario 5.1b: "child"). The cognitive interpretation is that for an average 19th century English speaker in America, "conservative," "socialist," "communist," and "nationalist" were not usually used or cognized in a reflective manner, and "republican" and "democratic" were typically used or cognized in a weakly reflective manner. Strong alignments between the fitted lines and cluster centers for all terms suggest that this pattern was also dynamically stable.

(ii). From 1920 to 2020, the regression coefficients for all political adjectives turn slightly or moderately positive. Ranked from the largest (0.244) to smallest coefficients (0.032), these terms can be arranged in the following order (from higher to lower levels of metacognitive awareness that each term entails): "democratic," "republican," "communist," "conservative," "nationalist," "socialist". This means that they can be characterized with either the metacognition scenario of "citizen" (scenario 5.1a) or the weak metacognition scenario of "child" (scenario 5.1b). The cognitive interpretation is that for an average 20th century English speaker in America, given the reference of the term "liberal," reflective awareness for other six terms of political attitudes systematically increased.

Together, the above observation shows that the reflective awareness of political attitudes has collectively increased from the 19th century to the 20th century in American English, and such an increase is not biased towards any specific political attitude, no matter the attitude pro-liberal or not. This observation can shed new light on the current research literature on the rise of contemporary political polarization (Levitsky & Ziblatt, 2018). The modeling results show that relationships between political attitudes can be measured at two language-mediated cognitive levels: while relationships between political attitudes in the 19th century are more prominently cognized as in the first-order cognitive tradeoff (scenario 5.1c: "scientist"), political attitudes in the 20th century are increasingly recognized at the higher-order, reflective level in cognition (scenario 5.1a: "citizen").

6.5.2. Religions

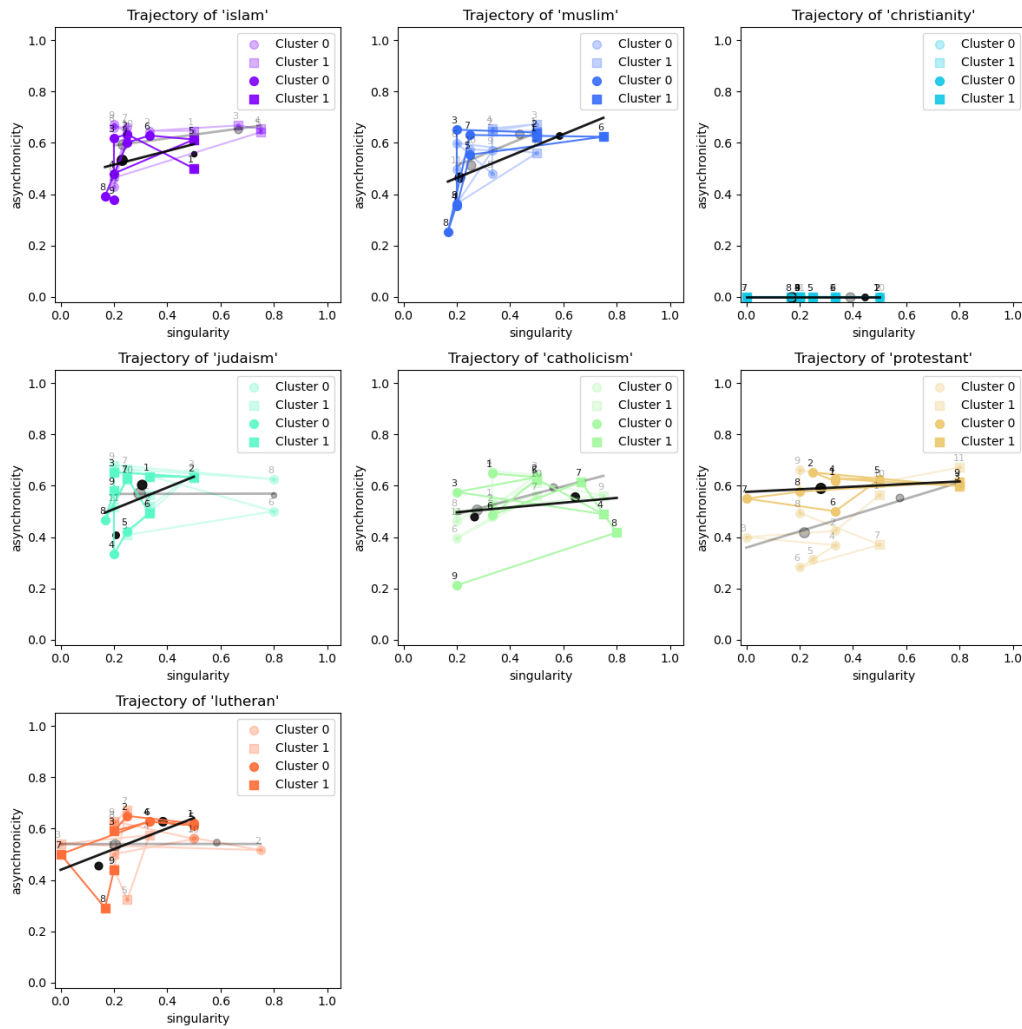


Figure 6.6. The results of Model II's simulation on seven religions from 1820 to 1920 (light colors) and from 1920 to 2020 (normal colors) are depicted. Regression coefficients reveal cognitive patterns over time. In the 19th century, "Muslim," "Islam," and "Catholicism" were

used in a strongly reflective manner, while "Protestant," "Judaism," and "Lutheran" were associated with weak or non-metacognitive modes. From 1920 to 2020, shifts occurred in coefficients and alignment patterns. "Islam" and "Judaism" underwent sharp changes and shifted between multiple metacognitive dynamics (5.1d: "politician" or "student"). "Catholicism" and "Lutheran" also transitioned between stable and unstable metacognitive modes. "Muslim" remained in the metacognitive scenario (5.1a: "citizen"), while "Protestant" remained weakly or non-metacognitive (5.1b: "child") throughout the past 200 years.

As Christianity plays a central religious role in Western societies, the term "Christianity" is chosen for the first target word. The similarity algorithm of Model I is then applied to the target word, generating other religion terms for the diachronically-based metacognitive mapping. In reference to "Christianity," the following terms are chosen by the similarity algorithm for the mapping by Model II: "Muslim," "Islam," "Judaism," "Catholicism," "Protestant," and "Lutheran." Other terms that are initially extracted by the similarity algorithm include "Buddhism," "Sunni," and "Hinduism," but their appearances in the American English corpus before the mid-20th century were negligible compared to other religions. Hence, they are not included in this analysis.

