

『計量国語学』アーカイブ

ID	KK300802
種別	研究ノート
タイトル	サーチエンジンの示すヒット件数の信頼性再び —Google検索をめぐる最新状況—
Title	Reliability of Hit Counts Returned by Search Engines Revisited: The Latest Situation Concerning Google Search
著者	田野村 忠温
Author	TANOMURA Tadaharu
掲載号	30巻8号
発行日	2017年3月20日
開始ページ	499
終了ページ	505
著作権者	計量国語学会

研究ノート

サーチエンジンの示すヒット件数の信頼性再び

—Google 検索をめぐる最新状況—

田野村 忠温 (大阪大学)

要旨

サーチエンジンの表示する検索ヒット件数には時間変動や論理的不整合の問題があることが知られている。筆者が 2008 年以来継続的に行ってきた調査によれば、Google および Yahoo! JAPAN のヒット件数の時間変動のあり方は時期によって異なり、現在 (2016 年秋) は比較的安定した状態から不安定な状態に移りつつあると見られる。論理的不整合についても依然として問題軽減の兆しは認められない。

キーワード: サーチエンジン, ヒット件数の信頼性, 時間変動, 論理的不整合, Google, Yahoo! JAPAN

1. はじめに

サーチエンジンが検索結果とともに表示するヒット件数は不安定で、同じ条件で検索しても日時によってしばしば数値が異なる (田中 (2003), 荻野他 (2007))。また、複合的な条件で検索してみると、しばしば論理的に不整合なヒット件数が表示される。これらの事実は、サーチエンジンの示すヒット件数は言語に関する学術的な議論の証拠として信頼性に問題があることを意味している。

筆者はサーチエンジンの信頼性に関する調査・考察の結果を本誌上で 3 度述べたことがある (拙論 (2008, 2009, 2012))。その内容を手短かにまとめれば次の通りである。以後、Yahoo! JAPAN のことを Yahoo! と略記する。

- ・ 2008 年当時、最も多量の日本語文書を検索対象とするサーチエンジンは Yahoo! であり、そのデータ量は Google の約 3 倍であった。また、Yahoo! のヒット件数は比較的安定していたが、Google のヒット件数には時間変動や論理的不整合が多かった。したがって、日本語の表現を検索し、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! であった。
- ・ 2010 年 11 月以後、Yahoo! は検索機能を Google のサーチエンジンに頼って提供するようになり、両者の検索結果は合一化した。それにより、Yahoo! の有していたデータ量およびヒット件数の安定に関わる優位性は失われた。

2. Google 検索は安定してきたか？

その後、サーチエンジンによる語句の検索とその日本語研究への応用を主題とする荻野 (2014) が刊行された。その第 1 部第 3 章において荻野は各種のサーチエンジンをヒット件数の安定性の観点から比較評価し、「日本語の検索エンジンとしては、2012 年 2 月現在のところ、goo が一番よいと判断する」と結論付けている。そして、「その後、2013 年には、Google や Yahoo! の検索エンジンでも、それ以前ほどの極端なヒット件数の増減は見られなくなった」と補足的に述べている。

この補足の趣旨は「ときに 1 万倍もヒット件数が違ってくるということはなくなった」ということであるが¹、ではヒット件数の時間変動の圧倒的大部分を占める数倍、数十倍までの規模の増減についてはどうなのであろうか。Google と Yahoo! の表示するヒット件数は安定してきたのであろうか。

筆者は 2008 年 1 月から現在 (2016 年秋) に至るまで Google と Yahoo! のサーチエンジンによる検索におけるヒット件数の調査を継続的に行ってきた。次節においては、9 年弱の期間にわたるその調査の結果に基づいて上の問いに関して知り得るところを述べる。結論を先に述べれば、Google と Yahoo! の示すヒット件数の時間変動はその様相自体が時間とともに変動しており、相対的に時間変動の少ない時期もあるが長期的に見れば状況は改善に向かってはいない、そして、現在はむしろ半年、1 年前に比べて状態が悪化している、ということになる。

なお、筆者の調査はもっぱら Google と Yahoo! を対象としているので、goo に関して荻野の述べた評価については確かめることができない。しかしいずれにせよ、荻野 (2014) の刊行後、goo のサーチエンジンは仕様の変更によりもはやヒット件数を表示しなくなった。² したがって、今も我々にとっての重要な関心事は Google と Yahoo! の示すヒット件数の現状の如何である。

