

『計量国語学』アーカイブ

ID	KK280502
種別	ノート
タイトル	日本語研究の観点から見た昨今のサーチエンジン事情 —GoogleとYahoo!の技術提携の結果—
Title	On Recent Situations regarding Japanese Search Engines
著者	田野村 忠温
Author	TANOMURA Tadaharu
掲載号	28巻5号
発行日	2012年6月29日
開始ページ	186
終了ページ	193（英文要旨:p.208）
著作権者	計量国語学会

ノート

日本語研究の観点から見た昨今のサーチエンジン事情

—Google と Yahoo!の技術提携の結果—

田野村 忠温 (大阪大学)

キーワード：サーチエンジン，検索ヒット件数，時間変動，検索結果の表示順，Google，Yahoo!JAPAN

1. はじめに

本誌の過去の号に掲載された拙論 2 編（拙論(2008,2009)）において，各種のサーチエンジンのうち特に日本語文書収集量の多い Google と Yahoo!JAPAN の 2 つを取り上げ，それらの表示する検索ヒット件数の調査・分析に基づいて日本語研究の観点からの比較評価を行った。

その後しばらくは，日本語サーチエンジンをめぐる状況に取り立てて注目すべき変化はなかったが，2010 年秋にその様相を一変させる出来事があった。その直後には，事態がどう展開するか不透明な部分があったが，さらに 1 年間余り観察を続けた結果，ほぼ状況の見極めが付いた。

ここでは，そうした経緯をふまえて，日本語サーチエンジンの最新状況について報告する。以後，Yahoo!JAPAN を Yahoo!と略記する。

2. 報告の背景

拙論(2008)において筆者は，2008 年の 1 月から 2 月にかけての約 1 か月間の調査に基づき，Google と Yahoo!のサーチエンジンをデータの量と検索の質という 2 つの面から比較評価した。その結果，日本語の表現を検索してそのヒット件数を見る目的に最も適したサーチエンジンは Yahoo!であるとの結論を得た。データ量に関しては，Yahoo!の収集している日本語文書は Google のその数倍の量に上るものと推定された。また，検索の質の面では，検索時に表示されるヒット件数の論理的整合性と時間的安定性のいずれに関しても Google には問題が大きく，Yahoo!のほうが相対的に信頼性が高いことが分かった。

拙論(2008)を執筆したあと調査はいったん打ち切ったが，その後同年 4 月に調査を再開した。その結果，Google の示すヒット件数の時間変動は，1 度目の調査で明らかになったところよりもさらに大幅で深刻なものであることが判明した。拙論(2009)では，2008 年 4 月から 7 月までの約 4 か月間の調査に基づいてそのことを明らかにし，報告した。もともと，Google と Yahoo!の相対評価に関しては，拙論(2008)の記述から変わるところはなかった。

しかし、それらの拙論で述べた Google と Yahoo!の相対評価は 2010 年の秋に意味を失った。Yahoo!は独自のサーチエンジンの使用をやめ、検索の機能を Google のサーチエンジンに依存して提供するようになったからである。このことは拙論(2011)の一部として簡単に報告したが、以下ではその後約半年間の観察にも基づき、昨今の日本語サーチエンジン事情について報告する。調査を開始した 2008 年 1 月以来約 4 年間の調査に基づく分析ということになる。

3. Google と Yahoo!の技術提携とそれがもたらした結果

2010 年 7 月 27 日に Google は Yahoo!に対する検索技術のライセンス提供を発表した。「Google Japan Blog」のサイトに Daniel Alegre アジア太平洋地域担当副社長（当時）名で掲載された「Yahoo! JAPAN のより良い検索と広告サービスのために」と題した文書に次のように記されている (<http://googlejapan.blogspot.jp/2010/07/yahoo-japan.html>)。

2001 年から 2004 年にはヤフー株式会社に対し、Google の検索エンジンを提供していました。そして、本日、ヤフー株式会社に対して再び検索技術をライセンス提供することになったことを発表いたします。