Figure 6.6 shows the results of Model II's simulation on the seven religions across two periods: from 1820 to 1920 and from 1920 to 2020 (respectively shown in light colors and normal colors in the figure). The results can be summarized as follows:

- (i). From 1820 to 1920, the regression coefficients for "Muslim," "Islam," and "Catholicism" are positive, and near-zero for "Protestant," "Judaism," and "Lutheran". This means that the former group can be characterized with the metacognitive scenario with two equilibria (scenario 5.1a: "citizen") and the latter

group can be characterized with the weak metacognition scenario (scenario 5.1b: "child"). The cognitive interpretation is that for an average 19th century English speaker in America, "Muslim," "Islam," and "Catholicism" were typically used or cognized in a strongly reflective manner, but "Protestant," "Judaism," and "Lutheran" were cognized in a weakly metacognitive or non-metacognitive mode. Strong alignments between the fitted lines and cluster centers for all these terms suggest that this pattern was also dynamically stable.

- (ii). From 1920 to 2020, the regression coefficients for "Islam" and "Judaism" changed sharply, and the alignments between the fitted lines and cluster centers turned from strong to weak. This means that in reference to the term "Christianity," both terms enter into a dynamical mode with shifts between multiple types of metacognitive dynamics (scenario 5.1d: "politician" or "student"). The coefficients and alignment patterns for "Catholicism" and "Lutheran" underwent the similar type of changes but with less intensity and appear to be transiting between stable and unstable metacognitive modes (between scenario 5.1a "citizen" and 5.1d "politician" or "student"). Dynamics for the terms "Muslim" and "Protestant" were relatively stable but they are characterized by different cognitive modes: while the term "Muslim" has been consistently in the metacognitive scenario 5.1a ("citizen"), "Protestant" has been consistently characterized by the weak or non-meta-level cognitive mode ("scientist") throughout the past 200 years.

An intriguing comparison that can be made with the results on political attitudes is that in the 19th century, terms of political attitudes are predominantly non-metacognitive, but terms of religious affiliations are predominantly metacognitive. While the metacognitive awareness of political attitudes systematically increased in the 20th century, the metacognitive awareness of religions shows more varied patterns of evolution. One can further interpret these patterns by incorporating two factors; the first factor is whether a religion belongs to Christianity or not, and the second factor is the semantic typicality of each religion's name in the secular discourse, the former of which can be measured by the typicality-contextuality mapping of Model I. Figure 6.7 shows the typicality-contextuality mapping of the seven religion terms. Together, these results demonstrate a cross-religion difference in relating typicality with shifts of metacognitive awareness. In the 20th century, non-Christian religions like Islam and Judaism with higher typicality and close semantic distances from Christianity still underwent tremendous shifts of metacognitive dynamics, whereas only Christian religions with much lower typicality and farther semantic distances from Christianity underwent metacognitive shifts, and these shifts were minor to moderate (Goddard, 2000; Tjørhom, 2020).

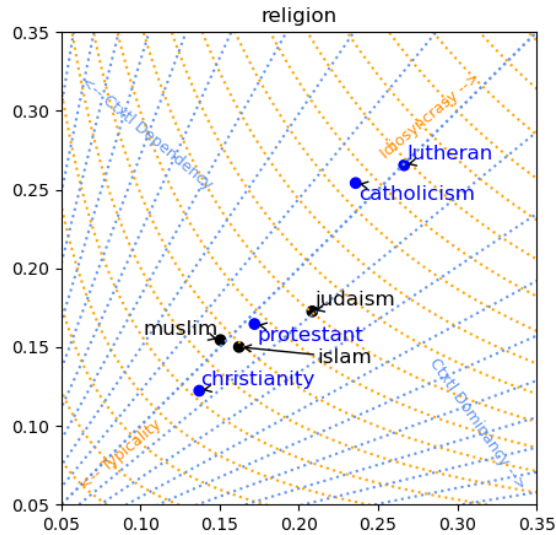


Figure 6.7. Typicality-contextuality mapping of the seven religion terms. Together, these results demonstrate a cross-religion difference in relating typicality with the shifts of metacognitive awareness of religious identities.

In short, the results from political attitudes and religious names together demonstrate the metacognitive model’s capacity of investigating language-related domain-specific sociocultural phenomena as well as performing sophisticated cross-domain comparisons. In addition, results on dynamic relationships between religions show the promise of employing both the distributional and dynamical models to investigate these questions in an integrative way.

6.5.3. Professions

Simulations in the domain of professions further shows the importance and potential of integrating the two semantic modeling approaches. Similar procedures are employed to extract semantically-related names of professions in American English.

In sociology of profession, one important research direction is about how the process of professionalization has been accelerated since the 19th century, particularly in fields relevant to natural science and technology (Abbott, 1988). In this preliminary study, I only select one profession ("poet") that is professionally distant from most fields of natural science and technology as the reference word for synchronicity in Model II, and apply the modeling approaches developed in this dissertation to briefly examine the process of professionalization from 19th to 21st centuries. Fifteen terms of professions are extracted by the similarity algorithm for further modeling: "mathematician," "physician," "astronomer," "architect," "philosopher," "theologian," "chemist," "historian," "painter," "diplomat," "poet," "novelist," "politician," "composer," "teacher." Figure 6.8 shows the results of using Model II to map the metacognitive dynamics for each of these terms across two periods: from 1820 to 1920 and from 1920 to 2020 (respectively shown in light colors and normal colors in the figure), with "poet" being the reference word for synchronicity.

