

E-7

言語間字体類似度が2言語併用者の語彙アクセスを妨げる：

中日同根語の視線計測による検討

Cross-language orthographic similarity hinders the bilingual lexical access:

An eye-tracking study on the processing of cognates by Chinese-Japanese bilinguals.

熊 可欣^{1,2}, 新国 佳祐³, 宋 凌峰¹, 卓 雅¹, 邑本 俊亮¹, 木山 幸子¹

¹東北大学, ²日本学術振興会, ³新潟青陵大学

xiongekexinkuma@gmail.com

要旨

2言語併用者は、言語間同根語（言語間で表記が同じまたは類似する同義語: cognate）¹を一方の言語で理解するとき、他方の言語の影響を受ける。頻度、音韻、統語的特性に加えて書字的特性も影響することが知られているが、漢字圏2言語併用者の言語間同根語の詳細な処理過程は、多くが未解明である。本研究は、漢字語の書字的特性が2言語併用時の語彙処理過程の間でどのような影響を及ぼすかを明らかにすることを目的とする。言語間の字体類似度に焦点を置きながら、中国語と日本語の2言語併用者（中日2言語併用者）が中日同根語（「店员」/「店員」等）の日本語での語彙性判断課題における2度の視線の注視を、判断までの反応時間とともに分析した。その結果、中日2言語併用者が同根語を理解する際には、前方左字の言語間字体類似度と左字の中国語での使用頻度が初期の語彙アクセスを妨げるが、その後の意味を統合する段階では処理を促進することが示された。本研究の知見は、2言語併用相互活性化プラス（bilingual interactive activation plus: BIA+）モデルにおける書字的特性が語彙アクセスの段階では言語間で競合するという仮定に、漢字同根語からの証拠を初めて与えた。

キーワード：言語間漢字同根語, 中日2言語併用者, 注視, 字体類似度, 使用頻度

人間の脳内には、単語や形態素の集合である心内辞書（mental lexicon）があると想定されている。言語間で表記が同じまたは類似する同義語、いわゆる同根語（cognate）の処理機序の検討は、2言語併用者が2つの心内辞書をどのように連絡させ、また制御しているのかの解明につながると考えられている。言語間同根語を一方の言語で理解する際には、対応する他方の言語における当該語の諸特性の影響を受ける。これまで、語彙使用頻度（フランス語と英語：Peeters, et al., 2013）、音韻的特性（日本語と英語：Nakayama, et al., 2012）、統語的特性（中国語と日本語：熊他, 2016）に加え、書字的特性の影響（オランダ語と英語：Dijkstra, et al., 2010）が報告されている。このような2言語の心内辞書の間で語彙の諸情報がどのように関係しあうかを説明するために、2言語併用相互活性化プラス（bilingual interactive activation plus: BIA+; Dijkstra & van Heuven, 2002）モデルが提案されている。このモデルは、語の認識における書字、意味、音韻の3つの表象を想定している。なかでも書字表象の役割については、単一言語使用における視覚的単語認知モデル（McClelland & Rumelhart, 1981）を踏襲して、視覚呈示された語（ターゲット）を理解するとき、その構成要素と類似する書字特性を持つ他のあらゆる語も同時に心内辞書内で賦活してターゲットと競合すると仮定されている。2言語併用者であれば、競合する書字特性は当該言語の心内辞書だけでなく、他方の言語の

