

自然言語処理を用いた
言語類型論と理論言語学の研究

Matsuura Yoshihiro

April 24, 2024

Contents

I	Introduction	1
1	Abstract	2
II	言語学の歴史	3
2	言語学史の概観	5
3	個別言語のみの時代	6
4	言語の比較研究の時代	7
5	一般言語学の時代	8
6	理論言語学の時代	9
III	言語学の諸課題	10
IV	本稿での定義	12
7	何を定義するか	13
8		14

<i>CONTENTS</i>	2
V 機械学習	15
9 機械学習の種類	17
VI 機械学習を用いた言語類型論研究	18
10 主題	19
11 手法	20
VII 機械学習を用いた理論言語学研究	21
VIII 参考文献	22
IX 謝辞・後書き	23
12 謝辞	24

Part I

Introduction

Chapter 1

Abstract

本稿では、機械学習の技術を用いて、言語学一般の問題にこれまでとは異なる視点からの切り口を提示する。まず、言語学の歴史を概観し、それを踏まえて言語学の諸課題を整理する。その後、本稿での定義を提示し、機械学習の概念を体系的に説明する。最後に、機械学習を用いた言語学研究が奏功する例として言語類型論と理論言語学を挙げ、筆者が卒業研究で行った具体的な研究成果の詳細について述べる。

Chapter 2

自然言語処理を用いる必然性

Part II

言語学の歴史

本章では、これまでの様々な文献を参考に、言語学の歴史を概観する。

Chapter 3

言語学史の概観

言語学の歴史は、以下のように概観される。

- 個別言語のみの時代
- 言語の比較研究の時代
- 一般言語学の時代
- 理論言語学の時代

Chapter 4

個別言語のみの時代

古代ギリシャでは、ギリシャ語文法のサンスクリット語に関しては、パーニニ

Chapter 5

言語の比較研究の時代

Chapter 6

一般言語学の時代

フェルディナン・ド・ソシュールの「一般言語学講義」

Chapter 7

理論言語学の時代

ノーム・チョムスキーの生成文法、ラネカーの認知文法

Part III

言語学の諸課題

言語学の課題は以下に集約される。

- そもそも定義する必要のない言語的単位を設定していること。
- 形式的・定量的な定義がない上に、論文や研究者の間で定義が揺れていることがしばしばあること。
- 研究対象が実際の言語活動の産物だけではないこと。

Part IV

本稿での定義

Chapter 8

何を定義するか

本稿では、ウィトゲンシュタインの『論考』に於ける用語を用いて言語学の問題・用語を包括的に定義することを試みる。

Chapter 9

Part V

機械学習

本章では、本稿で扱う機械学習の概念を体系的に説明する。そもそも機械学習とは、

Chapter 10

機械学習の種類

機械学習には以下のような種類がある。

- 教師あり学習
- 教師なし学習
- 強化学習

Part VI

機械学習を用いた言語類型論研究

Chapter 11

主題

ロマンス諸語とセム諸語を対象に、機械学習を用いて言語類型論的な研究を行った。語彙一致率同語源の語彙の意味の違い（空似具合）文法の違いの比較

Chapter 12

手法

Part VII

機械学習を用いた理論言語学研究

Part VIII

参考文献

Part IX

謝辞・後書き

Chapter 13

謝辞

本論文の作成にあたり、多くの方々にご指導ご鞭撻を賜った。
指導教官の長谷川先生には、突拍子もない研究テーマやアイデアも真剣に受けて止めて頂き、具体的なアクションプランを提示して頂いた。それによって後半の「機械学習を用いたロマンス諸語の定量的分布の調査」など本稿の最重要な研究を3年生の頃から推し進めることができた。また、日常的に研究室に訪問させて頂き、研究・進路の相談だけでなく、鉄道や他言語など趣味の話にも乗って頂き、2年間常に高いモチベーションを維持することができた。この場を借りて、心より感謝申し上げる。
京都大学の河崎靖先生には、1年生の頃から定期的に研究室に訪問させて頂き、過大評価とも言える程私に期待を寄せて頂いた。多くの書籍を貸して頂き、それらは本稿の私の言語学史・言語学全般の豊かな体系的知識となり、本稿の「言語学小史」編を充実させる基盤となった。また時には貴重な本を譲渡までして頂き、一生の宝物となった。ゲルマン語学のみならず、言語学一般において目覚ましい実績を残された河崎先生に多くの指導を仰げたことは、私のモチベーションを最高の状態に維持しただけでなく、これからの研究人生における大きな自信と財産になった。
大阪大学人文研究科博士課程でデジタルヒューマニティーズを専攻されている于拙氏には、機械学習・自然言語処理に関する非常に有益な知見を提供して頂いただけでなく、豊富な個別言語における事例を提示して頂き、
アラビア語専攻の依田先生には、2年の頃からアラビア語・ヘブライ語の授業を受講させていただき、日本では絶対に入手できない貴重な書籍を貸して頂いたり譲渡して頂いたりし、感謝の念が絶えない。また、本稿の主要部を成すSEM語関連の研究成果を多数紹介して頂き、本稿の完成に大きく貢献して頂いた。

同専攻に同期入学した藤井翔氏には、本研究が始まった当初から、幾度にも渡って非常に鋭く的確な指摘をして頂いた。常に客観的な目線を与えてくれ自分の研究の原点を思い出すきっかけを与えてくれた。その他プライベートでも仲良くして頂き、研究以外の面でも大変お世話になった。最後になったが、22年間様々な面でサポートして頂いた両親・祖母・弟、そして、高校の頃から交際しているパートナーに心の底から感謝の念を表したい。