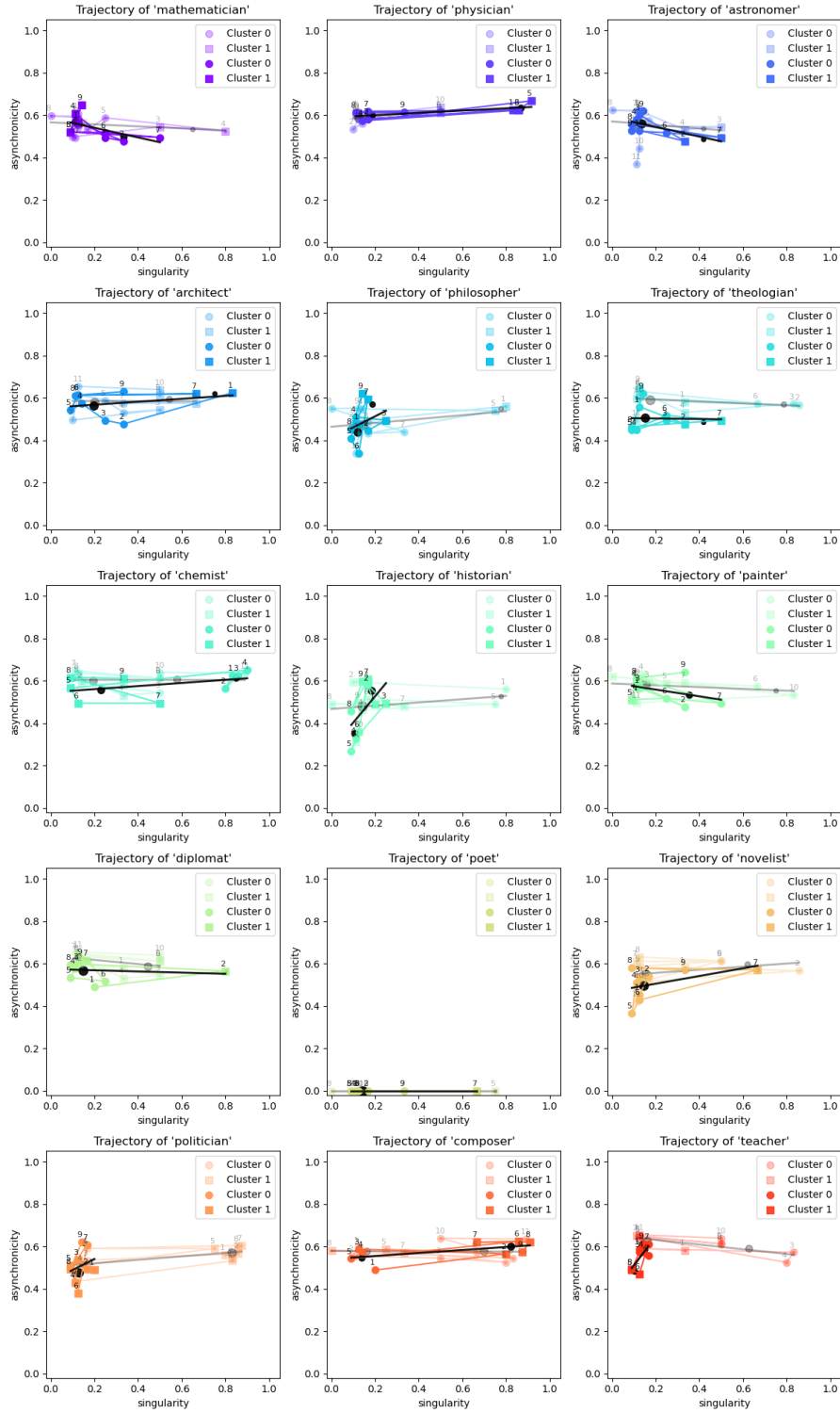


Figure 6.8. The results of using Model II to map the metacognitive dynamics for profession terms across two periods: from 1820 to 1920 and from 1920 to 2020 (respectively shown in light colors and normal colors in the figure), with "poet" being the reference word for synchronicity.

Results show that in the 19th century, an average American English speaker's metacognitive awareness in reference to "poet" was weak and stable for all other professions, which can be characterized by scenario 5.1b ("child"; with near-zero regression coefficients and good alignments between fitted lines and cluster centers). But in the 20th century, a number of profession names underwent categorical shifts of metacognitive dynamics: "historian," "philosopher," "teacher," and "politician," while for other terms, similar types of meta-level or non-meta-level cognitive dynamics are maintained or approximately maintained. These results cannot be accounted for by any obvious regularity of professionalization itself (such as scientific versus non-scientific professions). However, if these terms are mapped in the typicality-contextuality mapping by Model I, it is shown that "historian," "philosopher," "teacher," and "politician" are mapped as four out the five terms that have the closest hyperbolic distances with "poet" compared to the rest of professions (marked in red in Figure 6.9; the fifth one is "diplomat," but it is not shown to undergo major cognitive shifts across the two periods).

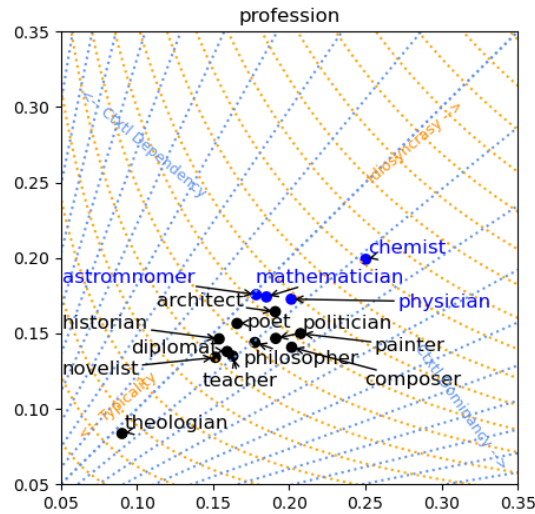


Figure 6.9. Typicality-contextuality mapping of profession names. "Historian," "philosopher," "teacher," and "politician" are mapped as four out the five terms that have the closest hyperbolic distances with "poet" compared to the rest of professions.

Though still preliminary, this result shows an intriguing hypothesis that professionalization as a social process has correlation with the semantic shifts of professions that can be measured in the hierarchical-similarity semantic space, which then can be driven by how an average social actor is metacognitively aware of the identity of each profession, and how stable such metacognitive awareness is in a given sociocultural spacetime. In our case study, terms that are semantically closest to "poet" in the cognition-driven hyperbolic space entail the more turbulent changes of metacognitive awareness. It will be intriguing to further explore how much this is one of multiple cognitive consequences of the process of professionalization, or a cognitive mechanism that contributes to the sociocultural phenomena of professionalization. The methodology and procedures shown in this section demonstrate a promising possibility to

connect sociology of professionalization and social psychology (Abbott, 1988; Lizardo et al. 2020) from the language, cognition, and modeling perspectives.

6.6. Discussion

The intent of this chapter has been to present the utility and potential of the two models in comprehending and interpreting various semantic domains other than color, kinship, and smell. In doing so, I have explored plural suffixes, pragmatic knowledge of place names, and metacognitive sensitivity across several semantic categories with social, political, and religious significance. In this discussion, I consider how these studies can be further related to relevant research areas in cognitive science and other social sciences.

The study on plural suffixes contributes to the broader literature in both cognitive science and linguistics on the heterogeneous distributions of plural markers and shape-driven classifiers across languages (Aikhenvald, 2000). Modeling the plural suffixes for specific basic-level semantic categories of body part, time, and length further propose a computational methodology to examine how semantics across two or multiple domains interact in the same linguistic phenomena and their cognitive consequences (*i.e.*, how language of numbers interact with language of body parts, time, and length, and how this interaction accordingly shapes perception in corresponding domains). In extending this work, future research could delve into computational studies of domain-specific plural markers in other languages and generate new empirical hypotheses for testing how domain-specific plural markers influence speakers'

perceptions and cognitions of body, time, and length, among other domains, in diverse cultural contexts.