その後、技術提携を実際に確かめ得る事実はなかなか観察されなかったが、発表の数か月後に Yahoo!は従前のサーチエンジンから Google のそれへの切り替えを行った。筆者が継続的に行ってきた調査における検索ヒット件数の記録によれば、その切り替えは 2010 年 11 月 17 日(水)の午前 10 時に行われた。ただし、両サーチエンジンによる検索を 30 分ごとに実行するという調査の方法上¹、最大で前後 30 分の誤差があり得る。また、巨大な Google, Yahoo!の検索システムの変更は、特定の時刻に一斉にではなく、ある程度の時間をかけて実施されたという可能性も排除できない。

3. 1 検索ヒット件数の共通化

Google と Yahoo!の技術提携の結果、両者の検索内容は非常に近いものになった。Yahoo!が自社の都合で Google サーチエンジンの提供する情報を意図的に操作している事例も現に観察されはしたが²、2つのサーチエンジンの返す検索結果は今やほぼ共通であ

¹ 当初、より短い時間間隔で検索をかけていたところ、Google から検索要求を拒否されるようになった。そこで、30 分間隔で調査するように変更した。

² 2010 年 12 月後半に「虫眼鏡 SEO」というキーワードで検索すると、Google では検索結果の上位に「虫眼鏡 SEO」を実施する業者のウェブサイトが複数表示されたが、Yahoo!ではその大半が表示されなかった。

Yahoo!には「関連検索ワード」と名付けられた機能があり、あるキーワードを指定して検索すると、よく検索される関連のキーワードが提示されるようになっている。例えば、「国語」で検索すると、検索窓の下に、「国語 勉強法」「国語 教科書」「国語力」「中学国語」などのキーワードで再検索するためのリンクが虫眼鏡のアイコンとともに表示される。この機能を利用し、企業名などを関連キーワードの候補として表示させるようにしむける行為が「虫眼鏡 SEO」である（SEO は search engine optimization（サーチエンジン最適化）の略）。これをシステムの悪用と見なす Yahoo!は、対策の一環として、Google の検索結果に含まれる「虫眼鏡 SEO」実施業者のウェブサイトへのリンクの削除（非表示

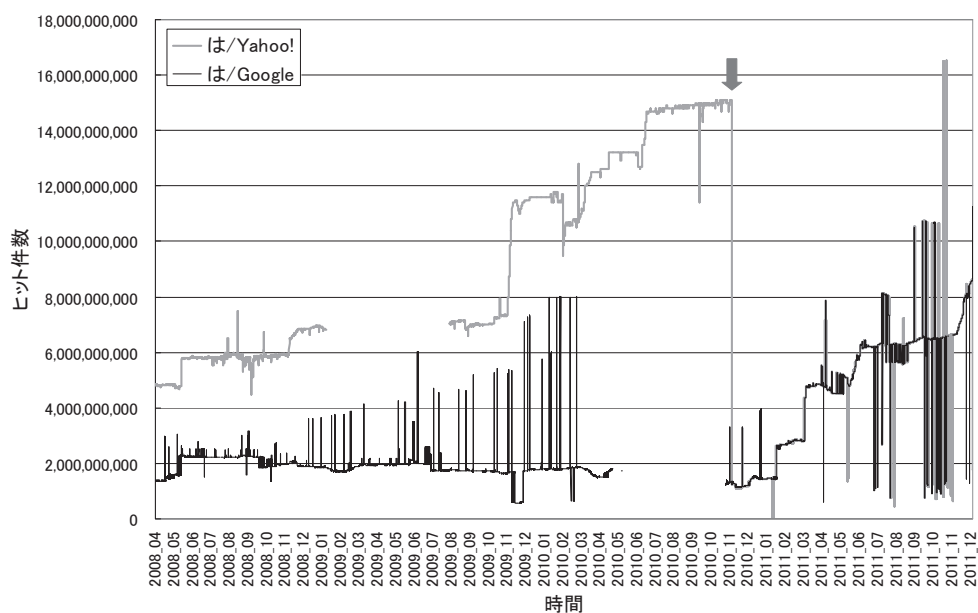
る。このことは、過去に Yahoo!のサーチエンジン検索が持っていた Google に対する量的、質的な優位性を失ったことを意味する。すなわち、Yahoo!のヒット件数は従前の数分の1に減少し、また、Yahoo!も Google 検索のヒット件数の不安定性の問題を抱え込むことになった。