¹ 本研究では、同根語における音韻的特性の役割については検討しない。

心内辞書にも及ぶという。

この仮定に基づけば、併用する 2 言語が類似する書字情報をより多く共有している同根語では、そうした競合がより激しくなるため、書字認知に長い時間を要するはずである。ところが、アルファベット言語（オランダ語と英語）における言語間同根語の語彙性判断（その語が当該言語の語彙であるか否かの判断）の反応時間の検証からは、綴りが類似しているほど語彙性判断が速くなると報告されている（Dijkstra et al., 2010 の実験 1）。漢字同根語（中国語と日本語の同根語）でも同様に、言語間で漢字の字体が一致していると、一致していない語より判断が速くなるという（Nakayama, 2002）。このような語彙性判断の反応時間は、刺激語が呈示されてから語彙を確定するまでの総合的な指標であり、初期の書字認知の他に、その後の意味処理も含んでいる。したがって、同根語の語彙性判断の反応時間が短いという知見は、語彙処理の初期段階の書字認知が迅速になされたことを反映しているのか、それともその後の意味理解が促進されたことを反映しているのかが判別できない。アルファベットの綴りであれ、漢字の字体であれ、言語間同根語の書字的特性の類似度が書字認知そのものにどのような影響を及ぼすのか（促進するのか干渉するのか）について、その後の意味理解の段階における影響と切り分けて明らかにするためには、語彙性判断に至るまでの過程を反映する視線行動が指標として有用である。本研究の目的は、語彙性判断の反応時間とともに、その判断の過程でなされる 2 度の注視時間を計測することにより、言語間漢字同根語の書字的特性の類似度がその語彙処理過程で及ぼす影響を明らかにすることである。

2 言語併用者の語彙処理において言語間で共通する書字特性が書字認知の段階で競合を起こすという BIA+モデルの提案は、主にアルファベット語の綴りの類似度の検討に基づくものであったが、この仮定は表音文字以外の書字認知にも適用できるだろうか。本研究では、漢字同根語の言語間の書字的特性の類似度が漢字書字認知に及ぼす影響を特定するために、字体類似度に注目し、中日 2 言語併用者の語彙処理過程を書字認知と意味統合の 2 段階に分けて検討する。視覚呈示された同根語の領域内のどこかを初めて注視してから別のどこかへ移動（サックード）しはじめるまでの時間（一次注視時間）を、初期段階の書字認知処理を反映する指標とみなす。続いて同領域内で起こる 2 度目の注視時間（二次注視時間）を、書字認知後の意味情報の処理を表す指標と考える。これらの注視の指標を、参加者が語彙性判断に要するまでの反応時間と関連付けて同根語の語彙処理過程を検討する。もし、BIA+モデルが中日 2 言語併用者の漢字同根語処理過程にも当てはめられるならば、言語間で漢字の字体類似度が高いほど、初期の書字認知の段階で、L1 の字体と L2 の字体が強く競合するので、一次注視が長くなると予想される。それに対して、書字認知後の意味統合の段階では、字体類似度が高ければ L1 と L2 がより強く結びつき意味理解を促進することから、二次注視および語彙性判断の反応時間は短くなると予測される。

なお、視線計測を用いた単一言語使用時の研究では、漢字 2 字熟語の処理は語全体の語彙的特性だけでなく、各々の漢字が持つ特性、とくに漢字使用頻度が大きく影響することが知られている（中国語：Yan et al., 2010；日本語：Miwa et al., 2014）。また字体類似度は、前方左字が異なる場合（“动物” / 「動物」等）もあれば、後方右字が異なる場合（“音乐” / 「音楽」等）もあり、様々に認識され得る。そこで、本研究は漢字 2 字の中日同根語に焦点を置き、前方の左字と後方の右字の漢字使用頻度と字体類似度を別個に操作し、どの特性がどちらの字により強い影響を及ぼすのかを検討する。あわせて、こうした中日同根語の処理が 2 言語併用者の言語能力に応じてどのように変化していくかも検討する。

方法

実験参加者

中国大陸出身で日本の東北大学に留学中の中国語母語話者 31 名（男性 7 名）が実験に参加した。参加者の平均年齢は 25.92 歳（ $SD = 1.74$ ）であった。全員が日本語能力試験 N1 に合格し、日本滞在歴 1 年以上（ $M = 30.84$ ヶ月, $SD = 13.03$ ）であった。参加者は、実験後に「筑波日本語テスト集 TTBJ」（<http://ttbj-tsukuba.org/p1.html>）のうち SPOT90、Grammar9 および漢字 SPOT50 による日本語能力テストを受け、満点 230 点のうち平均得点は 192.29 点（ $SD = 12.99$ ）であった。実験参加に対して謝礼が支払われた。実験は、東北大学文学研究科倫理審査委員会の承認を受けて行われた。

