

令和 2 年 6 月 18 日現在

機関番号：12612

研究種目：基盤研究(B) (一般)

研究期間：2017～2019

課題番号：17H02190

研究課題名(和文)大規模データ処理による網羅的データを用いた言語発達機構の解明とその応用

研究課題名(英文) Analysis of language development mechanism using comprehensive vocabulary data by large-scale data processing and its application

研究代表者

南 泰浩 (Minami, Yasuhiro)

電気通信大学・大学院情報理工学研究科・教授

研究者番号：70396208

交付決定額(研究期間全体)：(直接経費) 12,400,000円

研究成果の概要(和文)：本研究を通して、これまで、東京地区において800人規模の幼児の語彙データを収集してきた。また、幼児の語彙発達における変遷及び個人差を明らかにするための新たな指標を提案した。これを用いて、京都地区の幼児の語彙数を個人ごとに観察した結果、この指標は発話語彙数に強く影響されることが分かった。語彙数を基準に語彙獲得を分析すると、障害や、乳児院入居の区別になく、統一的な発達をするという全く新しい知見が得られた。

研究成果の学術的意義や社会的意義

本研究では、語彙発達の過程が、男女、障害や、乳児院入居の区別になく、語彙数にだけに強く依存し、統一的に解釈できることを発見したことは、学術的、社会的に意義が極めて大きい。これにより、語彙数を基準として統一的に、語彙発達の機構の分析を行うことができ、人間の言語獲得プロセスを解明する大きな手がかりとなる。また、発達の時間的な遅れを観測することにより、様々な障害の原因究明や、障害に対する学習手法の開発、生活環境に対する適応的な学習手法の開発など、様々な分野で新しい研究領域を広げることにつながる。

研究成果の概要(英文)：We collected data on vocabulary development based on psychological knowledge and analyzed a vocabulary development mechanism for children using data processing methods. Based on our analysis, we achieved the following in vocabulary development:

1. We collected cross-sectional data on about 900 people over a three-year span. 2. We reached the following conclusions: (1) a strong correlation exists between two regions: Kyoto and Tokyo, indicating almost no regional differences; (2) children with and without developmental disabilities whose vocabularies are about the same size tend to use similar words; (3) we found no significant differences between males and females; (4) in younger children, the effects of children abuses from their parents on vocabulary development appear before they entered the residential care institutions, and this effect decreases as the size of the acquired vocabulary increases. 3. We proposed a small vocabulary checklist to estimate the ability of young children.

研究分野：情報処理

キーワード：語彙獲得 言語発達 語彙発達 発達モデル

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

幼児の言語習得機構の解明には、これまで、語の習得時系列の詳細な観測や、集団としての語彙の習得時期の分析が重要であるとされてきた。しかしながら、縦断データ(特定児少数長期間データ)に関しては、少数のデータによる分析、目視による分析、時間分解能の荒い(一定期間で平均した)分析でのデータ収集が行われたのみで、精密で大規模な分析は行われていない。また、横断データ(不特定児多数ワンショットデータ)では、大規模なデータの収集はあるものの、適切なデータ分析手法を用いておらず、結果として標準となる基礎資料が不足している。この基礎資料の不足が、非定型発達児の言語習得に関する指標の不足にもつながっている。このため、現在、利用されている指標は、言語聴覚士の経験に基づくものが多く、標準的な指標として共有することが難しかった。研究協力者も、言語障害の幼児に関する研究を行ってきているが、このような問題に直面している。

2. 研究の目的

本研究では、定型および非定型発達児(言語発達障害児)の語彙習得に関する縦断データ、および、横断データを収集し分析することにより、研究者や実務者(言語聴覚士、保育士など)に有用な、語彙習得過程や集団としての語彙習得時期に関する標準的な指標を構築する。この構築したデータを用い、

- (1) 定型発達児の語習得の各地域(京都、福島、東京)における標準的な指標を構築し、
- (2) 非定型発達児に特徴的な語彙習得現象を発見し、
- (3) 非定型発達児の語彙発達をサポートする手法の開発、非定型発達児の様々なスペクトル(障害の多様性)の情報を利用した言語習得機構の解明、福島県の発達障害医療の現場へのフィードバックを行う。

