

Title	テキストマイニングによる日本の遺伝子組換え関連新聞記事の動向分析
Author(s)	標葉, 隆馬
Citation	年次学術大会講演要旨集, 23: 91-93
Issue Date	2008-10-12
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/7510
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

テキストマイニングによる日本の遺伝子組換え関連新聞記事の動向分析

○標葉隆馬（京都大学）

1：研究の目的

マスメディアは現代において極めて重要な言説空間となっている。マスメディアに登場する言説群は少なからず世論を反映し、更には社会における議論の方向性を左右する存在として陰に陽に社会に対して影響を与えている（e.g. Klapper, 1960: McCombs and Shaw, 1972: Noelle, 1973）。

科学技術に関しても、マスメディア上には数多くの言説が流通しており、その動向を知ることは「科学と社会」との関係を考える上で重要であると考えられる。このような観点から、近年ではマスメディアの動向に関する分析の重要性が、日本における政策決定に関わる場において議論されるようになってきた（e.g. 食品安全委員会 2006）。

そこで本研究では、「遺伝子組換え」に関連する新聞記事を、日本の科学を巡るマスメディア言説群の一例として取り上げ、分析を行う。1973 年に「遺伝子組換え」が登場して以来、「遺伝子組換え」に関連する様々な言説が登場しており、それはマスメディアの言説空間でも同様である。そのことは同時に、マスメディアの言説空間において「遺伝子組換え」に関する 10 年単位での言説の蓄積があることを意味し、科学を取り巻くマスメディア言説のモデルケースとして十分なデータ量が期待できる。しかし、「遺伝子組換え」を取り巻く新聞記事の分析については、これまでにいくつかの先行研究があるものの（e.g. 辻田と白楽, 2003: 西澤, 2006: Hibino and Nagata, 2006）、蓄積された膨大な記事量を生かした形で話題の全体像と時系列的なダイナミクスに対して十分にアプローチがなされているとはいえない。

そこで、本研究計画では、「遺伝子組換え」関連新聞記事について頻出する単語群に注目し、科学計量学的な分析を試みた。科学計量学の観点と手法を用いることで、大量のデータを一度に処理することができ、過去数十年における新聞記事の話題の全体像と時系列的なダイナミクスの提示が可能になる。

2：研究の対象

本研究では、日本の主要な新聞である、朝日新聞と読売新聞における遺伝子組換え関連記事を分析の対象とした。朝日新聞ならびに読売新聞の、オンライン新聞記事データベースを使用し、タイトル・本文中に遺伝子組換え関連キーワードを含む新聞記事を過去 20 年分に渡って収集した。尚、検索に使用したキーワードは、「遺伝子組換え」、「遺伝子組み換え」、「GMO」、「トランスジェニック」である。収集の結果、朝日新聞では 2886 件、読売新聞では 1981 件の新聞記事を収集した。記事数の変化について図 1 に示す。

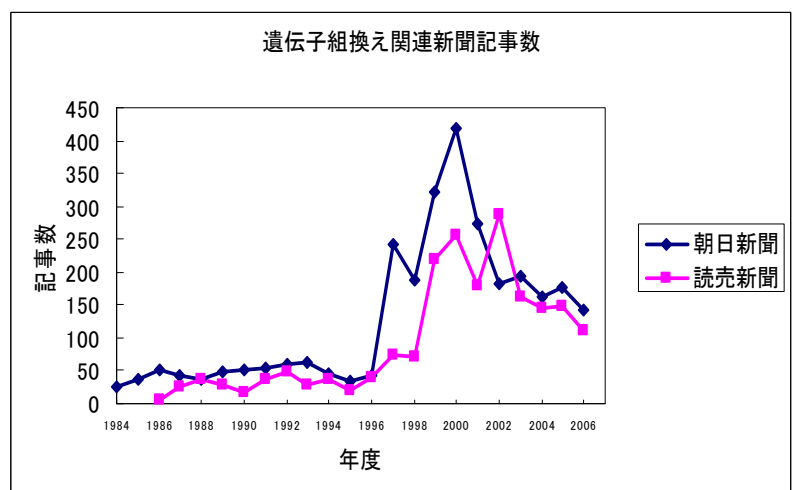


図 1: 遺伝子組換え関連新聞記事数の変遷

3：分析の方法

分析のステップは次のようになっている。

1. 各年度における頻出単語上位 150 語をリストアップⁱ
2. 各年度の頻出単語 150 語から、話題を特徴付けると考えられる 45 個の単語を抽出し、以降の分析の変数として用いるキーワードとして選択ⁱⁱ
3. 選択された 45 の単語が、記事中に含まれているかどうかについて検索し、行列データを獲得

