

国立国語研究所学術情報リポジトリ

助詞・助動詞接続の計量的分析：
言語処理における多変量解析の応用

メタデータ	言語: jpn 出版者: 公開日: 2020-06-29 キーワード (Ja): キーワード (En): 作成者: 米田, 正人 メールアドレス: 所属:
URL	https://doi.org/10.15084/00002869

助詞・助動詞接続の計量的分析

一 言語処理における多変量解析の応用一

米 田 正 人

1. この分析の目的

新聞語彙調査の結果のうち、助詞・助動詞の接続の資料をもとに、数量化理論第Ⅲ類を用いて、助詞・助動詞の接続のパターン分類を試みる。

2. 数量化理論第Ⅲ類（パターン分類）

数量化理論第Ⅲ類はパターン分類の数量化とも呼ばれ、「同じような反応をする対象は何らかの意味で共通性があり、また同じような性格をもつ対象から選ばれる項目は何らかの意味で共通性がある」という前提のもとに、対象あるいは項目を数量化する手法である。詳細は文献2, 3, 4にゆずる。

3. 分析の対象としたデータの特徴

この分析で用いたデータ（文献1, 26～28pp）で助詞・助動詞の接続といった場合、つぎのことを意味する。

- (1) 助詞および助動詞が直接結びついた結合、つまり、「助動詞+助詞（+助詞）」などといった結合だけを問題にしている。
- (2) したがって、「名詞+助詞+名詞」あるいは「動詞+助動詞+名詞」のような結合をする際の助詞・助動詞は、接続表から除外されている。
- (3) 助詞・助動詞が三つ以上結合する場合、直接隣り合う二つずつに分解し、接続表が作成されている。つまり、「助詞1+助動詞1+助動詞2+END」という結合があった場合「助詞1+助動詞1」「助動詞1+助動詞2」、および、「助動詞2+END」の三つの接続ができる。
- (4) ここでいう「END」とは、文末、句末の「。」「,」「」だけでなく、直後に詞がきて、助詞・助動詞の接続が終了することをも含んでいる。
- (5) 助詞・助動詞の接続表をもとに、接続の後に続く辞をアイテム・カテゴリーとし、前にくる辞をサンプルにみだてて数量化を行った。

4. 分析の対象とした項目

- (1) 前にくる辞

図1にあらわされた67の辞

- (2) 後に続く辞

〔助動詞〕なかつ、なく、ない、なけれ、られ、ず、ざる、ぬ、ん、よう、う、まし、まし、ます、た、たろ、たら、たい、らしい、らしく、べき、だつ、でし、でし、です、ように、よ

うで、ようだ、ようです、ような、そうだ、
〔助詞〕の、と、が、から、を、へ、より、で、に、だけ、まで、ばかり、くらい、ほど、など、
か、も、は、しか、でも、さえ、ながら、ども、ば、ので、し、のに、て、たり、とか、な、
よ、ね、もの、ても、
〔END〕

5. 分析の結果

図1, 図2は数量化理論第Ⅲ類により, 各項目に数量を与え, 前にくる語の第Ⅰ軸, 第Ⅱ軸について, 二次元のグラフを示したものである。

図1は後に助詞が接続する場合のグラフである。第Ⅰ軸では, 後に助詞をとりやすい語ととりにくい語とが分離し, 第Ⅱ軸では, そのうち係助詞を後にとりやすいものが正の領域に, 終助詞を後にとりやすいものが負の領域に描かれている。

図2は後に助動詞*が接続する場合のグラフである。第Ⅰ軸では, 後に助動詞をとりやすい語ととりにくい語とが分離し, 第Ⅱ軸では, 後にくる助動詞の用法の違いで各辞が分離されているようである。今回の発表では, この結果を含めて, さらに分析を試みる。