刺激材料

表 1 のとおり、肯定反応刺激として漢字 2 文字で構成される中日同根語 144 語（完全一致、左側一致、右側一致、両方不一致が 36 語ずつ）に加え、非同根語 36 語（フィルター刺激；「家賃」等）を使用した。全ての刺激語は「日韓中越同形二字漢字語データベース」（于・玉岡, 2015; <http://kanjigodb.herokuapp.com/>）から抽出した旧日本語能力試験 4 級から 2 級までの語彙である。これらの同根語の日本語と中国語における語彙使用頻度と漢字使用頻度を、それぞれ「漢字データベース」（Tamaoka et al., 2017; <https://www.kanjidatabase.com/>）と現代中国語コーパスである「BCC コーパス」（Xun et al., 2016）で調べた。また、否定反応刺激として、常用漢字を用いて作成した無意味語 180 語（「靴来」等）を用意した。課題では、これらの全 360 語の刺激語が参加者ごとにランダムに呈示された。字体類似度は、課題後に本実験の参加者に 7 段階で評定してもらった。

手続き

参加者は視覚呈示による日本語単語の語彙性判断課題を行った。実験は防音室で個別に行われ、実験中は参加者の眼球運動を経時計測した。実験開始前に、5 分程度の視線キャリブレーションを実施した。実験では、モニターの中央に注視点を 500 ミリ秒呈示した後、刺激語を呈示した。参加者は、その語が日本語として正しいかどうかをできるだけ速くかつ正確に判断し、キー押し反応を行った。参加者が語彙判断のキーを押すと同時に刺激語が消え、次の試行に入るように設定した。本番の試行に入る前に 20 試行の練習を行った。本番では、60 試行ごとに小休止をとった。刺激語が呈示されてから参加者が判断するまでの時間（反応時間）、およびその間の両眼の眼球運動が記録された。眼球運動の測定には Eyelink 1000 plus（SR research, Ottawa, Canada）を使用した。実験と機器の制御には Python（ver. 2.7.3）および PyGaze（Dalmaijer et al. 2014）を用いた。

表 1. 刺激語の例、漢字の字体類似度の平均、各言語における漢字使用頻度および語彙使用頻度の平均

タイプ	同根語の例		字体類似度		漢字使用頻度(自然対数)				語彙使用頻度(自然対数)	
					日本語		中国語		日本語	中国語
	日本語	中国語	左字	右字	左字	右字	左字	右字		
両方一致	湿度	湿度	6.92	6.95	11.79	12.36	11.30	11.92	8.48	10.51
左一致	音楽	音乐	6.93	4.09	12.29	11.76	11.92	11.69	8.75	10.88
右一致	動物	动物	4.23	6.89	12.06	12.10	11.79	11.91	8.96	10.99
両方不一致	鉛筆	铅笔	4.07	4.28	11.91	11.82	11.73	11.82	8.63	10.49