2008 年から 2011 年にかけての、Google と Yahoo!による 3 条件での検索におけるヒット件数の推移の様子を（図 1）～（図 3）に示す。それぞれ、「は」「日本語」「東京に行く」を検索語とした場合のヒット件数の推移で——これらは、助詞、名詞、句など種類を異にする 10 個の表現を適当に定めて調査した結果の一部である——、いずれの検索語も「" "」で囲んで指定して検索している。検索件数の絶対数の範囲は図によって大きく異なり、また、グラフ線の形状にも図による違いが見られるが、全体的に大きく見れば共通するところも大きい。最も重要な共通点は、いずれの図においても、増加を続けていた Yahoo!のヒット件数が 2010 年 11 月を境に一挙に落ち込み（矢印で示した箇所）、以後は Google のヒット件数とほぼ一致する値を示すようになったことである。そしてまた、それまでは時間変動の少なかった Yahoo!のグラフ線も、Google のそれに連動して激しい時間変動を示すようになった。

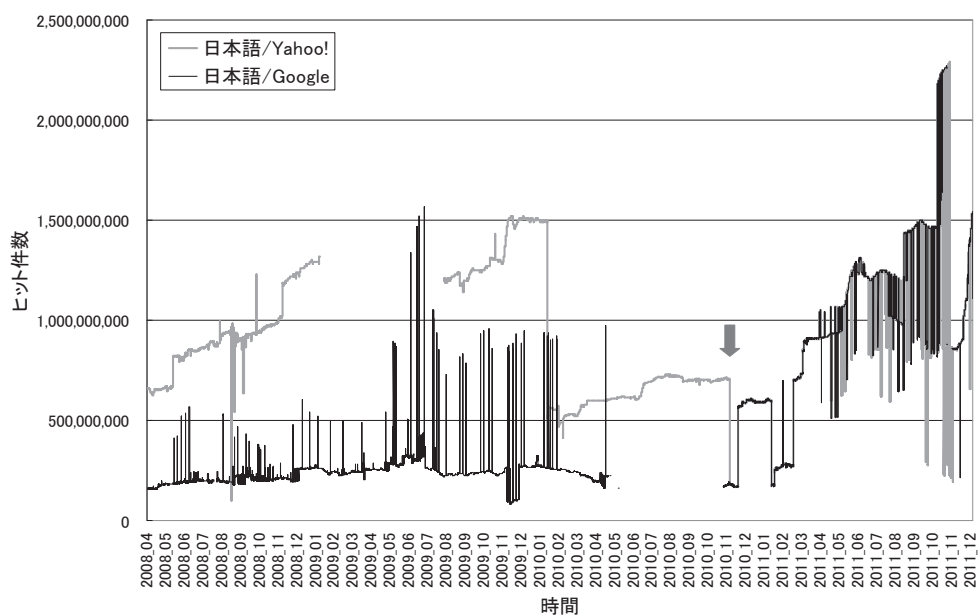
なお、いずれの図においても、2009 年には Yahoo!、2010 年には Google のグラフ線がそれぞれ数か月断続している。これは、筆者が各サーチエンジンの仕様変更気付くのに時間を要し、対処までの期間のヒット件数を記録できなかったことによる。

