社会的状況に基づいた日本語ビジネスメールコーパスの構築

Muxuan Liu[1]　石垣達也[2]　上原由衣[2]　宮尾祐介[3,2]　高村大也[2]　小林一郎[1,2]
[1] お茶の水女子大学大学院 [2] 産業技術総合研究所 [3] 東京大学
{liu.muxuan,koba}@is.ocha.ac.jp　yusuke@is.s.u-tokyo.ac.jp
{ishigaki.tatsuya, yui.uehara,takamura.hiroya}@aist.go.jp

概要

日本語の使用は，話者間の社会的地位の差異や親疎など多くの社会的状況に影響されるため，日本語のテキストを処理するモデルを構築する際に，これらの社会的状況を考慮する必要がある．本稿では，言語を社会記号体系として捉える選択体系機能言語学における社会的状況の記述方法を用いて，機械学習モデルの訓練のための社会的状況に関する情報を含むコーパスの構築を試みる．

1　はじめに

社会的状況には，聞き手の年齢・性別、話し手と聞き手の上下関係・親疎，談話の目的・内容，流れ，話の進行，談話状況の違い（電話か対面か），話しの様子（雑談的か討論的かという話のタイプ、好意的・対立的、気楽・緊張感），地域差などが含まれている [1]．日本語における言語使用は，これらの社会的状況を強く反映しているとされる [2]．

そのため，様々な自然言語処理タスクにおいて日本語テキストを処理する際，より正確なモデルを構築するためには，テキストを取り巻く社会的状況を考慮することが必要になる．例えば，メールを修正対象とする日本語誤り訂正タスクを考えた場合，文法的な間違いを訂正対象とするだけでなく，上司に休暇を申請する状況においては「明日は休みます．ありがとう．」と書いてしまうなど，社会的状況を踏まえていないために不適切な表現を使ってしまう誤りも存在する [3]．

機械学習モデルが個々の社会的状況を適切に捉え，利用するためには，それら社会的状況に関する属性情報を含むコーパスが必要である．そこで，本稿では，送信者と受信者の社会的状況に基づいた書面でのコミュニケーションと位置づけられる電子メール [4] というジャンルを選択し，社会集団における言語使用の観点から言語分析を行う選択体系機能言語学に基づき，より詳細な社会的状況に関する分析情報を含む日本語コーパスの構築を試みる．

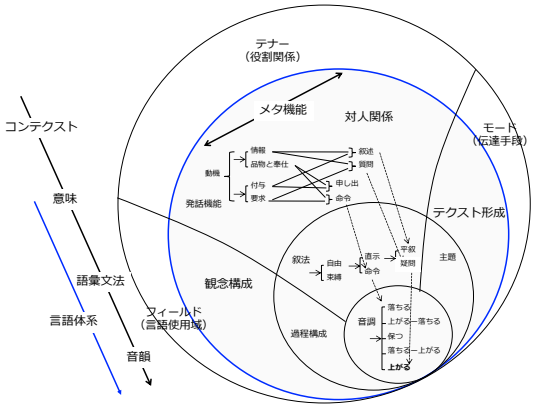

図 1: SFL による言語体系（図は [5] から引用）

2　社会的状況の表現

2.1　選択体系機能言語学 (SFL)

選択体系機能言語学（Systemic Functional Linguistics, SFL）は，Halliday によって確立された機能言語学の一分野であり，とくに言語体系を社会記号体系とみなしている．言語体系は，意味層、語彙・文法層、表現層という 3 つの異なる種類の記号体系が階層になっており，それらの記号体系はコンテクストによって包括されている．その体系は，社会状況下での言語使用を表すための選択肢からなるネットワーク（‘選択体系網’ と呼ぶ, 2.2 節で詳述）で記述されており，状況の選択が意味の選択を制約し，意味が語彙・文法の選択を制約するなど，選択の体系が，意味，語彙・文法，表現と異なる記号体系が連携して形作られている．SFL による言語体系の概観を図 1 に示す．

Halliday は，対話が発生する状況のコンテクスト（Context of Situation）を，「何が起きているか（フィールド・活動領域）」「誰が参加しているか（テナー・

役割関係)」「言語が使用される手段（モード・伝達様式)」という 3 つの枠組みで説明している [6].