* ここでは, 4-(2)の助動詞を用法の違いで代表形にまとめて分析した。

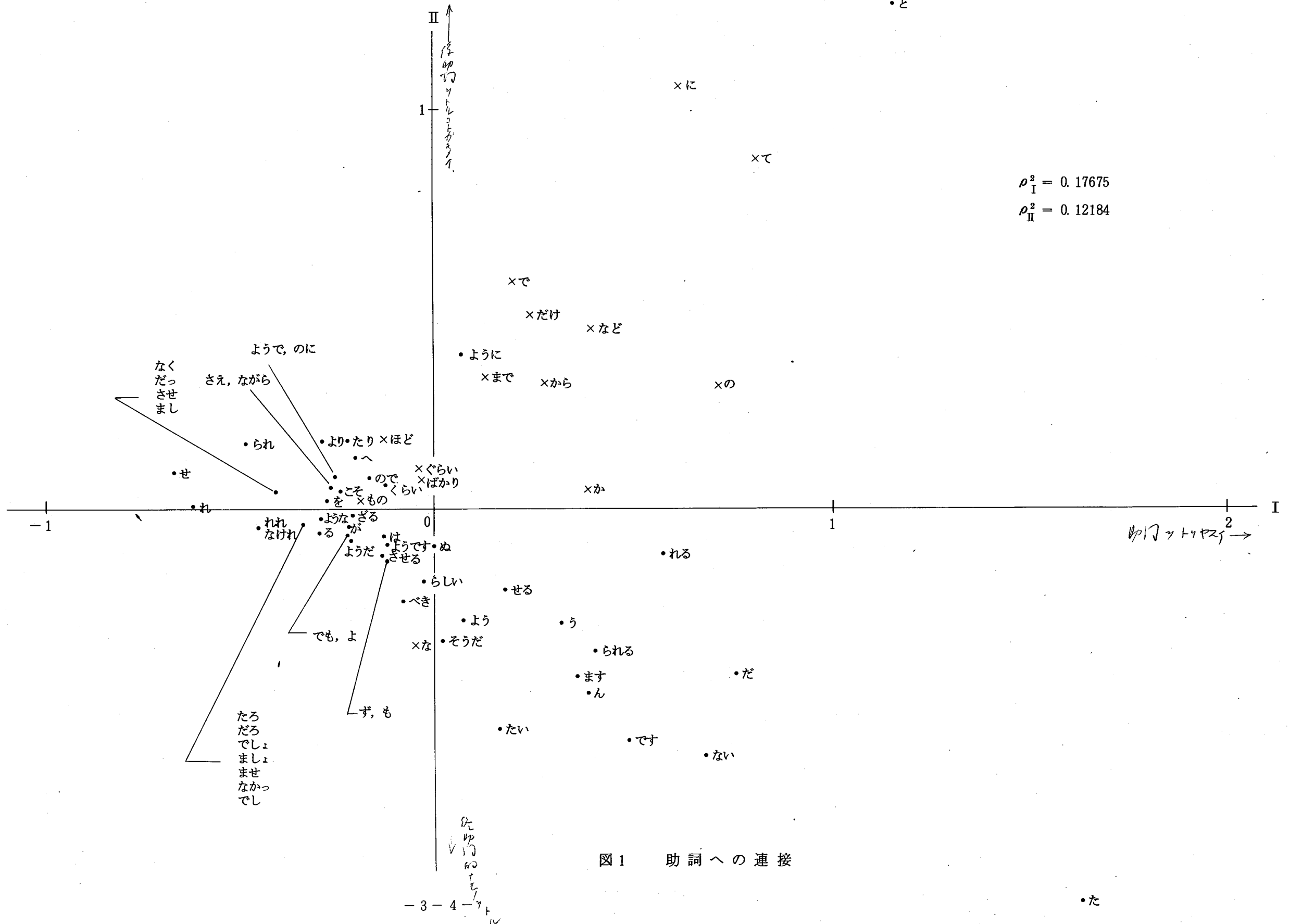
参考文献

1. 国立国語研究所報告42(1972)『電子計算機による新聞の語彙調査(Ⅲ)』
2. 林知己夫他(1970)『情報処理と統計数理』産業図書
3. 林知己夫(1974)『数量化の方法』東洋経済新報社
4. マーケティング・サイエンス研究会(1974)『マーケティング調査』有斐閣
5. 江川清・米田正人(1975)言語の多変量解析(一)『計量国語学』73号 18-28
6. 江川清・米田正人(1975)言語の多変量解析(二)『計量国語学』75号 23-41

