

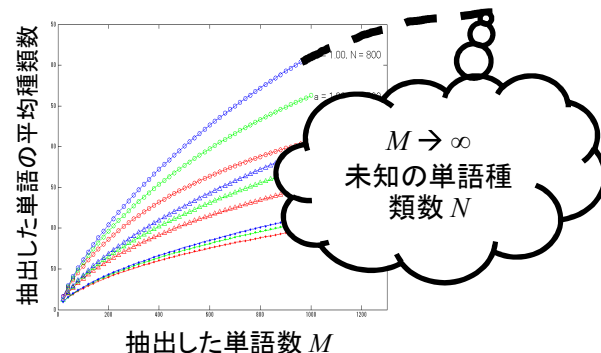
未知なる語彙の計数：単語頻度のロングテール分布の裾を測る



研究者氏名 ひだか しょうへい 日高 昇平	所属機関 北陸先端科学技術大学院大学 先端科学技術研究科	関連キーワード(複数可) 言語発達, 語彙獲得, 潜在種類数の推定
主な研究テーマ ・認知・言語発達、統計的モデリング、非線形力学系、人工知能に関する研究		主な採択課題 ・基盤研究(B)平成27-29年度(配分総額:20,670千円) 「非線形時系列解析による神経結合推定法の開発」 ・基盤研究(B)平成23-26年度(配分総額:8,060千円) 「行動情報解析による意味知識ネットワークの発達過程の解明」

① 科研費による研究成果

- ・幼児の言語学習メカニズムの研究の基礎として、まず幼児が理解あるいは発話する単語の数を正確に把握することが不可欠である。
- ・しかし、単語の発話頻度はジップ則(ロングテール分布)に従い、特に発達期の幼児の語彙獲得が速いため、発話中の単語の計数だけでは正確な語彙数の把握は難しい。
- ・観察データにまれにしか出てこない潜在語彙数を含めて、語彙数を見積もるにはどうしたらよいだろうか。我々は限られた数の発話単語データから、まだ発話されていない潜在的な語彙数の統計的な推定法を開発した。
- ・図は観察した単語数に対する理論的な潜在単語数である。データと類似した理論曲線から潜在語彙数がわかる。



② 当初予想していなかった意外な展開

- ・本プロジェクトで開発した手法により、発達期の子供の単語数をより正確に見積もることができるようになった。
 - ・実際の幼児の発話データに応用することで、観察されるより多くの語彙を知っていることが明らかになった。
1. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=10043982&fileId=S0305000915000094>
 2. 幼児の語彙力統計学で測定, 北國新聞 (2014年9月11日).
 3. 本学プレスリリース
http://www.jaist.ac.jp/whatsnew/press/2014/09/post_417.html

③ 今後期待される波及効果、社会への還元など

- ・発達期の語彙数は言語発達の主たる指標であるため、既存の計数法で低く見積もられていた幼児の言語能力をより正確に見積もることが可能となると期待される。
- ・定説では、18ヶ月前後の急速な語彙獲得期を経て、学習速度は次第に緩やかになるとされる。潜在語彙数が高い水準で増加するならば、新たな語彙発達の理論につながると期待できる。