The modeling results of the pragmatically-influenced semantic knowledge of nation-state names in American English can be specifically tied to debates between three major approaches in International Relations literature in political science - realism, constructivism, and liberalism (Miller & Saltzman, 2016). Embedding country names in contemporary American English in the typicality- contextuality space provides a computational methodology to map the entanglement and intersecting relationships of all three dimensions from corpus data. In the preliminary results shown in Section 6.4, different nation-state names are together mapped according their domestic and geopolitical power and impacts (the realist dimension), their significance in the collective memories of the United States (the constructivist dimension), and their affinity with the US's liberal value (the liberalist dimension). Future work will explore how the cognitive-computational methodology developed in this section can contribute to the three major contemporary theories in International Relations of political science.

The broad intent of the group of studies in this chapter on political attitudes, religious name and identities, and professions is to contribute to bridging two dichotomies, the first dichotomy is between semantics and pragmatics, which has been noted in various places in the earlier sections in this dissertation, and the second dichotomy is between social theory and cognitive psychology, which is noted by recent literature in cultural sociology and social psychology (Lizardo et al., 2020). In Chapter V, I explicitly or implicitly consider both aspects in the metacognitive modeling studies on color, kinship, and smell semantics, connecting them with

specific literature in color/race and kinship/family. In the third group in this chapter, the metacognitive modeling approach is first applied to the evolution of political attitudes in American English since the 19th century, which can be related to the question of political identity and polarization in today's political science literature (Levitsky & Ziblatt, 2018; Wedeen, 2019). The next study on religion names and religious dynamics with the historical dynamics of inter-religion relations can be related to studies on the history of interreligious dynamics within and across Christianity (Goddard, 2000; Tjørhom, 2020). Lastly, the investigation of professions can be related to the professionalization literature in cultural sociology and professions sociology, contributing to the investigations on how social cognition interacts with professional identities and roles and how such dynamics have evolved over time. All three sets of results in this group of study show the fruitfulness to consider a metacognitive perspective for connecting semantics and pragmatics, and social theory and cognition.

6.7. Chapter summary

Throughout this chapter, I have explored the potential of the two proposed models in shedding light on underrepresented or less emphasized semantic domains. These models are applied to three groups of tasks, demonstrating their utility in generating and testing hypotheses beyond the domains that are examined in previous chapters.

The first exploration using Model I focused on the cognitive mapping of plural suffixes in English, a notable contrast to the classifiers abundant in languages like Chinese and Japanese.

The interaction between plural suffixes and the domains of body parts, time, and length has been illuminating and invites extended research into other languages and their influence on speakers' cognitive and perceptual experiences. The second group of studies applying Model I dealt with pragmatic knowledge of place names, which, in particular, can be tied into the debates within International Relations literature. The third group of studies utilized both Model I and II to examine metacognitive sensitivity across semantic categories of political attitudes, religious names, and professions, which are important to politics, religion, and society. Together, studies in this chapter contribute to bridge the dichotomies between semantics and pragmatics and between social theory and cognitive psychology, from embodied, pragmatic, and metacognitive perspectives.

Chapter VII: Conclusion

This dissertation has set forth a pioneering approach in examining the coevolutionary interaction of cognition and domain-driven semantic change by taking a combined approach that leverages both distributional and dynamical models, filling a gap in the existing literature where these two methodologies have typically been pursued independently. The novel computational models proposed — the "typicality-contextuality" model (Model I) and the "asynchronicity-singularity" model (Model II) — have proven their utility in sets of empirical evaluation within the domains of color, kinship, smell, shape, and in the extended analysis of plural suffixes, pragmatic knowledge of nation state names, and metacognitive awareness of semantic categories regarding political attitudes, religions, and professions.

Model I, as discussed in Chapter II, stands as a novel semantic embedding model that successfully maps basic-level categories onto cognitively interpretable dimensions, namely typicality and contextuality, in hyperbolic space. Overcoming the limitations of traditional contextuality-based methods, this model presents an innovative way to simultaneously embed and interpret multiple cognitive-linguistic dimensions. Its evaluations within the domains of color and kinship semantics in particular demonstrate its effectiveness and promise in future research.

Chapter III saw the application of Model I to long-spanned ancient Chinese data, revealing unique insights into within-domain diachronic competitions of multiple basic-level color terms, as well as the intriguing intersection of smell and color semantics in ancient Chinese

language. Notably, it unearthed the evolving semantic structures of shape-based classifiers over time, offering a fresh perspective on a classical question in Chinese historical linguistics.

While Model I's strength lies in mapping domain-oriented semantic properties synchronically and diachronically, Model II, introduced in Chapter IV, presents a dynamical model focusing on how metacognition interacts with ensemble diachronic changes of basic-level semantic categories. Built on the premise that the metacognition of semantic change is both a response to and a driving force behind diachronic semantic transformations, the model captures the mutually causal dynamics between metacognition and cognitive knowledge of semantic categorization.

Chapter V applied Model II to diachronic corpus data from American English over the last 200 years, covering the domains of color, kinship, and smell. The analysis offered fascinating insights into how metacognitive awareness, sociocultural factors, and basic-level semantic shifts in these domains dynamically interact with one another. In doing so, it highlighted the promising potential of the metacognition-based model to bridge the gap between cognitive science, computational linguistics, and sociolinguistics and linguistic anthropology, in comparison to many traditional dynamics-driven approaches.

In Chapter VI, the adaptability and scope of both models were demonstrated. Model I's applicability was shown in the examination of plural suffixes in English and in the analysis of pragmatic knowledge in the context of place names. Furthermore, both models were employed to assess the metacognitive sensitivity within various semantic domains (political attitudes,

religions, and professions) based on corpus-based evidence of language change, suggesting the potential of these models to be applied across a broad range of semantic and pragmatic domains.

In conclusion, this dissertation contributes a pioneering framework for understanding the interplay between semantic cognition and language change, providing not only novel computational tools but also new insights into the cognitive underpinnings of language evolution as well as the how diachronic patterns in language variation and change shape semantic cognition. The models presented here have shown a robust capability to decipher both distributional and dynamic properties relating semantic cognition to language change. Moreover, they have demonstrated their applicability within and across diverse semantic domains, thereby establishing the potential of domain-oriented focuses to substantiate connections between domain-specific and domain-generic semantic questions in cognitive science. Broadly, this dissertation contributes to the understanding of how semantic knowledge and language change intersect with cognition, culture, and society, adding a crucial dimension to the growing body of knowledge in the interlocking fields of cognitive, computational, and social sciences. The study's implications and the computational models it introduces could serve as a catalyst for future research. They open avenues for more profound explorations into the intertwined landscape of semantic meaning and cognition, language change and variation, and contexts and conditions of culture and society.