本稿で取り上げた「電子メール」というジャンルにおけるコミュニケーションを想定した場合，フィールドは「電子メールによるコミュニケーション」という社会活動となり，コミュニケーションの内容は，様々な言語使用域として捉えられる．テナーは電子メールをやり取りをする関与者らの社会的役割となる．モードはテキストの形態を規定する伝達様式となるため「電子文書」となる．

2.2 選択体系網と選択肢

SFL の特徴の一つは，具現化される言語資源を選択肢（「特徴（feature)」と呼ばれる）の体系として表現することである．これらの体系は「選択体系網（system network)」と呼ばれ，言語体系を包括するコンテクスト層，言語体系を構成する意味層，語彙・文法層，表現層の各層の異なる記号体系の言語資源が選択体系網で表現され，コンテクスト層から順番に上位の層から選択された言語資源に基づき下位の資源が選択されるという選択の体系を構成している．例えば，「医療現場」という状況だと，「診査」「治療」のような出来事が存在し，それに対して「手術」「この薬を飲んでしばらく様子を見て」など語彙や文法が選択される．このように，選択体系網はテキストが具現される過程を表しており，テキストの具現過程を構成する各資源（特徴）が，どのような選択の関係にあるかを表す．また，「選択」においては，各特徴のどれか一つを選択する場合，‘[’を，同時に複数の特徴を選択する場合，‘{’ のそれぞれの表記を使用して選択体系網を記述する．

3 SFL に基づいたコーパスの構築

本研究では，上述した SFL によって捉えられる社会的状況を考慮した電子メール（とくに，ビジネスメール）のコーパスを構築する．構築の流れを以下に示す．

1. **選択体系網の構築**　SFL に基づきメールを対象とした社会的状況の選択体系網を構築する．
2. **場面の設定と収集**　クラウドソーシングを用いて，1 において構築した選択体系網の選択肢を反映する多様な場面を収集する．これにより，様々な状況を設定する．この際，選択体系網から選択肢を選んで場面を設定する作業は，3 において収集するメールに社会的状況をアノテーションすることに相当する．
3. **メールの収集**　2 で収集した場面を使い，クラウドソーシングでメール本文を収集する．
4. **SFL に基づくアノテーション**　3 で収集したメールに対して，SFL に基づくアノテーションを行う．

以下，それぞれについて詳細を説明する．

3.1 選択体系網の構築

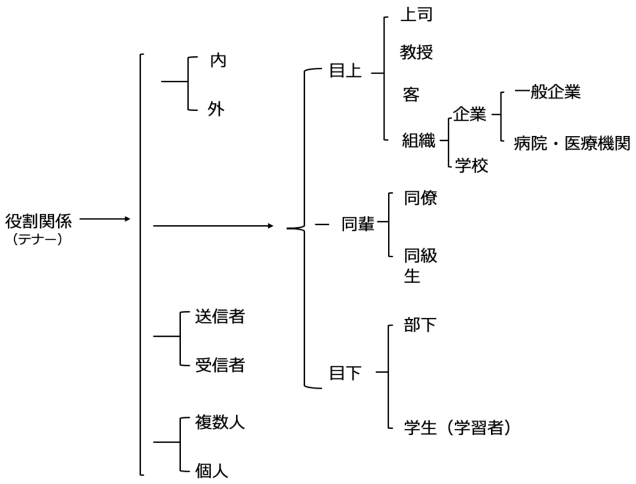

図 2: 「テナー（役割関係)」の選択体系網

テナー（役割関係）　テナー（役割関係）とは，言語表現のやり取りにおける話し手と聞き手，あるいはメールの送信者と受信者の関係である．テナーには，一般的なビジネスメールに見られる対話参加者の社会的立場を想定し，図 2 に示すような選択体系網を構築した．「内」と「外」は，対話参加者の所属に関しての内外の立場関係を表す属性である．一般的に，「内」は，「家族や自分の会社の人，自分が属するグループなど」を意味し，「外」は，「親しくない人や他人，他会社の人，他グループの人など」と説明されている [7]．また，話者の立場を表すため，ビジネスメールによく登場する人物と組織を，目上・同輩・目下の 3 つの属性に分けている．