4. 得られた行列データを元に、数量化Ⅲ類分析を実施

4：結果

数量化Ⅲ類分析の結果、日本の遺伝子組換えを巡る新聞記事においては「医療への応用」・「産業への応用」・「遺伝子組換え食品」・「植物研究」が支配的なトピックスとなっていることが示された（図2参照）。

また、日本の遺伝子組換えを巡る新聞記事において、支配的な話題の中心の変化が2度生じている。一度目の変化は1997年頃に生じており（「産業への応用」・「医療への応用」→「遺伝子組換え食品」）、二度目の変化は2003年頃に生じている（「遺伝子組換え食品」→「植物研究」）。これらのことから、遺伝子組換えを巡る新聞報道には三つのフェーズが存在していたと考えることができる。フェーズ1（1980年代～1996年）は「医療・産業」期、フェーズ2（1997年～2002年）は「食品」期、フェーズ3（2003年～2006年）は「研究」期と名づける（図2、3参照）。

フェーズ1からフェーズ2におけるキーワード出現割合の特徴として、①医療ならびに産業に関するキーワードの出現割合が減少（例、「医療：27.1%→13.1%」「免疫：12.9%→4.0%」「微生物：17.2%→5.3%」）、②食品に関するキーワードの出現割合が増加（例、「トウモロコシ：3.8%→25.3%」「ダイズ：3.6%→28.4%」「表示：3.4%→32.8%」）、といった点が見られた。

フェーズ2からフェーズ3におけるキーワード出現割合の特徴として、①食品に関するキーワードの出現割合が減少（例、「トウモロコシ：25.3%→7.3%」「ダイズ：28.4%→14.0%」「表示：32.8%→8.1%」）、②植物研究、特に「イネ」「野外栽培」の出現割合が増加（「イネ：10.4%→16.7%」「野外栽培：5.3%→11.7%」）、といった点が見られた。尚、ここで提示した割合の数値は全て朝日新聞における結果を提示している。