3. ヒット件数の時間変動

筆者の調査は、30 分ごとに Google と Yahoo! のそれぞれのサーチエンジンによって約 10 種類の語句を検索し、そのヒット件数を記録するという方法によって行っている。ただし、上述の通り Google と Yahoo! は今では検索内容が共通化しているのので、以下では Google のヒット件数についてのみ述べる。

図 1 は、Google で「は」という仮名 1 字を検索したときに表示されるヒット件数の 9 年弱における推移の様子である。³

¹ 1 万倍に及ぶヒット件数の増減は筆者の調査の範囲では確認できておらず、従来あったとしても稀な現象であったと思われる。

² 荻野氏 (2016 年 10 月私信) によれば、「goo でヒット件数が表示されなくなった時期はきちんと記録していないが、2 年くらい前であったろうか」とのことである。

³ 2010 年にグラフ線が数か月にわたって途切れた期間があるのは、筆者が Google の仕様変更にしづらく気付かず、調査用のソフトウェアをそれに対応させるまでヒット件数の記録を取れなかったことによる。

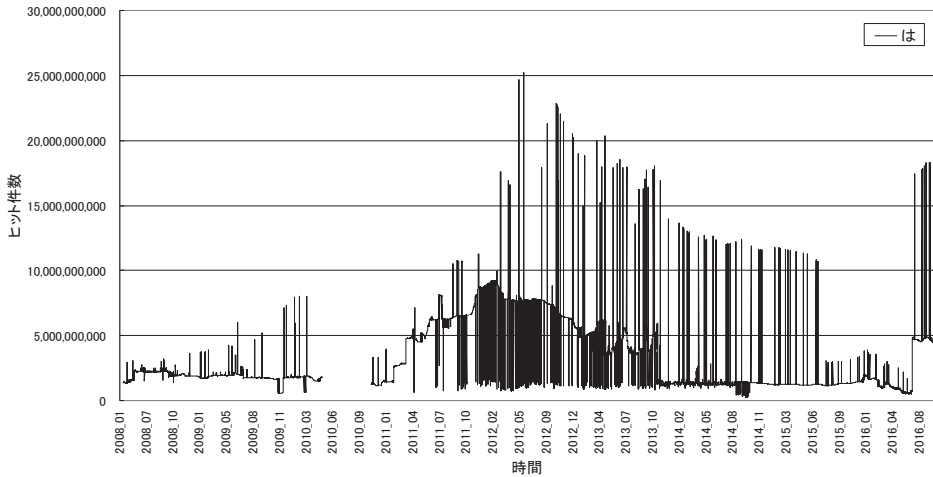


図 1: 「は」のヒット件数の推移

図 1 および以後の図においては、図の見やすさのために、30 分ごとに行っている検索で得られたすべてのヒット件数ではなく、そこから抜き出した 4 時間ごとのヒット件数だけを用いている。したがって、もし 30 分ごとのヒット件数をすべてグラフ化すれば、あるいは、より短い時間幅で調査してそれをそのままグラフ化すれば、時間変動はさらに高頻度で示されることになる。

図 1 から、ヒット件数の時間変動は時期によって様相が異なっていることが分かる。2011 年の半ばから 2015 年の半ばにかけて時間変動が最も頻繁で、特にその期間の前半においては面の一部が黒く塗りつぶされる形になっている。そして、2015 年の半ばからはしばらく時間変動の比較的少ない状態が続いていたが、2016 年の半ばから再び時間変動が増加している。⁴

拙論 (2008) でも述べた通り、ヒット件数の時間変動の様子とタイミングは不思議なことに検索語句によってさまざまに異なる。図 2 に示すのは「日本語」と「英語」を検索したときのヒット件数の推移である。

⁴ 図 1 では、「は」のヒット件数が 2016 年 7 月から急に数倍に増えたために時間変動が増えたように見えやすいという面もあるが、図 2 以下に見る通り、ヒット件数の絶対値には関わりなく現にヒット件数の時間変動は最近増加している。