References Cited

- Abbott, Andrew (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: The University of Chicago Press.
- Aikhenvald, Alexandra Y. (2000). *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press.
- Anderson, Philip. W. (1972). More Is Different. *Science*, New Series, Vol. 177, 4047. 393-396.
- Antonello, Richard, & Alexander Huth (2022). Predictive Coding or Just Feature Discovery? An Alternative Account of Why Language Models Fit Brain Data. *Neurobiology of Language*, 1-16.
- Aslin, Richard, Jenny Saffran, & Elissa Newport (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science* 9(4): 321–324.
- Barabási, Albert-László (2009). Scale-Free Networks: A Decade and Beyond. *Science* 325, 412-413. DOI:10.1126/science.1173299
- Banks, Briony & Louise Connell (2023). Category production norms for 117 concrete and abstract categories. *Behavior Research Methods* 55, 1292–1313. <https://doi.org/10.3758/s13428-021-01787-z>.
- Berditchevskaia, A., Cazé, R. & Schultz, S. Performance in a GO/NOGO perceptual task reflects a balance between impulsive and instrumental components of behavior (2016). *Scientific Reports* 6, 27389.
- Berlin, Brent, & Paul Kay (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

- Bruno, Aurelio, Jennifer Sudkamp & David Souto (2023). A metacognitive approach to the study of motion-induced duration biases reveals inter-individual differences in forming confidence judgments. *Journal of Vision* 23, 3: 15. doi:10.1167/jov.23.3.15
- Chambers, Deborah & Pablo Gracia (2022). *A Sociology of Family Life: Change and Diversity in Intimate Relations*. Cambridge: Polity Press.
- Chersoni, Emmanuele, Enrico Santus, Chu-Ren Huang, & Alessandro Lenci (2021). Decoding Word Embeddings with Brain-Based Semantic Features. *Computational Linguistics* 47(3): 663–698. https://doi.org/10.1162/coli_a_00412.
- Chen, I-Hsuan, Kathleen Ahrens, & Chu-Ren Huang (2022). The Chinese Classifier System as a Lexical-Semantic System. In Chu-Ren Huang, Yen-Hwei Lin, I-Hsuan Chen, & Yu-Yin Hsu (Eds.) *The Cambridge Handbook of Chinese Linguistics* (pp 550–757). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108329019.029>.
- Chao, Yuen Ren (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chomsky, Noam (1966). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. Cambridge: Cambridge University Press.
- CEOWorld (2021.2.10). Ranked: World's Most Influential Countries : <https://ceoworld.biz/2021/02/10/ranked-worlds-most-influential-countries-2021/>
- Clark, Robin, & Steven O. Kimbrough (2017) Social Structure, Opportunistic Punishment and the Evolution of Honest Signaling. PLoS ONE 12(12): e0188249.
- Cleland, Thomas A, & Ayon Borthakur (2020). A Systematic Framework for Olfactory Bulb Signal Transformations. *Frontiers in computational neuroscience* vol. 14 579143. 23, doi:10.3389/fncom.2020.579143

- Coontz, Stephanie (1993). *The Way We Never Were: American Families And The Nostalgia Trap*. Basic Books.
- De Sa, Christopher, Albert Gu, Christopher Re, Frederic Sala (2018). Representation Tradeoffs for Hyperbolic Embeddings. *Proceedings of the 35th International Conference on Machine Learning*, 80, 4460-4469.
- Detges, Ulrich & Richard WALTEReit (Eds.) (2008). *Paradox of Grammatical Change: Perspectives from Romance*. John Benjamins Publishing Company.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arxiv:1810.04805*
- Dhingra, Bhuwan, Christopher J. Shallue, Mohammad Norouzi, Andrew M. Dai, & George E. Dahl (2018). Embedding Text in Hyperbolic Spaces. *TextGraphs@NAACL-HLT*.
- Dumais, Susan T (2005). Latent Semantic Analysis. *Annual Review of Information Science and Technology*. 38: 188–230. <https://doi.org/10.1002/aris.1440380105>.
- Dunifon, Rachel, & Ashish Bajracharya (2012). The Role of Grandparents in the Lives of Youth. *Journal of family issues* 33, 9: 1168-1194. doi:10.1177/0192513X12444271
- Eckert, Penelope (2000). *Linguistic Variation as Social Practice*. Wiley-Blackwell.
- Eckert, Penelope (2019). The individual in the semiotic landscape. *Glossa: a journal of general linguistics* 4, no. 1: 1–15.
- Emirbayer, Mustafa & Matthew Desmond (2015). *The Racial Order*. The University of Chicago Press.

- Ettinger, Allyson, Naomi H. Feldman, Philip Resnik, & Colin Phillips (2016). Modeling N400 amplitude using vector space models of word representation. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.
- Ettinger, Allyson (2020). What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8, 34–48.
- Fellbaum, Christiane (Ed.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- FitzHugh, Richard (1962). Impulses and physiological states in theoretical models of nerve membranes. *Biophysical Journal*. 1, 445–466.
- Flanders, Judith (2015). *The Victorian City: Everyday Life in Dickens' London*. St. Martin's Griffin.
- Fleming, Stephen M, and Nathaniel D Daw (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological review*. 124,1: 91-114. doi:10.1037/rev0000045
- Frederick, Donald, Scott M. Rojas-Libano, Leslie M. Kay (2011). Rat Behavior in Go/No-Go and Two-alternative Choice Odor Discrimination: Differences and Similarities. *Behavioral Neuroscience* 125. 4:588-603.
- Freeman, Walter J. (2001). *How Brains Make Up Their Minds*. Columbia University Press.
- Freeman, Walter J. (2008). Nonlinear Brain Dynamics and Intention According to Aquinas. *Mind & Matter*. 6(2), 207–234.
- Freeman, Walter J. (2013). Consciousness, Intentionality, and Causality, in Susan Pockett, William P. Banks, and Shaun Gallagher (eds), *Does Consciousness Cause Behavior?* (Cambridge, MA, 2006; online edn, MIT Press Scholarship Online, 22 Aug. 2013).

García, Mario T. (2015). *The Chicano Generation: Testimonios of the Movement*. University of California Press.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky & James Zou (2018). Word Embeddings Quantify 100 Years of Gender & Ethnic Stereotypes. *Proceedings of the National Academy of Sciences*, April, 115 (16) E3635-E3644.