注: 字体類似度は本実験参加者による「1. 全く一致しない」から「7. 完全に一致する」までの 7 段階評定。

分析

正しく判断された同根語の反応時間のうち、平均から $\pm 2.5SD$ を超えたデータ(4.24%)を外れ値として分析から除外し、すべての連続変数を標準化した。一次注視時間、二次注視時間、反応時間のそれぞれを従属変数とする線形混合効果モデリング(Baayen, 2008)を行った。各分析では、漢字レベルの要因として、左字の字体類似度(LeftOrthSim)、右字の字体類似度(RightOrthSim)、左字の日本語の使用頻度(LeftJF)、右字の日本語の使用頻度(RightJF)、左字の中国語の使用頻度(LeftCF)、右字の中国語の使用頻度(RightCF)を固定効果要因として含めた。単語レベルの要因として、日本語の語彙使用頻度(JWF)と中国語の語彙使用頻度(CWF)も固定効果要因として用いた。さらに、個々の参加者の日本語テスト(TTBJ)の得点も固定効果要因に含めた。また、一次注視時間と二次注視時間の分析では、次の交互作用を仮定した: LeftOrthSim*TTBJ、RightOrthSim*TTBJ、LeftJF*TTBJ、RightJF*TTBJ、LeftCF*TTBJ、RightCF*TTBJ、JWF*CWF*TTBJ。反応時間の分析では、一次注視と二次注視時間を固定効果に含めるとともに、注視時間の分析と同様の交互作用効果を仮定した。いずれのモデルにも、これらの固定効果要因とともに、ランダム効果要因として参加者の個人差と刺激項目を加えた。すべての分析の有意水準は $\alpha = .05$ とした。分析には、R ver. 3.5.1上でパッケージ lme4 (Bates et al., 2014) および lmerTest (Kuznetsova et al., 2017) を用いた。

結果

参加者の平均正答率は、肯定反応刺激($M=97.17\%$, $SD=16.59$)と否定反応刺激($M=96.95\%$, $SD=17.19$)の両条件とも95%以上であった。肯定反応刺激のうち、同根語のみの平均正答率は99.10% ($SD=0.09$)で、非同根語の平均89.43% ($SD=0.31$)より高かった($p<.000$)。正答された同根語の一次注視時間、二次注視時間と反応時間の各モデリングにおいて、有意となった固定効果を表2に示す。総じて、参加者の日本語能力の有意な影響は見られなかった。

一次注視時間では、2字漢字同根語の前方左字の特性による影響が有意であった。左字が言語間でよく類似するほど、注視時間は長くなった($\beta = .092$, $p < .000$; 図1A)。同様に、左字の中国語での使用頻度が高いほど、注視時間が長くなった($\beta = .053$, $p = .032$; 同B)。その他の左字の特性、右字の諸特性の効果はいずれも有意ではなかった。

二次注視時間では、一次注視と同様に左字の字体類似度および左字の中国語頻度の影響が見られたが、その影響の方向は逆転していた。二次注視時間では、字体類似度が高いほど注視時間が短かく($\beta = -.054$, $p = .045$; 図2A)、また、左字の中国語使用頻度が高いほど注視時間が短かった($\beta = -.066$, $p = .016$; 同B)。また、日本語の単語使用頻度が高いほど、注視時間が短かった($\beta = -.083$, $p = .008$)。後方右字の諸特性による有意な影響は観察されなかった。

反応時間では、左字の字体類似度の主効果が有意であった($\beta = -.061$, $p = .007$)。影響の方向は二次注視と同様で、字体類似度が高いほど反応時間が短かった(図3A)。また、左字の中国語頻度の主効果($\beta = -.085$, $p < .000$)と右字の日本語頻度の主効果($\beta = -.054$, $p = .029$)が有意であった。左字が中国語で頻繁に使われているほど判断が速くなった(同B)。また、右字の日本語での頻度が高いほど判断が速かった(同C)。さらに、単語全体の中国語における使用頻度の主効果も有意であった($\beta = -.097$, $p < .000$)。中国語で単語使用頻度が高い同根語は、頻度が低い語の反応時間よりも短かった。