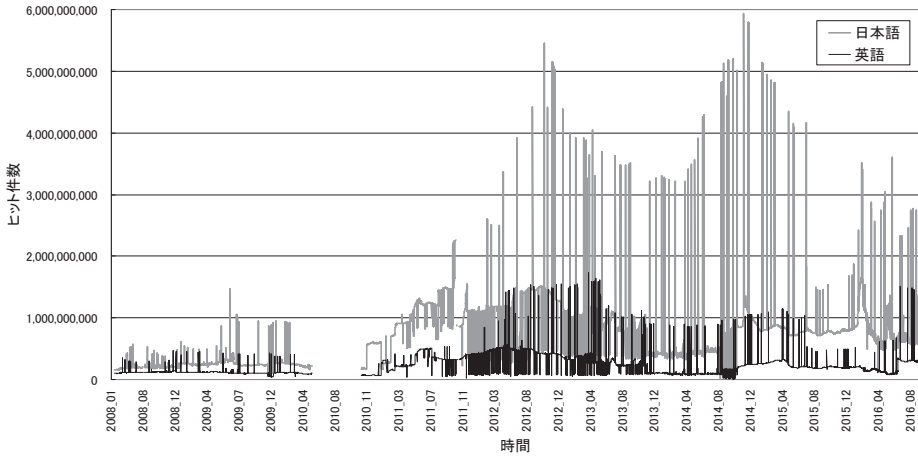


図2: 「日本語」「英語」のヒット件数の推移

図2のグラフは大まかに見れば図1に共通するとも言えるが、細かく見ればヒット件数の増減の様子とタイミングには多くの不一致の要素がある。そうした不一致は図2の「日本語」と「英語」のヒット件数のあいだにも同様に認められ、しばしば2本のグラフ線が交差し、2つの表現のヒット件数の大小が逆転することは今も過去の通りである。

図3、図4はそれぞれ「飲まされた」「飲ませられた」、「東京に行く」「東京へ行く」のヒット件数の推移である。これらについても、図1、図2と共通する点もあれば、それぞれに固有の特徴も認められる。