Gell-Mann, Murray & Merritt Ruhlen (2011). The Origin and Evolution of Word Order. *Proceedings of the National Academy of Sciences*. October, 201113716; DOI: 10.1073/pnas.1113716108.

Geertz, Clifford (1973). *The Interpretation of Cultures*. Basic Books.

Ghirardato, Paolo, and Massimo Marinacci (2001). Risk, Ambiguity, and the Separation of Utility and Beliefs. *Mathematics of Operations Research*: 26(4), 864-890.

Ginns, Paul & Jimmie Leppink (2019). Special Issue on Cognitive Load Theory: Editorial. *Educational Psychology Review* 31, 255–259. <https://doi.org/10.1007/s10648-019-09474-4>.

Goddard, Hugh (2000). *A History of Christian-Muslim Relations*. New Amsterdam Books.

Goldstein, Ariel, Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., ... & Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience* 25(3), 369-380.

Goldsmith, John A. & Bernard Laks (2019). *Battle in the Mind Fields*. The University of Chicago Press.

Griffiths, Thomas. L., & Michael. L. Kalish (2007). Language Evolution by Iterated Learning with Bayesian Agents. *Cognitive Science*, 31, 441–480.

Haeberli, Eric (2002). Inflectional morphology and the loss of V2 in English. In Lightfoot, ed., *Syntactic Effects of Morphological Change*. Oxford University Press.

Hamilton, William L., Jure Leskovec, & Dan Jurafsky (2016a). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Hamilton, William L. , Jure Leskovec, & Dan Jurafsky (2016b). Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2116–2121, Austin, Texas. Association for Computational Linguistics.

Hart, Betty, & Risley, Todd R. Risley (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Brookes.

Hawksley, Lucinda (2013). *March, Women, March: Voices of the Women's Movement from the First Feminist to Votes for Women*. André Deutsch.

Hirschfeld, Lawrence (1986). Kinship & Cognition: Genealogy and the Meaning of Kinship Terms. *Current Anthropology* 27(3), 217–242.

Holme, Petter (2019). Rare and everywhere: Perspectives on scale-free networks. *Nature communications* vol. 10, 1 1016. 4, doi:10.1038/s41467-019-09038-8.

Holmes, Selina J. & Andrew W. Ellis (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition* 13:7-8, 884-910.

Hopkins, Nicholas A (2012). The Noun Classifiers of Cuchumatán Mayan Languages: A Case of Diffusion From Otomanguan. *International Journal of American Linguistics*, vol. 78, no. 3, 411– 427.

Hopper, Paul J., and Elizabeth Closs Traugott (2003). *Grammaticalization*. 2nd ed. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press, 2003.
doi:10.1017/CBO9781139165525.

Hovy, Dirk, & Shannon L. Spruit (2016). The Social Impact of Natural Language Processing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers), 591–598, Berlin, Germany. Association for Computational Linguistics.

Irvine, Judith & Susan Gal. *Signs of Differences* (2020). Cambridge University Press.

Jackson, Alice, & Donald Bolger (2014). Using a high-dimensional graph of semantic space to model relationships among words. *Frontier in Psychology* 5:385.
<https://doi.org/10.3389/fpsyg.2014.00385>.

Kahneman, Daniel, and Amos Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2: 263–91.

Kay, Leslie, M (2014). Olfactory system oscillations across phyla. *Current Opinion in Neurobiology*. 31: 141-147.

Kardan, Omid, Kirsten C.S. Adam, Irida Mance, Nathan W. Churchill, Edward K. Vogel, Marc G. Berman (2020). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *NeuroImage*, 211, 116622.

Kelsey, David, and Sara Le Roux (2015). An experimental study on the effect of ambiguity in a coordination game. *Theory and Decision* 79 (4): 667–688.

Ji, Eugene Yu. Working paper (a). A metacognitive framework for modeling state-led reform decision-making via dynamical systems.

Ji, Eugene Yu. Working paper (b). Memorials, chronotopes, and maximal quantifiers: a semiotic approach to public and hidden transcripts in the endgame of Chinese zero-COVID lockdown.

Ji, Eugene Yu. In review. How does the state imagine? - Two socio-semiotic models of state governance in Hong Kong in the 2000s.

Jiang, Hao, Lixia Li, Yuanyuan Zeng, Jiajun Fan, & Lijuan Shen (2022). Low-Complexity Hyperbolic Embedding Schemes for Temporal Complex Networks. *Sensors* 22(23): 9306. <https://doi.org/10.3390/s22239306>.

Kemp, Charles & Terry Regier (2012). Kinship categories across languages reflect general communicative principles. *Science* 336(6084): 1049-54. doi:10.1126/science.1218811

Kovács, Bianka., Balogh, Sámuel G., & Palla, Gergely (2022). Generalized popularity-similarity optimisation model for growing hyperbolic networks beyond two dimensions. *Nature Scientific Reports*, 12, 968. <https://doi.org/10.1038/s41598-021-04379-1>

Krioukov, Dmitri, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, & Marián Boguñá (2010). Hyperbolic geometry of complex networks. *Physics Review*. E 82, 036106

Kumar, Sreejan, Theodore R. Sumers, ... & Samuel A. Nastase (2022). Reconstructing the cascade of language processing in the brain using the internal computations of a transformer-based language model. *BioRxiv*, 2022-06.

Levitsky, Steven & Daniel Ziblatt (2018). *How Democracies Die?* Crown.

Lizardo, Omar, Brandon Sepulvado, Dustin S. Stoltz, Marshall A. Taylor (2020). What can cognitive neuroscience do for cultural sociology? *American Journal of Cultural Sociology* 8, 3–28.

Liu, Ninghao, Qiaoyu Tan, Yuening Li, Hongxia Yang, Jingren Zhou, & Xia Hu (2019). Is a Single Vector Enough? Exploring Node Polysemy for Network Embedding. *In*

Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). New York, NY, USA, 932–940.
DOI:<https://doi.org/10.1145/3292500.3330967>

Lucy, John (1997). The Linguistics of “Color”. C. L. Hardin & L. Maffi (Eds.), *Color Categories in Thought and Language*. Cambridge: Cambridge University Press, 320–346.

Lucy, John. A., & Susanne Gaskins (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman and S. C. Levinson (Eds.), *Language acquisition and conceptual development: 257–283*. Cambridge: Cambridge University Press.

Majid, Asifa, and Niclas Burenhult (2014). Odors are expressible in language, as long as you speak the right language.” *Cognition*. 130,2: 266-70.
doi:10.1016/j.cognition.2013.11.004.