また、「イネ」「野外栽培」「医療」「産業」「コーン」「ダイズ」などのキーワードの出現割合の変遷において、新聞社に違いによる有意差は認められなかった。

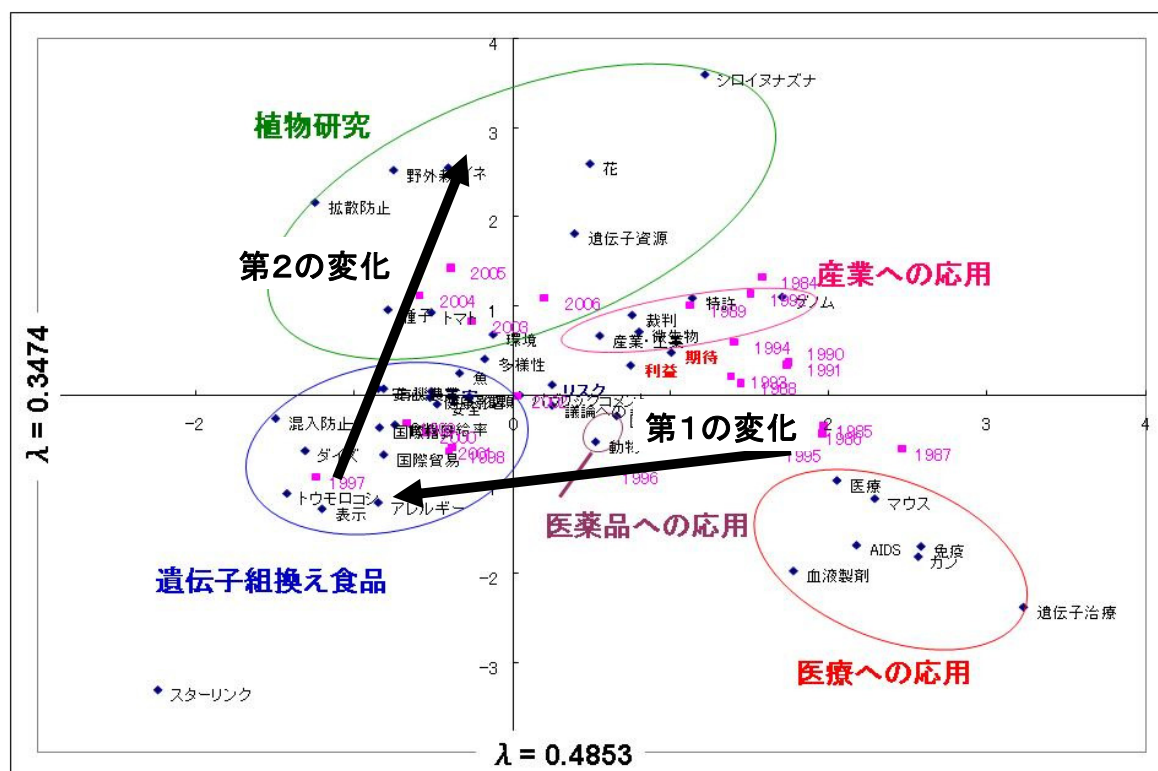


図2:数量化Ⅲ類分析の結果（朝日新聞）

5：考察

日本における遺伝子組換え関連記事において、支配的な話題の中心の変化が2度生じていることが示された。一度目は1997年頃に生じており（「産業への応用」・「医療への応用」→「遺伝子組換え食品」）、二度目は2003年頃に生じている（「遺伝子組換え食品」→「植物研究」）（図3参照）。

一度目の変化に際しては、新聞記事の急激な増加の時期と一致していることが見出せる。1997年前後は、日本において遺伝子組換え食品の安全性の認可が初めてなされた時期であり、生協を始めとする団体による遺伝子組換え食品反対運動が盛んになった時期である。1度目の話題の変化は、この社会的な状況を反映している可能性が考えられる。

2003年頃に生じた二つ目の変化（「遺伝子組換え食品」→「植物研究」）において特に加筆すべき事項としては、①新聞記事の全体数の減少、②「食品」に関するキーワード出現割合の急激な減少、③植物研究関連キーワード（特に「イネ」、「野外栽培」）の増加、以上の3点が挙げられる。これらの結果の解釈としては、遺伝子組換え食品反対運動等により生じた「食品」に関する話題が沈静化し始めたことで、以前から一定割合で存在していた「遺伝子組換えイネの野外栽培試験」などのトピックスが表面化し、目立つようになってきたというものである。この解釈は同時に、「遺伝子組換えイネの野外栽培試験」などの重要なトピックスが、他の話題に埋もれてしまう可能性があることを示唆している。

6：今後の方向性

これまでの研究から、過去20年間における日本の遺伝子組換え関連新聞記事の動向が示された。しかし、その一方で、各記事のトーンや各単語の使われている文脈といった点までは踏み込めていないのが現状である。この、単語の置かれた文脈やニュアンスといった質的な問題に対して、単語の文脈が単語間のつながりに反映されるものと捉え、単語間のネットワーク構造を分析することで、科学計量学の観点から単語のおかれた文脈を分析・可視化できるのではないかという試みが近年なされるようになってきた（e.g. Leydesdorff and Hellsten, 2005, 2006）。そのような研究の知見を踏まえ、日本の新聞記事における各単語の置かれた文脈についての単語間ネットワーク分析を試みることは意味があるものと考えられる。また、同じ単語であっても媒体や領域の違いにより、その置かれる文脈に違いがあるものと予想される。そのような単語の文脈の差異における位置づけの違いについても、単語間ネットワーク構造分析などを通して、示唆を得ることができるかもしれない（e.g. Leydesdorff and Hellsten, 2005, 2006）。単語の文脈という質的な領域に、量的な観点から切り込むことで、異なる領域間における文脈の違い、更には文法や論理構造の違いを分析することが今後の本研究の方向性として考えられる。

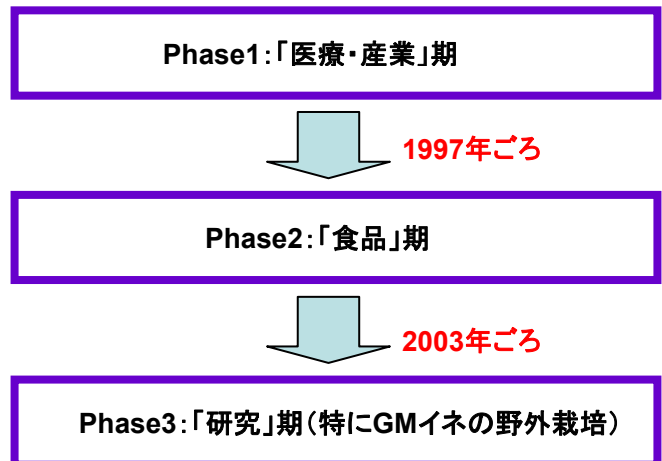


図3：遺伝子組換え関連記事の変遷

- i テキストマイニングのための前処理は正規表現を用いて行い、単語のリストアップにはフリーのテキストマイニングツールである KH Coder (<http://khc.sourceforge.net/>) を使用した。
- ii 頻出単語上位150語を変数としたクラスター分析（ユークリッド距離・Ward法）の結果も参考にした。