化)を行った。これにより、検索結果は Google と基本的に共通でも、一部の結果が Yahoo!では表示されないという状況がもたらされた。

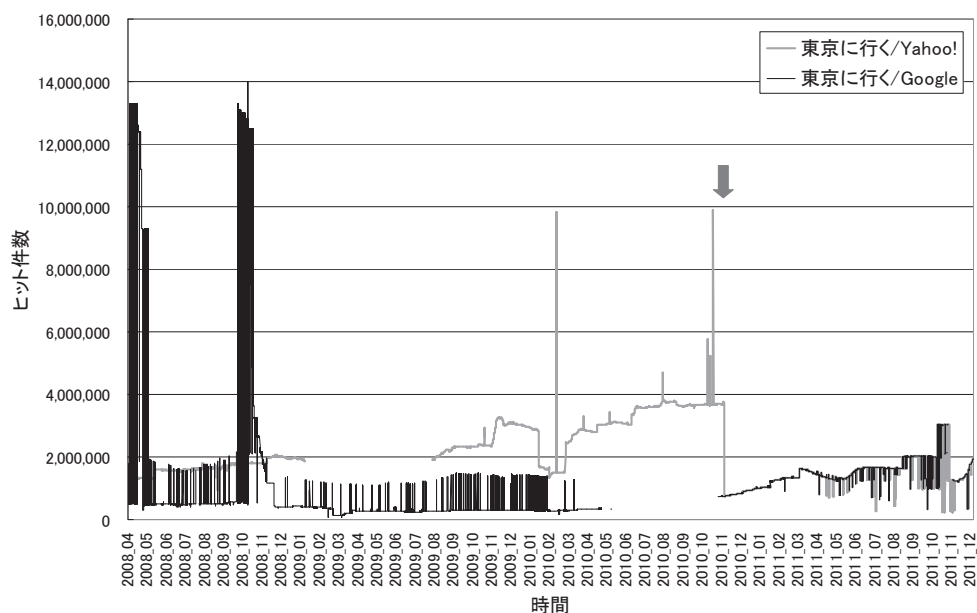
Yahoo!は、「Yahoo!検索スタッフブログ」の 2010 年 12 月 10 日の記事「Yahoo!検索の品質向上に対する取り組み」において、「スパム行為に対してシステム的な措置を講じて排除する」と宣言している (http://searchblog.yahoo.co.jp/2010/12/yahoo_102.html)。実際、同年 12 月 14 日には、「ASCII.jp」に「[速報] ヤフー、虫眼鏡 SEO 業者に制裁措置、インデックスから削除」という記事が掲載されている (<http://ascii.jp/elem/000/000/577/577084/>)。



(図1) 「は」のヒット件数の推移



(図2) 「日本語」のヒット件数の推移



(図3) 「東京に行く」のヒット件数の推移

3. 2 検索結果表示順の共通化

サーチエンジンは、利用者の指定した検索語を含むウェブページを豊富に提示することに加えて、それらをより適切な順序で提示することも重要な開発課題としている。すなわち、利用者にとって有益である可能性の高いページを優先的に示すということである。

この検索結果の表示順も、2010年11月17日以後は Google と Yahoo! とでほぼ共通になった。(図4)は2010年12月に「入学試験」という語を検索したときに各サーチエンジンが上位に提示したウェブページの一覧(ページタイトルと URL)と、両サーチエンジンにおけるその順位の対応関係を示したものである。ページタイトルの右に「※」を添えているのは、他方のサーチエンジンの検索結果上位50位以内に当該のページが現れないものである。

この図に見る、各サーチエンジンによるページの表示順に合理的な説明を与えることは困難であろう。実際、時期を隔てて同じ検索をすると、ページが大幅に入れ替わった順序で表示される。それにもかかわらず、同一の日時に Google と Yahoo! で検索すればかなりの程度に表示順の一致が見られ、偶然の産物として理解できるところを明らかに超えている。3.1 で見たヒット件数の点だけからでも両サーチエンジンの共通化は明らかであるが、検索結果の表示順の点においても同じことが確かめられることになる。