発話機能　SFL では，テナー（役割関係）が，意味層における「発話機能」に影響を与える．

発話機能について，照屋 [8] は，SFL を用いて人間関係と対人的な意味を分析し，発話機能における対人的役割をまとめている（付録 3 参照)．

本稿はその対人的関係と発話機能を参照し，ビジネスメールにおいてよくある発話動機に基づき，図 3 の選択体系網を構築した．

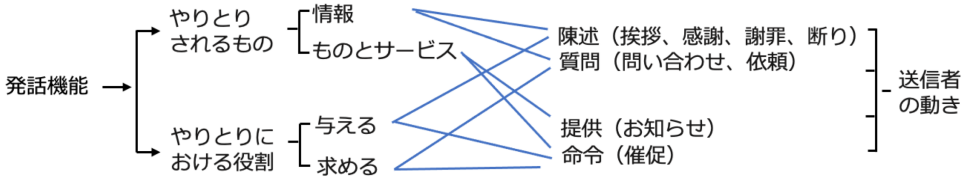

図 3: 「発話機能」の選択体系網

3.2 場面の設定と収集

ビジネスメールによるコミュニケーションという社会的状況を表すために，前節で示した選択体系網の選択肢を元に，コーパスの属性を設定する．

「ビジネスメールによるコミュニケーション」というフィールドの下，多用な言語使用域のコーパスを収集するために，メールでのコミュニケーション場面もクラウドソーシングで作成依頼をかけ，場面の収集を行なった．具体的には，図 2 で示されている役割関係を元に，ビジネスメールによくある送信者と受信者の関係を 20 ペア設定した（付録：表 4 参照）．また，図 3 で示されている「送信者の動き」について，ビジネスメールによくある送信者の目的を「挨拶」「感謝」「謝罪」「断り」「問い合わせ」「依頼」「お知らせ」「催促」の 8 種類で設定している．

クラウドソーシングで 52 名の作業者を集め，1 人あたり約 20 場面を作成してもらい，合計 1040 種類の場面を集めた．また，クラウドソーシングのデータの質を高めるために，発注時に付録の表 5 で示されているような例を大量に提示した．また，今後のデータ分析等で受信者の特定を容易にするため，発注の際には受信者の称呼を「A さん・A 様・A 社長など」と表記することを求めた．

3.3 メールの収集

前ステップで得られた場面の中から，有効な 770 の場面を選び，クラウドソーシングで各場面毎に 5 通のメールを取得した．データの質を担保するために，発注時に，付録の表 6 で示されているような例を提示し，「上司には敬語を使う」、「友人にはため口でも大丈夫」など、常識的な範囲で、対人関係や社会的上下関係を考慮して作成することを求めた．また，形態素解析などで利用しやすいように，件名と宛名の記入，また，「XX 大学 XX 学部の XX さんは私の友達です」のように，場所や自分の名前などの固有名詞を「XX」で記入することを求めた．

3.4 SFL に基づくアノテーション

表 1 に，コーパスの全体像の例を示す．機械学習での活用を容易にするため，3.1 節で列挙した選択体系網の選択肢の名称をコーパスのアノテーション名としてそのまま踏襲するのでなく，一定の構造を簡略化している．

表 1 に例示したメール本文にあるように，これは従業員から顧客へのお知らせのメールである．図 2 で示されている対話参加者の選択体系網の「目上」「同輩」「目下」という上下関係を，「目下（送信者）」と「目上（受信者）」というアノテーションで表現している．また，送信者と受信者のそれぞれの具体的な身分および話者間の所属に関する内と外の関係も「内外関係」で表現する．