3. 研究の方法

本研究では、目的に示した

- (1) 定型発達児の語習得の各地域(京都、東京)における標準的な指標を構築、
- (2) 定型発達児、非定型発達児に特徴的な語彙習得現象を発見、
- (3) 非定型発達児の語彙習得をサポートする手法の開発等に対して、以下の手順で研究を行う。 -1 定型発達児調査の計画策定とアプリの構築、 -2 各拠点での定型発達児の横断データ収集、 -3 各拠点での定型発達児の横断データ分析、 -1 非定型発達児調査の計画策定とアプリの構築 -2 非定型発達児の縦断・横断データ収集、 -3 非定型発達児固有の言語習得現象の分析、 簡易語彙テストの開発、言語習得の認知科学的モデル化、

4. 研究成果

本研究の目的は、心理学的知識に基づいた語彙発達に関するデータを収集し、これまで心理学分野では用いられてこなかったデータ処理手法を用いて、子どもの語彙発達のメカニズムを分析することである。さらに、機械学習法を用いてメカニズムのモデル化を行う。それに加え、その分析結果に基づいて語彙テストを作成し、医療分野の実務者に提供する。これらの目的のために、本研究では、1)語彙発達データの収集、2)語彙発達データの解析、3)小語彙テストの開発に対して以下の成果をあげた。

(1) 語彙発達データの収集

語彙発達データの収集のため、以下の手順を実施した。調布市の保育園、児童センター、こどもセンターなどにおいて、パンフレットを配布し、実験参加者を募集した。応募していただいた

実験参加者を対象に、電気通信大学において、3年間で約900名の横断的データを収集した。

(2) 語彙力発達データの分析

乳幼児の語彙発達について、以下のような分析と・モデル化の研究を行った。

基本的な分析方法の提案

我々は、子どもの言語発達のための語彙共通性の指標（共通ボキャブラリ指数）を提案した。この指標を用いて、京都と東京での横断データを用いて、以下のような子どもの語彙発達パターンを明らかにした。

・語彙のサイズが大きくなるにつれて、語彙発達の初期段階にある子どもは、一定の安定した割合で共通語を発声する傾向がある。

また、品詞による分析でも同様の結果が得られた。この知見から、語彙発達過程は、年齢よりも語彙数に強く依存していることが分かった。また、以前から、言われていた幼児語彙獲得タイプの存在は、この語彙数の違いに由来する可能性が高いことが分かった。この発見は、今後の子どもの言語発達の研究において重要な役割を果たす。

語彙獲得における性差の言語間比較分析

様々な言語において、男女間の語彙獲得順序の違いを調査した。その結果、単語の発達には男女間で有意差がないことが確認されたが、男女間の社会的格差が語彙の発達に影響を与えている可能性があることが確認された。

語彙発達の地域差の分析

幼児の語彙発達の地域差を評価するために、東京と京都の2つの地域間の共通ボキャブラリ指数の相関関係を調べた。その結果、2つの地域間には強い相関関係があり、地域差がほとんどないことが確認できた。また、マニアックな言葉の共通ボキャブラリ指数の相関は、基本的な言葉に比べて弱いことも分かった。このことから、地域差はマニアック語の方に強く表れる可能性があることが示唆された。

非定型発達児の語彙発達の分析

非定型発達児の共通ボキャブラリ指数を調査した。本研究では、26名の非定型発達児の語彙発達データを収集した。その内訳は、自閉症スペクトラム障害の子ども7名、ダウン症の子ども4名、筋ジストロフィーの子ども2名、言語発達遅滞の子ども5名、知的障害の子ども8名であった。この結果、年齢は非定型発達児の発達の遅れを示す結果であったが、語彙の大きさに関しては、非定型発達児と定型発達児の間で共通ボキャブラリ指数は大きく乖離していないことが示された。つまり、定型・非定型に関わらず、語彙数が同じ子どもでは、同じような単語を話す傾向があることが示された。

バイリンガル児の語彙発達過程の分析

本研究では、日英バイリンガルの子どもたちがどのように日英両言語の語彙を獲得するのかに着目し、両言語の詳細な項目別分析を行った。ここでは、生後9~45ヶ月の子ども2,417名の日本語語彙データ、Wordbank から取得したアメリカ英語データ、および両親が日本人または英語の子ども18名の日英バイリンガルデータを用いて分析を行った。この結果、バイリンガル児の英語語彙数は非常に少ないが、日本語語彙数の増加に伴い、バイリンガル児とモノリンガル児

の日本語語彙数の差は縮小していることが分かった。また、バイリンガルの子どもの英語 + 日本語の合計語彙数が、バイリンガルの子どもの語彙数とほぼ同じであることも確認できた。この結果から、日本語と英語の語彙は初期段階では重複せず、環境的により優位な言語（子どもに多くのインプットを与える言語）の語彙が早く獲得されていることがわかる。

