

平成30年6月5日現在

機関番号：13901

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00201

研究課題名(和文) 言語進化論的アプローチによる法令用語の語義に関する通時的変化のモデル化

研究課題名(英文) Evolutionary linguistic approaches to a model of diachronic changes in the sense of legal terms

研究代表者

中村 誠 (Nakamura, Makoto)

名古屋大学・法学研究科・特任准教授

研究者番号：50377438

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：言語進化研究において、実際に使用される単語の通時的な語義の変化や、その関連用語の動的変化を表したデータベースはこれまで存在しなかった。そこで、本研究では法令用語に着目した。本研究の目的は、語義が通時的に変化する法令用語のオントロジーを構築し、語義変化のモデル化を行うことである。本研究のタスクは、法令用語オントロジーの構築、通時的変化の描画ツールの作成、および実際の言語進化のモデル化のように、大きく3つに分けられる。法令用語オントロジーに関しては、延べ27,737用語の定義語と36,698個の用語巻の上位下位関係が得られた。これをもとに言語進化のモデル化を行い、その有用性を示した。

研究成果の概要(英文)：Simulation studies have played an important role in language evolution. Although a variety of methodologies have been proposed so far, they are typically too abstract to recognize that their learning mechanisms properly reflect actual ones. One reason comes from the lack of empirical data recorded for a long period with explicit description. Our purpose in this paper is to show simulation models adapt to actual language change. As empirical diachronic data, we focus on a statutory corpus. In general, statutes define important legal terms with explanatory sentences, which are also revised by amendment. This study consists of 3 tasks: compiling a legal term ontology, developing a drawing tool for diachronic change in legal terms, and simulation modeling with actual language change. We successfully extracted legal terms, their explanations, and their relations. We found 27,737 terms and 36,698 relation. We showed this database is useful in the study of language evolution.

研究分野：法情報学

キーワード：法令用語 通時的変化 ターミノロジー シミュレーション 言語進化

1. 研究開始当初の背景

言語進化研究とは、分野横断的な学術分野であり、言語学、認知生物学、ロボット工学など多くの研究分野の研究者が携わっている。その中でも、計算機によるシミュレーションが理論の検証に大いに役立っているが、現実と比較してモデルが過度に抽象的になりがちである。ここで、実際の単語の通時期的変化を扱ったオントロジーがあれば、より具体的なモデル化が可能となる。

語彙の通時期的変化を抽出した研究として、WEB上の文書から文を構文解析したコーパスをもとに、ネットスラングによる語彙の変化を発見、分析した研究がある。また、語義変化のモデル化は、このような同じ語義をもつ単語の変化を発見するだけでなく、単語に含まれる語義の変化を発見することが必要である。出現単語ごとにその語義が明示されていれば発見しやすいのは自明である。語義が明示されていれば、その変化を追うことで語義の通時期的変化をとらえることが可能である。

2. 研究の目的

言語進化研究において、実際に使用される単語の通時的な語義の変化や、その関連用語の動的変化を表したデータベースはこれまで存在しなかった。版を重ねた電子辞書を用いてもそれを構築することは難しい。そこで、本研究では法令用語に着目した。法令用語は、その法令で使用する前に語釈文を用いて定義される。また、法令は社会情勢を反映して改正または廃止される。改正には定義規定も含まれることもあるため、定義語の変化と関連用語の増減を通時的に表すことができる。これを体系的に扱うことで、オントロジーを構築することが可能となる。したがって、本研究の目的は、語義が通時的に変化する法令用語のオントロジーを構築し、語義変化のモデル化を行うことである。

法令において重要な用語は、その使用前にあらかじめ定義される。その一方で、法改正によって定義語は追加、変更、削除される(図1参照)。ガス事業法(昭和29年法律第51号)は、これまで30回の改正(うち最後の改正は未施行)が行われ、そのうち定義語の改正は7回あった。(1)制定された当初、定義規定によって定義された用語は2個であった。(2)第1回目の改正によって、もともと「ガス事業」の定義として与えられていた語釈文が「一般ガス事業」の定義となった。「簡易ガス事業」が追加され、これらの上位語として「ガス事業」が新たに定義された。つまり、「ガス事業」の語義が変更され、関連用語が増加したことを意味する。(3)平成26年10月現在、定義規定によって定義されている用語は15個ある。この間に「卸供給」という用語が定義されたが後に削除された。また、「接続供給」という用語は「託送供給」に名称が変更された。これらの消滅した語は、