順位	Google	Yahoo!
1	http://ja.wikipedia.org/wiki/入学試験 入学試験 - Wikipedia	http://ja.wikipedia.org/wiki/入学試験 入学試験 - Wikipedia
2	http://www.admissions.keio.ac.jp/ 慶應義塾大学 学部入学案内	http://www.admissions.keio.ac.jp/ 慶應義塾大学 学部入学案内
3	http://www.nyusi.kansai-u.ac.jp/ 関西大学 入学試験情報総合サイト	http://www.nyusi.kansai-u.ac.jp/ 関西大学 入学試験情報総合サイト
4	http://www.dnc.ac.jp/ 大学入試センター	http://www.utsunomiya-u.ac.jp/admission/index.html 宇都宮大学 入学試験
5	http://www.utsunomiya-u.ac.jp/admission/index.html 宇都宮大学 入学試験	http://www.it-chiba.ac.jp/entrance/index.html 入学試験案内 - 千葉工業大学
6	http://www.it-chiba.ac.jp/entrance/index.html 入学試験案内 - 千葉工業大学	http://www.adm.zim.tottori-u.ac.jp/ 鳥取大学 入学試験情報
7	http://www.adm.zim.tottori-u.ac.jp/ 鳥取大学 入学試験情報	http://www.tms.ac.jp/guidance/examination.html 入学試験について 宝塚音楽学校
8	http://www.okayama-u.ac.jp/tp/admission/index.html 入試 - 国立大学法人 岡山大学	http://www.kyoto-u.ac.jp/ja/education/admissions/... 学士入学試験 - 京都大学
9	http://www.juntendo.ac.jp/ac/exam/index.html 順天堂大学入学試験情報	http://www.juntendo.ac.jp/ac/exam/index.html 順天堂大学入学試験情報
10	http://www.nagoya-u.ac.jp/admission/applicant/ ※ 学部入学選抜方法の概要 名古屋大学	http://www.sophia.ac.jp/jpn/admissions/in_ad/... 法科大学院入学試験 上智大学 公式サイト
11	http://www.tms.ac.jp/guidance/examination.html 入学試験について 宝塚音楽学校	http://nightbook.himegimi.jp/escape/room1.htm 入学試験編
12	http://www.aoyama.ac.jp/admission/college/... 青山学院大学 - 青山学院大学 - 入試・入学案内 / ...	http://www.okayama-u.ac.jp/tp/admission/index.html 入試 - 国立大学法人 岡山大学
13	http://www.u-ryukyu.ac.jp/admission/index.html ※ 琉球大学:学部入試課 Entrance Examination Section	http://www.matsuyama-u.ac.jp/nyusi/siken/... 松山大学 入学試験日程
14	http://www.matsuyama-u.ac.jp/nyusi/siken/... 松山大学 入学試験日程	http://www.dnc.ac.jp/ 大学入試センター
15	http://www.kusa.ac.jp/admission/about/ 入学試験について 倉敷芸術科学大学	https://www.wnp8.waseda.jp/goukaku/index.html 早稲田大学入学試験合格者発表掲示板
16	http://www.kyoto-u.ac.jp/ja/education/admissions/... 学士入学試験 - 京都大学	http://www.pu-kumamoto.ac.jp/nyusi/ 熊本県立大学入学試験情報 - TOP
17	http://www.sophia.ac.jp/jpn/admissions/in_ad/... 法科大学院入学試験 上智大学 公式サイト	http://www.kobepharm-u.ac.jp/adms/adms_00.html 入学試験情報インデックス
18	http://textopen.ironmannet.com/ ※ 【入試の鉄人】過去の入試問題(入学試験の過去問題)・...	http://www.iot.ac.jp/examination/index2.html ※ 受験案内: 入学試験情報 - ものづくり大学
19	http://nightbook.himegimi.jp/escape/room1.htm 入学試験編	http://www.kokushikan.ac.jp/future/result/... 入学試験合格発表 入学をお考えの方へ 国士館大学
20	http://www.kobepharm-u.ac.jp/adms/adms_00.html 入学試験情報インデックス	http://www.youtube.com/watch?v=UkdHnE4ynI ※ YouTube - 【お笑い】U字工事『入学試験』
21	http://www.pu-kumamoto.ac.jp/nyusi/ 熊本県立大学入学試験情報 - TOP	http://www.dokuritsugakuen.com/entrance.htm 学校見学、入学試験
22	http://www.u-tokai.ac.jp/admission_app/ ※ 入試情報 東海大学	http://www.t-kougei.ac.jp/admission/arts/exam/ 一般入学試験(前期)
23	https://www.wnp8.waseda.jp/goukaku/index.html 早稲田大学入学試験合格者発表掲示板	http://www.midorigaoka.ac.jp/kushirojo/nyusi/yoko/... ※ 入学試験要項
24	http://www.ris.ac.jp/admission/summary/ ※ 入学試験概要・入試データ立正大学	http://www.keiwa-c.ac.jp/nyushi/nyushi.html ※ 入学試験 - 敬和学園大学
25	http://www.gakushuin.ac.jp/univ/adm/adm/ ※ 入試情報(受験生の方): 学習院大学	http://www.aoyama.ac.jp/admission/college/... 青山学院大学 - 青山学院大学 - 入試・入学案内 / ...