表 2. 線形混合効果モデリングによる一次注視、二次注視および反応時間に有意に影響する効果

対比	標準偏回帰係数： β	t (df)	p
一次注視時間			
(切片)	.012 [-.117, .141]	0.178 (30)	.860
左字体類似度	.092 [.045, .139]	3.844 (1557)	< .000
左字の中国語使用頻度	.053 [.005, .102]	2.151 (1558)	.032
二次注視時間			
(切片)	-.017 [-.168, .135]	-0.213 (31)	.833
左字体類似度	-.054 [-.104, -.004]	-2.027 (124)	.045
左字の中国語使用頻度	-.066 [-.118, -.014]	-2.435 (127)	.016
日本語の単語使用頻度	-.083 [-.141, -.025]	-2.714 (126)	.008
反応時間			
(切片)	-.027 [-.170, .115]	-0.368 (30)	.715
一次注視時間	.137 [.095, .181]	6.238 (1572)	< .000
二次注視時間	.359 [.316, .404]	15.949 (1581)	< .000
左字体類似度	-.061 [-.104, -.019]	-2.730 (128)	.007
左字の中国語使用頻度	-.085 [-.129, -.041]	-3.682 (130)	< .000
右字の日本語使用頻度	-.054 [-.101, -.008]	-2.209 (127)	.029
中国語の単語使用頻度	-.097 [-.142, -.052]	-4.079 (126)	< .000