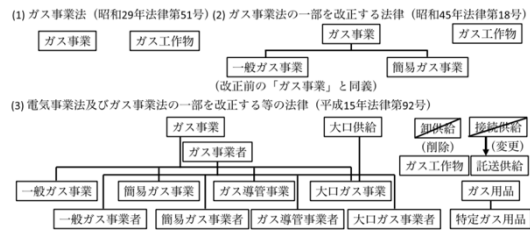


図1 ガス事業法における定義語の変化

世相を反映して淘汰されたと考えることができる。

図1にある「ガス事業」の語義が図中(1)から(2)に変化する部分を該当法律から抜粋すると、以下ようになる。

ガス事業法(昭和29年法律第51号)

第二条 この法律において、「ガス事業」とは、一般の需用に応じ導管によりガスを供給する事業をいう。

ガス事業法(昭和29年法律第51号)(昭和45年法律第18号による改正)

第二条 この法律において、「一般ガス事業」とは、一般の需要に応じ導管によりガスを供給する事業(第三項に規定するガス発生設備においてガスを発生させ、導管によりこれを供給するものを除く)をいう。
(中略)

3 この法律において「簡易ガス事業」とは、一般の需要に応じ、政令で定める簡易なガス発生設備(以下「特定ガス発生設備」という。)においてガスを発生させ、導管によりこれを供給する事業であつて、一の団地内におけるガスの供給地点の数が七十以上のものをいう。
(中略)

5 この法律において「ガス事業」とは、一般ガス事業及び簡易ガス事業をいう。

3. 研究の方法

本研究は、法令文中に現れる定義語の変化を通時的にとらえたモデル化を目的とし、以下の二つのサブタスクからなる。

オントロジーの構築: 定義語の抽出、法改正による語義の変更を行う。

通時変化のモデル化: オントロジーを利用して言語進化のシミュレーションに実装し、通時変化のモデル化を行う。

オントロジーの構築に関しては、定義規定コーパスを作成する。ここに記載されている定義語とその語釈文から(1)上位下位関係の抽出、(2)法律を超えた類似定義語の抽出、(3)法改正による変化の追跡を考慮したネットワークの構築を行う。

に関しては、で獲得した語義や関連語彙の変化を言語進化モデルに取り込み、シミュレーションを行う。

4. 研究成果

4.1 定義規定コーパスの整備

民間の法令データベースから法律の定義規定をクローリングした。ここでは、平成13年以降に有効な全法令および、主要な法令については新規制定以降格納されている。また、廃止法によって定義規定を含む法律が廃止されたものを加えた。結果として、定義規定1,121箇条（ひとつの法律に複数の定義規定を含む場合があるため、法律の本数と定義規定の数は一致しない）を獲得した。このうち、定義規定に変化があるものは537箇条、変化がないものは533箇条、廃止されたものは51箇条だった。

取得した定義語の数は、延べ27,737用語の定義語と36,698個の用語巻の上位下位関係が得られた。関係の精度は、100個のランダムサンプリングによる抽出で88.0%だった。

4.2 ターミノロジーの構築

平成27年4月1日現在有効な定義語延べ6,890語（異なり5,077語）に対するターミノロジーを構築した。それに合わせて、検索・表示を行うためのツールを開発した。

4.3 アニメーション描画ツールの開発

定義語の語釈文から抽出される定義語間の関係について、通時変化のアニメーション描画を行うツールを開発した。定義語の法改正があった法律588本に含まれる定義語5,208語を対称とする。このうちの任意の定義規定について選択、描画が可能であり、同じ時間に共有していた定義語間の関係を可視化することができた。実行画面を図2に示す。

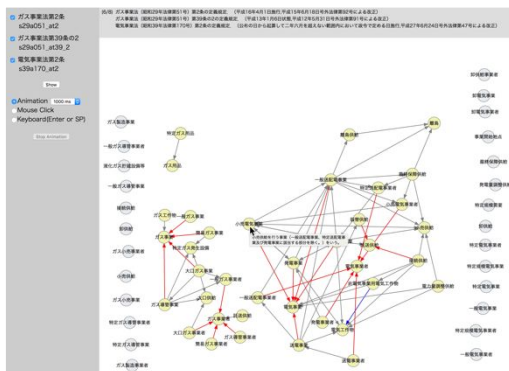


図2 定義語の通時変化描画ツール

4.4 通時変化のモデル化

実際の通時変化を利用した言語進化のモデル化を行った。モデルには、Kirbyによる繰り返し学習モデル(KILM)をベースとした認知バイアスを考慮した文法獲得を行うように修正したモデル(MSILM)を採用した。

KILMとは、親子の間で行われる言語獲得モデルである。話者である親は、自身が持つ文法をもとに発話をするのに対し、聞き手である子は、発話から文法の推論を行う。一定の発話を聞いたあとで世代交代が行われ、子は

