

この子にぴったりの絵本を探す

～幼児単語に重みを置いたグラフベースの類似絵本検索～

どんな研究

子どもの言語・認知発達を促す手段として、**絵本の読み聞かせ**が注目されています。今回、たくさんの出版された絵本の中から子どもにあった絵本を容易に選べるようにするために、今読んでいる絵本と**類似の絵本を簡単に検索できるシステムを構築**しました。

どこが凄い

これまで困難だった**ひらがな・カタカナが多く含まれるテキストに強い形態素解析器**を開発し、絵本の高精度な解析が可能になりました。さらに、幼児の各発達段階での単語情報を加味するとともに、我々の開発した**グラフ探索アルゴリズム**による、高速な類似探索を実現しています。

めざす未来

絵本を介した子どもの国語力アップのための教育プログラムでの利用、子どもや養育者を対象にした絵本購入のための推薦システム、それぞれの子どもの合った新しい絵本の創作支援システムなどへの応用が考えられます。

・絵本の読み聞かせ

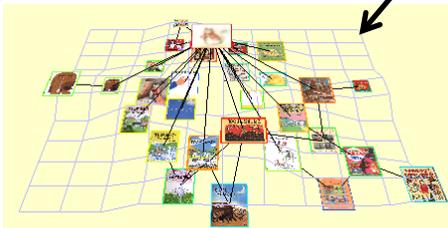
- ・子どもの言語・認知発達に良い影響
- ・しかし、適切な絵本選びは困難



↓そこで

・絵本選びのための類似絵本検索システム

絵本同士の関連性が素早くわかる



出現する単語分布の近さをもとにグラフを作成

ポイント1 ひらがな・カタカナに強い形態素解析器

- ・大規模な辞書を使い、ひらがな・カタカナ語の元の漢字を考慮し、解析器を再学習

入力文(絵本のテキスト)

ぼくらの もりに はらが きたよ

(小手鞠い著、たかすかずみ絵「はじめてのもり」、金の星社、2010年より)

従来の形態素解析システム(かな部分で誤り)

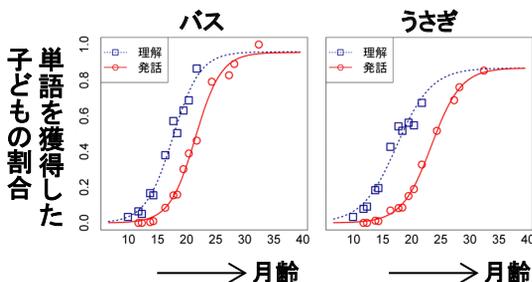
ぼく	ら	の	もり	に	はら	が	きた	よ
名詞	接尾	名詞	動詞	助詞	動詞	助詞	名詞	助詞
ぼく	ら	の	も	に	はら	が	きた	よ

今回開発した形態素解析システム(正しく解析)

ぼくら	の	もり	に	はら	が	き	た	よ
名詞	助詞	名詞	助詞	名詞	助詞	動詞	助動詞	助詞
僕ら	の	森	に	春	が	来る	た	よ

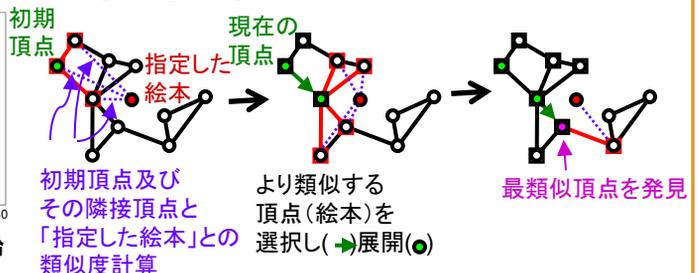
ポイント2 幼児の語彙発達研究の成果を活用

- ・幼児の各月齢で、どのような単語を理解しているかについての研究データの活用



ポイント3 グラフを索引とする類似探索アルゴリズム

- ・有力な従来法(ハッシュ法)と比べても6倍以上高速
- ・検索結果表示を従来のリストだけでなくグラフでも行えるため直感的理解が容易



関連文献

- [1] 平博順, 藤田早苗, 小林哲生, “絵本テキストに現れる語彙の分布,” 日本赤ちゃん学会第12回学術集会, p. 92, 2012.
 [2] 平博順, 藤田早苗, 小林哲生, “絵本テキストにおける高頻度語彙の分析,” 平成24年度情報処理学会関西支部大会, F-103, 2012.
 [3] 藤田早苗, 平博順, 小林哲生, “絵本のテキストを対象とした形態素解析,” 言語処理学会第19回年次大会, A6-4, 2013.
 [4] K. Aoyama, K. Saito, H. Sawada, N. Ueda, “Fast approximate similarity search based on degree-reduced neighborhood graphs,” in Proc. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 1055-1063, 2011.

連絡先

平博順 (Hirotoishi Taira) 協創情報研究部 言語知能研究グループ
 E-mail : taira.hirotoishi[at]lab.ntt.co.jp ({at}の部分をも@に置き換えてください)