発話機能を表すアノテーションに関しては，図 3 で示されている選択体系体系網を元に設定している．今回，送信者目線でメール本文を収集したため，アノテーションを設定する際に「受信者の動き」を除外した．「送信者の動き」は「陳述・質問・提供・命令」の 4 項目があり，それぞれの詳細を「送信者の動き（詳細）」というアノテーションで表現する．「やりとりにおける役割」に関して，送信者がやりとりしたいものやことを「与える」か「求める」かで選択され，対話参加者によってやりとりされるのは，「情報」か「モノとサービス」のいずれかである [8]．例示のメール本文の場合，従業員が情報を顧客に与えている．「モノとサービス」のやりとりの場合，例えば，「本を返して」という要求で開始されるやりとりは，本を送信者に返せば目的を達成されるため，言語の果たす役割において「情報」のやりとりと「モノとサービス」とでは異なっている [8]．

4 コーパスの分析

コーパスの場面数やメール数などの統計量を表 2 に示す．延べ語数および異なり語数の計算について，国立国語研究所の形態素解析ツール「Web 茶ま

表 1: コーパスの概要：従業員が顧客に対してお知らせする場面の例

場面	本文	対話参加者							発話機能			
		目上(受信者)	目下(送信者)	送信者身分	受信者身分	内外関係	送信者数	受信者数	送信者の動き	送信者の動き(詳細)	やりとりにおける役割	やりとりされるもの
顧客 A 様からご依頼頂きました商品が入荷しました。あなたは A 様に商品の入荷をお知らせするメールを書いて下さい。	件名: 商品入荷のお知らせ A 様 日頃より弊社の製品をご愛顧くださり、まことにありがとうございます。先月ご注文いただきました商品ですが、本日入荷いたしましたので、ご都合のよろしいときにご来店いただければと存じ上げます。 ご自宅へのご配送をご希望される場合は、お手数でございますが、下記の配送センターまでご連絡くださいますようお願い申し上げます。 配送センター　担当 XX	上	下	従業員	顧客	外	個人	個人	質問	問い合わせ	与える	情報

表 2: コーパスの特徴を示す統計量

送信者の動き	場面数	場面数（割合）	メール数	平均文長	延べ語数	異なり語数
断り	70	0.09	350	16.69	23406	1086
依頼	100	0.13	500	17.60	35961	1561
謝罪	100	0.13	500	17.14	42326	1576
催促	100	0.13	500	20.57	42758	1130
感謝	100	0.13	500	15.82	37232	1499
挨拶	100	0.13	500	15.62	38370	1641
お知らせ	100	0.13	500	18.31	44822	1903
問い合わせ	100	0.13	500	19.07	40734	1614
合計	770	1	3850	17.67	302521	3869

め」[1)]を用いて，メール本文のテキストを解析し，記号を含めて計算を行った．また，送信者の動きにある「断り」について，「依頼主体」と「断り主体」の人間関係の調査 [9] によると，一般的に「断る」という行為は個人の送信者と個人の受信者の間で起こるものだと考えられる．1 人の送信者が複数人の受信者全体を断ること（例えば、ある学生がサークルのメンバー全員を断ること）は稀であるため，コーパス構築の際には，そのような一対多の対人関係のペアを除外した．そのため，他の項目と比べ，「断り」の場面は 70 個しかない．

5　おわりに

本研究は，選択体系機能言語学に基づき，ビジネスメールに具現される社会的役割関係の情報をアノテーションした日本語コーパスを作成した．アノテーションとして用いられたタグの集合として，SFL における状況のコンテクストの要素および言語体系の一部（発話機能）の選択体系網における選択肢を採用した．今回は，とくに社会的役割，とりわけ社会的上下関係の立場が明確なビジネスメールを対象にコーパスを作成した．作成されたコーパスは，アノテーションのタグとして用いられる SFL の選択体系網の選択肢をすべて使っているわけではなく，社会的役割関係を重視したものとなっている．

今後の課題として，作成したコーパスの機械学習課題における利用と性能評価を行うとともに，SFL での状況のコンテクストからテキストが具現される過程を捉えたアノテーション手法の確立を目指す．

1） https://chamame.ninjal.ac.jp

謝辞

この成果は, 国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の助成事業 (JPNP20006) およびアバナード研究奨励金による支援の結果得られたものである.

参考文献

[1] 李舜炯. 談話分析からみた日本語学習者と母語話者の聞き手言語行動の実証的研究. 首都大学東京大学院人文科学研究科・人間科学専攻日本語教育学教室・博士学位論文, 2016.