親となり、新たな子に発話を行う。これを何世代か繰り返すことで文法が精緻されるというモデルである。

MSILMでは、KILMに2つの次の制約を加えた。(1)親子間でブロックワールドを共有し、発話内容はその世界の出来事に限定する。(2)発話から意味を推測する際、複数の候補が得られたときは、認知バイアスによって曖昧性を解消する。ここでは、ブロックワールドにガス事業法における定義語の関係を用いた。その際、現実の1年は、モデルの1世代に対応することとした。これにより、実際の社会変化の様子をモデル化することで言語獲得にどのような影響を及ぼすかを観察することができる。

実験結果を図3に示す。世代に対する言語の表現能力を表している。実験を50回ずつ行った平均をプロットした。このモデルのポイントは、エージェントがいかに構成的な文法を獲得するかにかかっている。現実の言語変化は不安定であり、30世代付近で大きな変化が起こっていることを示している。このときKILMは、表現能力がそれほど伸びていないのに対し、MSILMでは、表現能力が飛躍したことを表している。このように、実際の言語変化の実態をモデルに組み込むことで、より現実的な実験を行うことが可能となった。

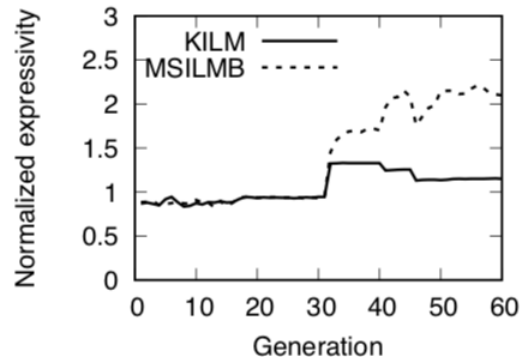


図3 実験結果

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計7件)

Makoto Nakamura, Ryuichi Matoba, Satoshi Tojo: Simulation of Emergence of Local Common Languages Using Iterated Learning Model on Social Networks. International Journal on Advances in Intelligent Systems, **8**(3&4), IARIA, pp.374-384 (2015)

Makoto Nakamura, Yasuhiro Ogawa, Katsuhiko Toyama: Development of Diachronic Terminology from Japanese Statutory Corpora. Journal of Open Access to Law, **4**(1), 16 pages (2016)

Kouhei Okada, Yasuhiro Ogawa, Makoto Nakamura, Tomohiro Ohno, Katsuhiko Toyama: Improvement of Translation Accuracy for the Outlines of Japanese Statutes by Splitting Parenthesized Expressions. Journal of Open Access to Law, **4**(1), 16 pages (2016)

Gen Kawachi, Akira Nagai, Makoto Nakamura, Yasuhiro Ogawa, Tomohiro Ohno, Katsuhiko Toyama: Applying the Akoma Ntoso XML Schema to Japanese Legislation. Journal of Law, Information and Science, **24**(2), pp.49-102 (2016)

Satomi Sakamoto, Yasuhiro Ogawa, Masahiro Ohno, Makoto Nakamura, Katsuhiko Toyama: Utilization of Multi-Word Expressions to Improve Statistical Machine Translation of Statutory Sentences. Lecture Notes in Artificial Intelligence (New Frontiers in Artificial Intelligence), Vol. 10091, pp.249-264 (2017)

Yasuhiro Ogawa, Makoto Nakamura, Tomohiro Ohno, Katsuhiko Toyama: Extraction of legal bilingual phrases from the Japanese Official Gazette, English Edition. Journal of Information and Telecommunication (2018) (in printing)

Takahiro Yamakoshi, Tomohiro Ohno, Yasuhiro Ogawa, Makoto Nakamura, Katsuhiko Toyama: Hierarchical Coordinate Structure Analysis for Japanese Statutory Sentences Using Neural Language Models. 自然言語処理 (2018)(in printing)

[学会発表](計 17 件)

Kouhei Okada, Yasuhiro Ogawa, Makoto Nakamura, Tomohiro Ohno, Katsuhiko Toyama: Improvement of Translation Accuracy for the Outlines of Japanese Statutes by Splitting Parenthesized Expressions. the Seventh International Conference on Knowledge and Systems Engineering (KSE2015) (国際学会), Ho Chi Minh, Vietnam (2015)

Gen Kawachi, Makoto Nakamura, Yasuhiro Ogawa, Tomohiro Ohno, Katsuhiko Toyama: Applying the Akoma Ntoso XML Schema to Japanese Legislation. 2015 Law via the Internet Conference (国際学会), Sydney, Australia (2015)