入所型養護施設における子どもの言語発達の分析

入所型養護施設で養育されている乳幼児と施設に入所していない乳幼児の語彙発達データを比較し、施設での養護環境が語彙発達にどのように影響しているかを分析した。また、施設入所前に行われた虐待が乳幼児の語彙発達データに与える影響を調査した。これらのデータは、日本語マッカーサー言語発達質問紙の語彙リストを用いて収集した。これらのデータを、語彙の大きさ、単語の獲得順序、品詞の分類の観点から分析した。その結果、虐待を受けた子どもは語彙の大きさに遅れがあるが、語彙獲得順序は普通の家庭の子どもとほぼ同じであった。品詞による分析では、施設の子どもは人に関する単語の獲得が遅れていたが、動詞の獲得は一般家庭の子どもよりも早かった。また、虐待に関する結果は、施設に入る前に起こった虐待が語彙の発達に及ぼす影響が低年齢の子どもに現れ、この影響は後天的な語彙数の増加に伴って減少することを示していた。

語彙学習モデルの提案

本研究では、機械学習法を用いて乳幼児の語彙発達における様々な心理現象を説明するモデルを提案した。このモデルを評価するための実験を行った結果、モデルの妥当性が示された。

(3) 小規模語彙テストの開発

幼児の能力を推定するための語彙数の少ない語彙チェックリストを提案した。これを実現するため日齢 400 日から 1500 日までの区間で、単語の獲得時期を分割し、各区間から単語を選択することで、適切な語彙チェックリストの項目セットを構築した。これにより、2688 項目から項目数を減らして、回答者の負担を軽減し、幼児の能力推定を効率的に行うことができた。また、乳幼児の語彙発達タイプの特徴に合わせて、チェックリストから推定した年齢を補正した。その結果、この補正により推定年齢のばらつきが小さくなることが確認できた。このことから、従来のテストでは外れ値とされていた推定年齢の乳児を減らすことが可能となった。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Yan Cao, Yasuhiro Minami, Yuko Okumura, and Tessei Kobayashi
2. 発表標題 Vocabulary Size As Explanatory Variable for Japanese-Speaking Children's Vocabulary Development
3. 学会等名 ICPS (国際学会)
4. 発表年 2019年

1. 発表者名 Yasuhiro Minami, Tessei Kobayashi and Yuko Okumura
2. 発表標題 Infant Word Comprehension-to-Production Index Applied to Investigation of Noun Learning Predominance Using Cross-lingual CDI database
3. 学会等名 LREC 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Yan Cao, Yasuhiro Minami, Yuko Okumura and Tessei Kobayashi
2. 発表標題 Analyzing Vocabulary Commonality Index Using Large-scaled Database of Child Language Development
3. 学会等名 LREC 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 野口 輝, 南 泰浩
2. 発表標題 ニューラルネットワークと強化学習による幼児の語彙獲得のモデル化
3. 学会等名 電子情報通信学会ヒューマンコミュニケーション基礎研究会 (HCS)
4. 発表年 2018年

1. 発表者名 曹 妍, 南 泰浩, 奥村優子, 小林哲生
2. 発表標題 幼児の言語発達における共通ボキャブラリー指数の提案
3. 学会等名 電子情報通信学会ヒューマンコミュニケーション基礎研究会 (HCS)
4. 発表年 2018年

1. 発表者名 藤田浩貴, 南 泰浩, 小林哲生, 奥村優子
2. 発表標題 多言語における幼児語彙獲得時期の男女間関連の比較
3. 学会等名 電子情報通信学会ヒューマンコミュニケーション基礎研究会 (HCS)
4. 発表年 2018年

1. 発表者名 藤田浩貴, 南泰浩, 小林哲生, 奥村優子
2. 発表標題 多言語コーパスを用いた幼児語彙獲得時期での男女間関連の特性
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

1. 発表者名 塚田元春, 南泰浩, 小林哲生, 奥村優子
2. 発表標題 幼児の簡易語彙能力チェックリスト作成における幼児分類の効率化
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

1. 発表者名 野口輝, 南泰浩
2. 発表標題 DRQNによる幼児の語彙獲得のモデル化
3. 学会等名 言語処理学会第24回年次大会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	小林 哲生 (Kobayashi Tessei) (30418545)	日本電信電話株式会社NTTコミュニケーション科学基礎研 究所・協創情報研究部・主幹研究員 (94305)	