Majid, Asifa, Burenhult Niclas, Stensmyr Marcus, de Valk Josje and Hansson Bill S (2018). Olfactory language and abstraction across cultures. *Phil. Trans. R. Soc. B*. 373: 20170139.

Majid, Asifa (2021). Human olfaction at the intersection of language, culture, and biology. *Trends in Cognitive Sciences*, 25(2), 111–123. doi.org/10.1016/j.tics.2020.11.005

Meyer, Madonna Harrington (2014). *Grandmothers at Work: Juggling Families and Jobs*. New York: NYU Press.

Mikolov, Tomas, Kai Chen, Greg Corrado, & Jeffrey Dean (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, Benjamin, Saltzman, Ilai (2016). Beyond the three ‘isms’: Rethinking IR and the post-cold war order. *International Politics* 53, 385–414.

- Miller, Vincent P., & John M. Quigley (1990). Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area. *Urban Studies* 27, 1: 3–21.
<http://www.jstor.org/stable/43084068>.
- Misra, Kanishka, Allyson Ettinger & Julia Taylor Rayz (2021). Do language models learn typicality judgments from text?. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.
- Nagumo, Jin-ichi, Suguru Arimoto, and Shuji Yoshizawa (1962). An Active Pulse Transmission Line Simulating Nerve Axon. *Proc. IRE* 50: 2061–2070.
- Ngai, Mae M (2014). *Impossible Subjects: Illegal Aliens and the Making of Modern America*. Princeton University Press.
- Nguyen, Dong, Laura Rosseel, & Jack Grieve (2021). On learning and representing social meaning in NLP: A sociolinguistic perspective. In *Proceedings of NAACL*.
- Nickel, Maximillian, & Douwe Kiela (2017). Poincare' embeddings for learning hierarchical representations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett, editors, *Advances in Neural Information Processing Systems* 30, 6338–6347. Curran Associates, Inc.
- Niyogi, Partha. (2006). *The Computational Nature of Language Learning and Evolution*. MA: MIT Press.
- Nowak, M.A. & Krakauer, D.C. (1999). The Evolution of Language. *PNAS*, July, 96 (14), 8028-8033.
- Olofsson, Jonas & Jay Gottfried (2015). The Muted Sense: Neurocognitive Limitations of Olfactory Language. *Trends Cognitive Science*, 19, 314–321.

- Ortiz, Vilma & Edward Telles (2012). "Racial Identity and Racial Treatment of Mexican Americans." *Race and social problems* 4(1): 10.1007/s12552-012-9064-8. doi:10.1007/s12552-012-9064-8
- Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. (2022). Training language models to follow instructions with human feedback. *arXiv preprint*. arXiv:2203.02155.
- Owen, Stephen (1986). *Remembrances: The Experience of the Past in Classical Chinese Literature*. Cambridge, MA: Harvard University Press.
- Papadopoulos, Fragkiskos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá & Dmitri Krioukov (2012). Popularity versus similarity in growing networks. *Nature* 489, 537–540. <https://doi.org/10.1038/nature11459>
- Peterson, Joshua, Dawn Chen & Thomas L. Griffiths (2020). Parallelograms revisited: Exploring the limitations of vector space models for simple analogies. *Cognition*, 205, 104440.
- Petruck, Miriam R. L. & Valia Kordoni (2015). Robust Semantic Analysis of Multiword Expressions with FrameNet. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts, Lisbon, Portugal*. Association for Computational Linguistics.
- Plant, Christopher, Janet Webster & Anne Whitworth (2011). Category norm data and relationships with lexical frequency and typicality within verb semantic categories. *Behavior Research Methods*, 43, 424–440.
- Poeppel, David, & Philip J. Monahan (2008). Speech Perception: Cognitive Foundations and Cortical Implementation. *Current Directions in Psychological Science*, 17(2), 80–85.
- Prigogine, Ilya. (2000). *The Networked Society*. *Journal of World-Systems Research*. November, 892-898.

- Regier, Terry, & Yang Xu (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8:e1440.
- Regier, Terry, Charles Kemp, & Paul Kay (2015). Word Meanings across Languages Support Efficient Communication. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence*. Hoboken, NJ: Wiley-Blackwell, 237-263.
- Rosch, Eleanor (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233.
- Rosch, Eleanor, Carol Simpson & Scott Miller (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance* 2(4), 491–502.
- Ruggles, Steven (1994). The Transformation of American Family Structure. *The American Historical Review* 99, : 103–28. <https://doi.org/10.2307/2166164>.
- Saffran, Jenny (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*. 12 (4): 110–114.
- Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake, & Dan Flickinger (2002). Multiword Expressions: A Pain in the Neck for NLP. In *Computational Linguistics and Intelligent Text Processing*, 1–15. Berlin, Heidelberg: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/3-540-45715-1_1.
- Said, Edward W (1979). *Orientalism*. Vintage.
- Sajjad, Hassan, Nadir Durrani, & Fahim Dalvi (2022). Neuron-level Interpretation of Deep NLP Models: A Survey. *Transactions of the Association for Computational Linguistics* 2022, 10: 1285–1303. doi: https://doi.org/10.1162/tacl_a_00519

- Schneidman, Laura, Michelle E. Arroyo, Susan C. Levine & Susan Goldin-Meadow (2013). What counts as effective input for word learning? *Journal of Child Language*, 40, 672–686.
- Silverstein, Michael (1976). “Shifters, linguistic categories, and cultural description.” In K. Basso & H. Selby, eds., *Meaning in Anthropology*, 11–55. Albuquerque: University of New Mexico Press.
- Silverstein, Michael (1993). Metapragmatic Discourse and Metapragmatic Function. *Reflexive Language: Reported Speech and Metapragmatics*, edited by John A. Lucy, 33–58. Cambridge: Cambridge University Press.
- Silverstein, Michael (2004). “Cultural” concepts and the language-culture nexus. *Current Anthropology* 45(5):621–52.
- Silverstein, Micheal (2022). *Language in Culture: Lectures on the Social Semiotics of Language*. Cambridge University Press.
- Sun, Vincent C. & Chien-Chung Chen (2018). Basic color categories and Mandarin Chinese color terms. *PLoS ONE* 13(11): e0206699. <https://doi.org/10.1371/journal.pone.0206699>
- Steyvers, Mark, & Joshua Tenenbaum (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41–78. doi.org/10.1207/s15516709cog2901_3.
- Strogatz, Steven (2000). *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*. CRC Press.
- Tanweer, Anissa , Emily Kalah Gade, PM Krafft, & Sarah K Dreier (2021). Why the data revolution needs qualitative thinking. *Harvard Data Science Review*.
- Taylor, Ann (1999). The change from SOV to SVO in Ancient Greek. *Language Variation and Change*, 6: 1-37.