Makoto Nakamura, Yasuhiro Ogawa, Katsuhiko Toyama: Development of the Diachronic Terminology from a Japanese Statutory Corpus. 2015 Law via the Internet Conference(国際学会), Sydney, Australia (2015)

Satomi Sakamoto, Yasuhiro Ogawa, Makoto Nakamura, Tomohiro Ohno, Katsuhiko Toyama: Utilization of Multi-Word Expressions to Improve Statistical Machine Translation of Statutory Sentences. the Ninth International Workshop on Juris-Informatics (JURISIN2015)(国際学会), Yokohama, Japan (2015)

Makoto Nakamura, Katsuhiko Toyama: Diachronic and Synchronic Analyses of Japanese Statutory Terminology. The Workshop on Legal Texts, Document, and Corpus Analytics (LTDCA2016)(国際学会), San Diego, USA (2016)

Makoto Nakamura: Development of Diachronic Legal Terminology and Its Application. Law via Internet (Lvi) Conference 2016(国際学会), Limassol, Cyprus (2016)

Yasuhiro Ogawa, Makoto Nakamura, Tomohiro Ohno, Katsuhiko Toyama: Extraction of Legal Bilingual Phrases from the Japanese Official Gazette, English Edition. 8th International Conference on Knowledge and Systems Engineering (KSE2016)(国際学会), Hanoi, Vietnam (2016)

Makoto Nakamura, Yuya Hayashi, Ryuichi Matoba: Simulation of Language Evolution Based on Actual Diachronic Change Extracted from Legal Terminology. 9th International Conference on Agents and Artificial Intelligence (ICAART2017)(国際学会), Porto, Portugal

Kazuya Fujioka, Makoto Nakamura, Yasuhiro Ogawa, Tomohiro Ohno, Katsuhiko Toyama: Search Method for Ordinances and Rules in Japanese Local Governments Based on Distributed Representation. the Ninth International Conference on Knowledge and Systems Engineering (KSE2017)(国際学会), Hue, Vietnam (2017)

Takahiro Yamakoshi, Tomohiro Ohno, Yasuhiro Ogawa, Makoto Nakamura,

Katsuhiko Toyama: Coordination Analysis for Japanese Statutory Sentences Using Neural Language Models. the Eleventh International Workshop on Juris-Informatics (JURISIN2017)(国際学会), Yokohama, Japan (2017)

研究者番号：90272989

山腰貴大, 大野誠寛, 小川泰弘, 中村誠, 外山勝彦: 文脈自由文法に基づく法令文の並列構造解析. 平成 28 年度電気・電子・情報関係学会東海支部連合大会, 豊田工業高等専門学校 (2016)

山腰貴大, 大野誠寛, 小川泰弘, 中村誠, 外山勝彦: ニューラル言語モデルを用いた法令文の並列構造解析. 言語処理学会第 23 回年次大会(NLP2017), 筑波大学 (2017)

小酒井款雄, 小川泰弘, 大野誠寛, 中村誠, 外山勝彦: 新旧対照表の利用による法令の英訳修正. 言語処理学会第 23 回年次大会(NLP2017), 筑波大学 (2017)

佐藤充晃, 小川泰弘, 大野誠寛, 中村誠, 外山勝彦: 統計的機械翻訳の利用による法令のあらましの自動生成. 平成 29 年度電気関係学会東海支部連合大会, 名古屋大学 (2017)

栗本雄太, 小川泰弘, 大野誠寛, 中村誠, 外山勝彦: 法令文の統計的機械翻訳における訳語統一の効果. 平成 29 年度電気関係学会東海支部連合大会, 名古屋大学 (2017)

藤岡和弥, 小川泰弘, 大野誠寛, 中村誠, 外山勝彦: 分散表現を用いた類似例規検索. 平成 29 年度電気関係学会東海支部連合大会, 名古屋大学 (2017)

山腰貴大, 小川泰弘, 中村誠, 外山勝彦: CBOW 言語モデルを用いた契約用語の校正手法. 言語処理学会第 24 回年次大会 (NLP2018), 岡山大学 (2018)

6. 研究組織

(1) 研究代表者

中村 誠 (NAKAMURA, Makoto)
名古屋大学・大学院法学研究科・特任准教授
研究者番号：50377438

(3) 連携研究者

外山勝彦 (TOYAMA, Katsuhiko)
名古屋大学・情報基盤センター・教授
研究者番号：70217561

東条 敏 (TOJO, Satoshi)
北陸先端科学技術大学院大学情報科学系・教授