〔付記〕

この度の発表にあたり, 統計数理研究所の林文氏に大変お世話になった。深く感謝する。

・と



•た(1.00,2.41)

•ない

$$\rho_I^2 = 0.28039$$

$$\rho_{II}^2 = 0.21742$$

断定
推定 7.14727

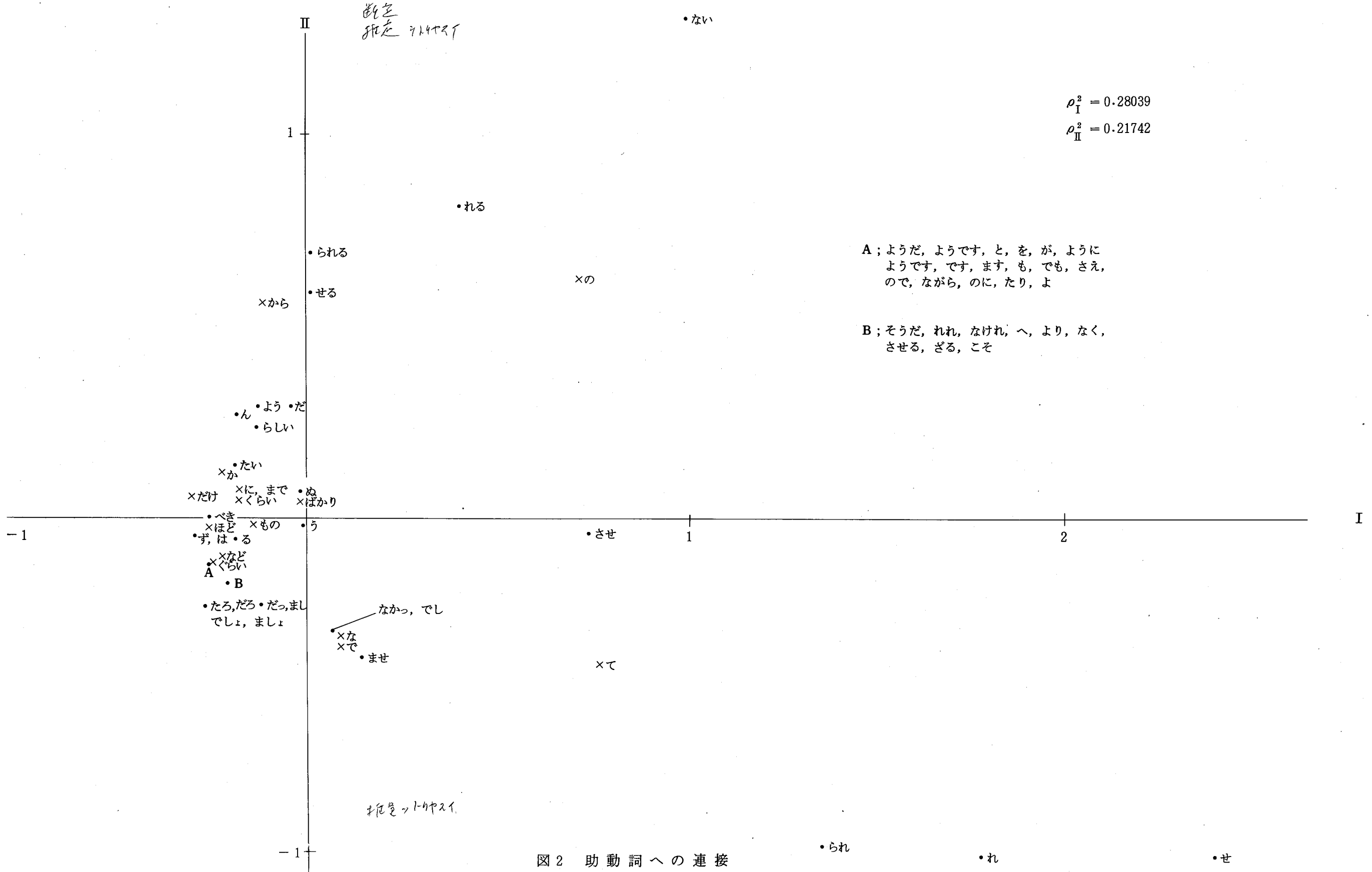


図2 助動詞への接続