- Tamkin, Alex, Miles Brundage, Jack Clark, Deep Ganguli (2021). Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. arXiv:2102.02503
- Tenenbaum, Joshua. et al. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, 331(6022), 1279–85.
- Tikochinski, R., Goldstein, A., Yeshurun, Y., Hasson, U., & Reichart, R (2021). Fine-tuning of deep language models as a computational framework of modeling listeners' perspective during language comprehension. *bioRxiv*, 2021-11.
- Tjørhom, Ola. (2020). Fifty Years of International Catholic–Lutheran Dialogue: Much Consensus, Little Fellowship? *Theological Studies*, 81(1), 65–87.
- Tong, Zhihe (1984). 'Xiang' he 'chou' de ciyi yanbian ji qiweici de ciyi xitong de fazhan (The Semantic Developments of 'Xiang' and 'Chou' and the Smell Word System). *Journal of Hangzhou University*, June, 13(2), 99 - 107.
- Trousdale, Graeme (2008). A constructional approach to lexicalization processes in the history of English: Evidence from possessive constructions, *Word Structure*, 1: 2, 156-177.
- Urban, Greg (2006). Metasemiosis and metapragmatics. In K. Brown (Ed.), *Encyclopedia of language and linguistics*. 2nd ed., 8, 88– 91. Amsterdam: Elsevier.
- Uyeda, Katherine, & George Mandler (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods & Instrumentation*, 12(6), 587–595.
- Vincze, Orsolya (2015). Learning multiword expressions from corpora and dictionaries. *Doctoral Thesis*.
- Volo, James M. & Dorothy Volo (2007). *Life in 19th-Century America*. Greenwood.

- Wagner, Claudia, David Garcia, Mohsen Jadidi, & Markus Strohmaier (2015). It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia. *In Proceedings of the AAAI conference on web and social media*.
- Wallerstein, Immanuel. (1998). The Heritage of Sociology, The Promise of Social Sciences, Presidential Address. *XIVth World Congress of Sociology*. Montreal, July 26.
- Wedeen, Lisa (2019), *Authoritarian Apprehensions: Ideology, Judgement, and Mourning in Syria*. The University of Chicago Press.
- Wittgenstein, Ludwig. *Philosophical Investigations*, 1953, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.
- Wu, Frank H (2003). *Yellow: Race in America Beyond Black and White*. Basic Books.
- Wu, Jianshe (2012). Hanyu jiben yanseci de jinghua jieduan yu yanse fanchou (The Evolutionary Stages and Perceptual Categories of the Basic Color Terms in Chinese). *Guhanyu yanjiu (Studies on Ancient Chinese)*, 1, 10-17.
- Wu, Xiu-Zhi Zoe. (2004). *Grammaticalization and Language Change in Chinese: A formal view* (1st ed.). Routledge. <https://doi.org/10.4324/9780203420645>
- Wu, Qinxuan & Allyson Ettinger (2021). Variation and generality in encoding of syntactic anomaly information in sentence embeddings. *Proceedings of the Fourth Blackbox NLP Workshop on Analyzing and Interpreting Neural Networks for NLP*.
- Xie, Bin, & Salvendy, Gavriel (2000). Prediction of mental workload in single and multiple task environments. *International Journal of Cognitive Ergonomics*, 4, 213–242.
- Yao, Xiaopin (1988). Jiben yanseci lilun shupin -- jianlun hanyu jiben yanseci de yanbianshi (Review on the Basic Color Term Theory: Comments on the Evolutionary History of the

Basic Color Terms in Chinese). *Waiyu jiaoxue yu yanjiu (Pedagogy and Research on the Foreign Language)*, 1, 19-28.

Yang, Charles (2002). Grammar competition and language change. In Lightfoot, ed., *Syntactic Effects of Morphological Change*. Oxford University Press.

Yeung, Nick, and Christopher Summerfield (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* vol. 367, 1594: 1310-21.

Zaslavsky, Noga, Terry Regier, Naftali Tishby, & Charles Kemp (2019). Semantic categories of artifacts and animals reflect efficient coding. *In Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Zhao, Xiaochi (2016). *Shanggu - zhonggu hanyu yanseji yanjiu (The study of Chinese Color Terms from Ancient to Mid-Ancient Times)*. Beijing: Chinese Social Science Publisher.

Appendix: Supplementary data of typicality scoring in Table 2.2 in Chapter II

1. Typicality scoring of basic color terms in English predicted by the typicality-contextuality model and compared with empirical data from the MTurk studies in this paper and other studies in the literature. All data are weighted to the range of [0,1], where the score for the word that is ranked as the most typical among all the words equals to 1.

	red	yellow	green	blue	orange	brown	gray	purple	pink
Typicality-contextuality model (this paper)*	1.000	0.967	0.965	0.965	0.284	0.703	0.326	0.471	0.195
Weighted production frequency with words limit (Mturk study of this paper)	0.923	0.931	0.923	1.000	0.603	0.206	0.137	0.389	0.405
Weighted production frequency with time limit (Banks & Connell, 2022)	1.000	0.414	0.357	0.75	0.375	0.080	0.109	0.214	0.264
Weighted typicality rating (Banks & Connell, 2022)	1.000	0.949	0.931	1.000	0.677	0.965	0.913	0.931	1.000
Weighted typicality rating (Uyeda & Mandler 1980)	1.000	0.96	0.879	1.000	0.771	0.472	0.388	0.674	0.514

Table A.1.

2. Typicality scoring of selected basic kinship terms in English predicted by the typicality-contextuality model and compared with empirical data from the MTurk studies in this paper and other studies in the literature. All data are weighted to the range of [0,1], where the score for the word that is ranked most typical among all the words equals to 1 (except for pairwise first production frequency, for which raw production frequencies are shown).

	father	mother	son	daughter	brother	sister	grand father	grand mother
Typicality-contextuality model (this paper)*	0.983	0.980	1.000	0.923	0.934	0.844	0.499	0.498
Weighted production frequency (within a words limit; Mturk study of this paper)	0.976	0.915	0.329	0.280	1.000	0.927	0.329	0.280
Pairwise first production frequency (within a words limit; Mturk study of this paper)	33	59	17	7	58	23	10	22
Weighted production frequency (within a time limit; Banks & Connell, 2022)	0.721	1.000	0.136	0.149	0.623	0.626	0.069	0.062
Weighted typicality rating (Banks & Connell, 2022)	0.982	0.966	1.000	1.000	1.000	0.984	0.95	1.000

Table. A.2.