[2] 松村瑞子・因京子. 日本語談話におけるスタイル交替の実態とその効果. 『言語科学』, No. 33, pp. 109–118, 1998.

[3] 藤原安佐・阿部仁美・大井裕子・椿原博子・吉田則子. 日本語教育における配慮に関わる表現の指導. 『北海道大学大学院教育学研究院紀要』, No. 108, pp. 85–98, 2009.

[4] 文化審議会国語分科会. 分かり合うための言語コミュニケーション (報告). 国語分科会（第 67 回）, 2018.

[5] 小林一郎. 意味へのアプローチ：ハリデー言語学の観点から, 2017.

[6] M.A.K. Halliday. **Language as Social Semiotic: The Social Interpretation of Language and Meaning**. Open University set book. Edward Arnold, 1978.

[7] 平林周祐, 浜由美子. 外国人のための日本語例文・問題シリーズ 10　敬語. 荒竹出版, 1988.

[8] 照屋一博 (編）, クリスチャン・マティスン, ジョン・ベイトマン, ハイジ・バーンズ, M.A.K. ハリデー, 奥泉香, 水澤祐美子. 意味がよくわかるようになるための言語学ー体系機能言語学への招待. くろしお出版, 2022.

[9] 蔡胤柱. 日本語母語話者のＥメールにおける「断り」――「待遇コミュニケーション」の観点から. 『早稲田大学日本語教育研究』, No. 7, pp. 95–108, 2005.

A　付録

表 3: 発話機能と応答（ [8] より引用し一部改変）

やりとりにおける役割	やりとりされるもの	話し手の動き	聞き手の動き	
		開始	応答	
			期待に応じた応答	期待に応じない応答
与える	情報	陳述	承認	否認
求める		質問	返答	忌避
与える	ものとサービス	提供	受容	拒絶
求める		命令	遂行	拒否

表 4: 送信者と受信者の関係に関する設定

	送信者	受信者
ペア 1	学生	教授
ペア 2	学生	会社の人事
ペア 3	学生	サークル全員
ペア 4	学生	アルバイト先の上司
ペア 5	学生	先輩
ペア 6	学生	友人
ペア 7	学生	クラスメイト
ペア 8	学生	アルバイト先全員
ペア 9	学生	大学のスタッフ
ペア 10	学生	自分の大学の学生全員
ペア 11	従業員	上司
ペア 12	従業員	顧客
ペア 13	従業員	自社のある部門全員
ペア 14	従業員	クライアント会社のある部門全員
ペア 15	従業員	同僚
ペア 16	従業員	部下
ペア 17	従業員	社長
ペア 18	教員	同僚
ペア 19	教員	学生
ペア 20	教員	学生全体

表 5: 場面の設定例

【従業員が『自社のある部門全員』にメールで催促する必要がある場面】の解答例
「あなたは経理部に所属しています。あなたは最近、旅費の払い戻し期限が 12 月 10 日 17:00 であることを、全社員に電子メールで送りました。しかし、期限を過ぎても、まだ提出していない社員がいます。全員に対してできるだけ早く提出するよう促すメールを書いてください。」

表 6: メール収集の回答例

場面	あなたは大学生です。今回家庭の事情で半年休学するようになったことを、お世話になっている A 教授にメールでこの件を知らせてください。
回答例	件名：休学について A 先生 いつもお世話になっております。XX 学部１年生の XX です。 実は先月、私の母が交通事故に遇い、大腿骨骨折で入院することになりました足を骨折して歩けなくなりました。 私は母子家庭で、家では私しか面倒を見る人がいないので、学校に通いながら、同時に面倒を見るようにしてきました。 しかし、1 ヵ月間を試して、やはり介護と学業との両立が難しいと感じており、先日、半年休学を申請しました。 これまでいつも丁寧にご指導いただきありがとうございます。来年復学したら、もう一度先生の授業を履修させていただきます。 これからも何卒よろしくお願いいたします。 ーーーーーーーーーーーーーーーーーーーー XX 学部１年生の XX