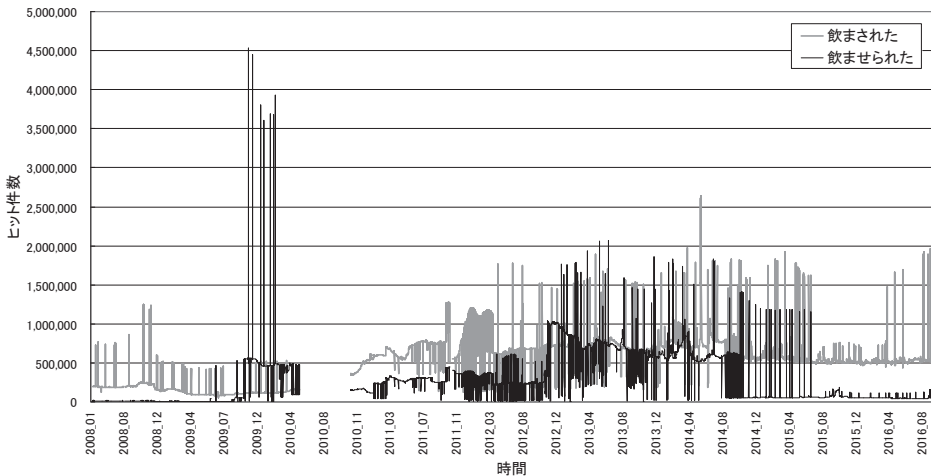


図3: 「飲まされた」「飲ませられた」のヒット件数の推移

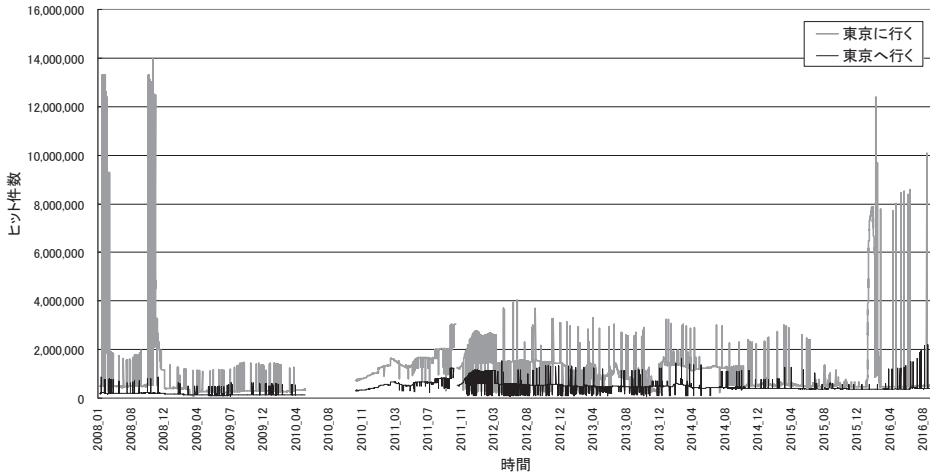


図 4: 「東京に行く」「東京へ行く」のヒット件数の推移

このように、Google のサーチエンジンの表示するヒット件数の時間変動の様相は過去 9 年弱の期間において一定せず、また、2016 年からは再び不安定の度を増していることが確かめられる。このことは 2010 年 11 月以後の Yahoo! による検索にも共通する。

4. ヒット件数の論理的不整合

ヒット件数の時間変動が前節で見たような状態にある以上、ヒット件数の論理的不整合についてもそれが改善されたということはそもそも期待しがたい。また、かりに論理演算 AND, OR, NOT を含む複合的な検索条件が正確に処理されるようになったとしても、時間変動によってヒット件数が不整合になることもあるはずである。

実際、今もヒット件数の論理的整合性に問題のあることは簡単な検索によって確かめることができる。表 1 は、Google を使って OR に関わる 4 通りの条件で検索したときに表示されるヒット件数の不可解な組合せの一例である。ヒット件数の前に表示される「約」は省いて示す。

表 1: OR に関わる検索のヒット件数 (2016 年 10 月 6 日検索)

検索条件	ヒット件数
白	1,640,000,000
黒	910,000,000
白 OR 黒	10,090,000,000
黒 OR 白	1,640,000,000

表1に見るようなヒット件数の明白な論理的不整合は任意の類似の条件による検索を試してみれば高い確率で観察することができる。

5. おわりに

Google, Yahoo!のサーチエンジンの表示する検索ヒット件数の信頼性には依然として問題があり、ヒット件数は2016年から再び不安定の度を増していることを述べた。

田中(2016)は、本誌前々号(第30巻第6号)の特集の趣旨として、サーチエンジンによる検索には不明の要素、不確定な要素が多く、常に知見のアップデートが望まれることを述べている。特集には残念ながらその趣旨に沿った論考や報告が掲載されていなかったため、ここではヒット件数の信頼性をめぐる最近の状況の一端を報告してみた。

なお、gooのサーチエンジンがヒット件数の表示をやめたのはインターネットの流動的特性の1つの現れと言える。そのことを思うと、GoogleやYahoo!のサーチエンジンが今後もヒット件数の表示を続けてくれることを願わずにはいられない。不安定の問題はあるにせよ、ヒット件数が言語に関心を抱く者にとって大きな情報価値を有することには変わりはないからである。

文献

- 荻野綱男(2014)『ウェブ検索による日本語研究』朝倉書店。
- 荻野綱男・末永絵梨・下重秋弓・三好亜萌(2007)「WWWの検索による日本語研究(2)」『東京女子大学日本文学』103:147-166。
- 田中ゆかり(2003)「ネット検索は日本語の研究に有用か」『日本語学』22(5):111-123。(2003年4月臨時増刊号『コーパス言語学』)
- 田中ゆかり(2016)「特集 WWW コーパスを用いた研究 巻頭言」『計量国語学』30(6):325。
- 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』26(5):147-157。
- 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・統一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』26(8):290-294。
- 田野村忠温(2012)「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—」『計量国語学』28(5):186-193。

(2016年10月7日受付)

Note

Reliability of Hit Counts Returned by Search Engines Revisited: The Latest Situation Concerning Google Search

TANOMURA Tadaharu (Osaka University)

Abstract

Hit counts returned by search engines are known to be subject to problems such as temporal fluctuation and logical inconsistency. According to the survey which I have been conducting for some nine years since 2008, the extent of temporal fluctuation of hit counts returned by Google and Yahoo! JAPAN differs from period to period, and presently (autumn 2016) the situation seems to be changing from a relatively stable state to an instable one. There is not yet a sign of reduction in logical inconsistency of hit counts, either.

Keywords: search engine, reliability of hit counts, temporal fluctuation, logical inconsistency,
Google, Yahoo! JAPAN