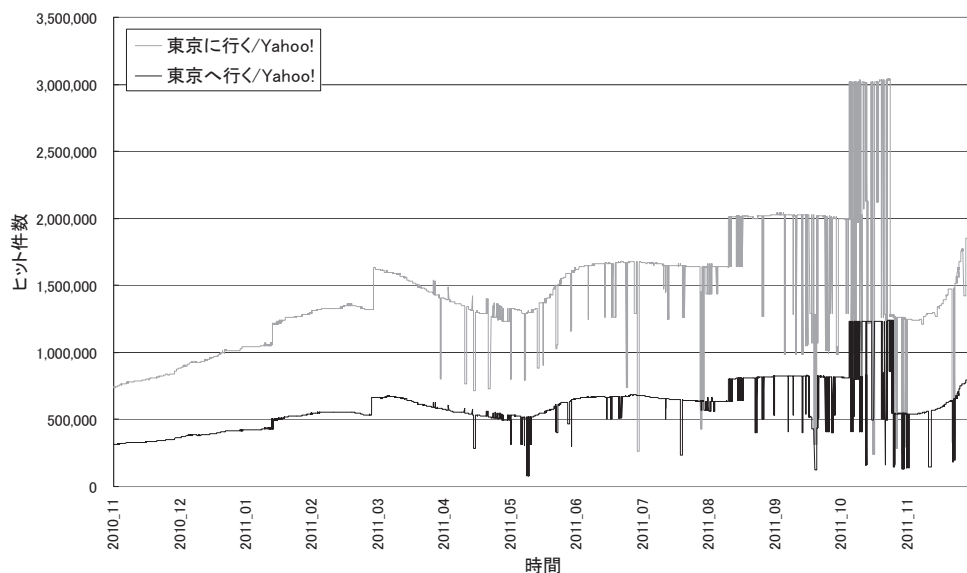
(図4) 検索結果表示順の対応

4. Google と Yahoo!の合一化のその後

2010 年 11 月 17 日以後 Yahoo! と Google のヒット件数と検索結果の表示順が共通化したことを確認してからしばらくのあいだ、その後の状況の展開について予測の立てにくいところがあった。それは、(図 1) ～ (図 3) にも見るように、両サーチエンジン合一化後の数か月間は Google 特有のヒット件数の時間変動が比較的少なく、サーチエンジンが安定性を増した可能性について期待を抱かせる状態であったからである。ただ、合一前にも Google の時間変動が少なかった期間はあるので、合一後の安定した状態がサーチエンジンの改良によるものなのか、単なる偶発的な一時的現象であるのか、判断が付かなかった。

しかし、2011 年後半にもなると、(図 1) ～ (図 3) に見る通り、やはり従前のようにヒット件数の激しい時間変動が見られるようになった。これにより、結局、ヒット件数の時間的変動に関して実質的な改善は実現されていないことが判明した。

(図 5) は、合一化からの 1 年余りの期間における Yahoo! による「東京 {に／へ} 行く」の検索ヒット件数の推移を示したものである。期間の前半ではグラフ線が比較的安定しているが、後半になると過去の Google に似たヒット件数の時間変動が見られるようになる。2 本のグラフ線が交差することもあり、したがって、2 種類の表現のヒット件数の大小が検索日時によって逆転することになる。³



(図 5) 合一化後の Yahoo! による「東京 {に／へ} 行く」の検索ヒット件数

³ 合一前の Google では、比較的安定したヒット件数の“ベースライン”からときに件数がはね上がる形での時間変動が大半であったのに対し、合一後の Google, Yahoo! では、むしろベースラインからヒット件数が落ち込む変動が多い印象がある。したがって、ヒット件数の激しい時間変動が見られるという事実は合一化前後で共通であるが、サーチエンジン内部の処理としては何らかの実質的な変更があった可能性がある。

5. おわりに

以上、Google と Yahoo! の技術提携の結果として、2010 年 11 月より両サーチエンジンの検索結果が共通化したこと、そして、Yahoo! による検索にも従前の Google と同様の検索の時間変動が見られるようになったことを述べた。拙論(2008,2009)で報告した調査・分析の結果のうち、Yahoo! に関わる記述の大半が今日的な意味を失ったことになる。

なお、ヒット件数の信頼性には、拙論(2008)で述べた通り、時間的安定性のほかに論理的整合性の要素もある。後者については継続的な調査は行っていないが、今も前者と同様に問題が大きいことはわずかな検索の試行で容易に確認できる。例えば、「山」を検索するとヒット件数は約 1,330,000,000 件と表示されるが、「山 AND 山」のヒット件数は約 2,190,000,000 件となる⁴。

2010 年 11 月の二大サーチエンジン合一化までは、Google を避けて Yahoo! を使えばそれなりに安心して日本語表現の使用頻度の傾向に探りを入れることができたが、今では頼りになるサーチエンジンのない状態となった。将来の改善に期待したいところではあるが、サーチエンジンはその進歩によってますます“気を利かせた”検索をする——例えば、指定した語句そのものだけでなくその表記上の変異や類義表現も同時に検索する——ようになるので、日本語研究者の望む方向に状況が進むことはあまり期待できないかも知れない。

文献

田野村忠温(2008) 日本語研究の観点からのサーチエンジンの比較評価——Yahoo! と Google の比較を中心に——『計量国語学』第 26 巻第 5 号

田野村忠温(2009) 日本語研究の観点からのサーチエンジンの評価・続——検索ヒット件数の時間変動のその後と Web 文書量の推計の修正——『計量国語学』第 26 巻第 8 号

田野村忠温(2011) 大規模コーパスとしてのインターネット『「現代日本語書き言葉コーパス」完成記念講演会予稿集』（JA 共済ビル・カンファレンスホール，2011 年 8 月 2～3 日）

謝辞 お二方の査読者からの有益なコメントに基づき、記述を見直し、説明を追加した。記して謝意を表したい。

(2012 年 2 月 24 日受付)

⁴ 2012 年 5 月 14 日に Google および Yahoo! で検索したところによる。ヒット件数は 2 つのサーチエンジンで完全に一致した。

Note

On Recent Situations regarding Japanese Search Engines

TANOMURA Tadaharu (Osaka University)

Keywords: search engine, number of hits, temporal fluctuation,
order of search results, Google, Yahoo!JAPAN

Abstract:

In my two previous reports based upon a survey and analysis of search results of Google and Yahoo!JAPAN, I demonstrated that the latter search engine is more suitable for the purpose of searching for Japanese phrases and getting a grasp of their relative use frequency by way of seeing the displayed numbers of hits.

However, the relative estimation of the two search engines lost its validity due to the fact that Yahoo!JAPAN employed Google's search engine in November 2010. Yahoo!JAPAN now returns virtually the same search results as Google does, which implies that Yahoo!JAPAN lost the superiority it used to have over Google.