注：標準偏回帰係数のカッコ内は 95%信頼区間の下限と上限値。

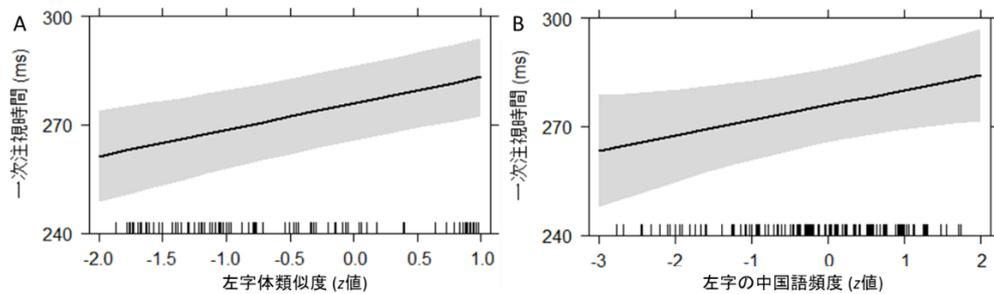


図 1. 左字体類似度 (A) と左字の中国語頻度 (B) が一次注視時間に及ぼす影響

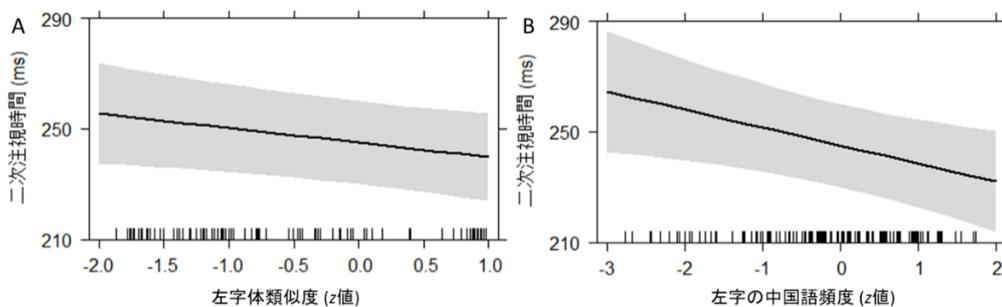


図 2. 左字体類似度 (A) と左字の中国語頻度 (B) が二次注視時間に及ぼす影響

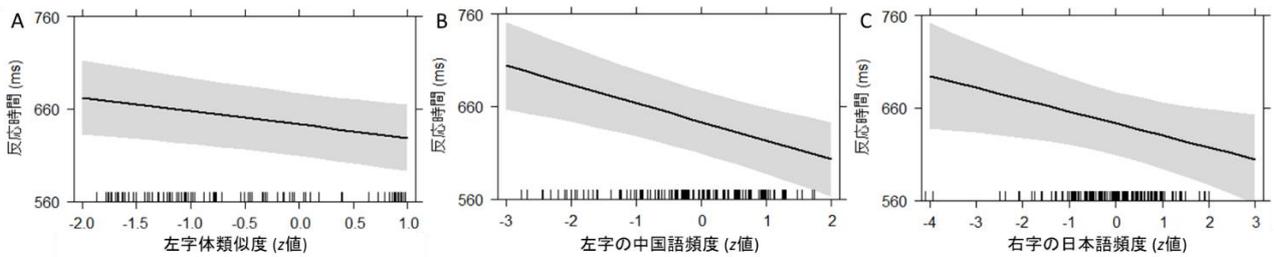


図 3. 左字体類似度 (A)、左字の中国語頻度 (B) と右字の日本語頻度 (C) が反応時間に及ぼす影響

考察

本研究の結果は、中日 2 言語併用者が漢字 2 字の同根語を理解するまでの過程で、一貫して前方左字の言語間字体類似度と中国語における頻度の影響を受け続けることを示している。しかしその影響の仕方は、語彙処理過程の初期と後期とで逆転している。初期の書字認知段階では L1 中国語の情報が L2 日本語の情報と競合するが、後期の意味統合段階では L1 中国語と L2 日本語の語彙情報がともに理解を促進することを示唆している。

具体的には、語彙処理過程の初期の書字認知を反映すると考えられる一次注視時間は、漢字 2 字同根語の前方左字の字体が中国語と日本語で似ているほど、また中国語での使用頻度が高いほど長くなり、処理負荷が高まるようである。言語間で字体が似ていると、またそれが L1 中国語で頻繁に使われる字であると、L2 日本語の処理でも両言語の書字情報がともに活性化し、語彙処理過程の初期段階では干渉すると考えられる。

一方、後期の意味統合段階を表すと考えられる二次注視時間では、一次注視とは反対に、前方左字の字体類似度が高いほど、またその中国語での頻度が高いほど注視が短くなっている。これは、はじめは干渉する情報でも、いったん語彙にアクセスした後は両言語の情報がともに同根語の意味理解を促進し、処理を効率化できるようになるためであると考えられる。

これらの諸段階を含む統合的な語彙処理過程を表す反応時間では、2 字同根語の漢字レベルの特性とともに、語レベルの特性の影響も認められた。漢字レベルでは、左字の字体類似度が高いほど、左字の中国語での頻度が高いほど、さらに右字の日本語での頻度が高いほど、語彙判断にかかる時間が短い、すなわち処理負荷が低いことが示された。単語レベルでは、中国語における単語使用頻度が高いほど頻度の低い語より語彙判断が速くなっている。実験の課題は日本語の語彙判断であったにもかかわらず、参加者の L1 中国語の語彙情報も自動的に活性化され、初期の書字情報の言語間での競合を経て、最終的には中日同根語の理解を促進するのに役立っているといえる。この結果は、アルファベット表記の 2 言語併用者において L1 と L2 のあらゆる語彙情報が非選択的に活性化される (language non-selective activation) という多数の知見 (Peeters et al, 2013 等) を裏付けるものである。漢字圏 2 言語併用者における言語間の語彙情報の非選択的活性化の証拠としては、視線計測による漢字同根語処理の統語情報 (lemma) のレベルでは確かめられていたが (熊他, 2016)、本研究はそれが字体類似度という書字情報のレベルでも起こることを明らかにした。

以上のように、本研究の視線行動と語彙性判断による検討で、漢字同根語の字体類似度が意味理解を促進するまでの間、視覚提示後の初期の語彙認知の段階においては、前方の字体類似度と L1 での使用頻度が処理の妨げになることを初めて示した。この知見は、書字的特性が語彙アクセスの段階において言語間で競合するという BIA+モデルの仮定を、漢字語処理の面から支持するものである。

謝辞

本研究は、日本学術振興会外国人特別研究員奨励費（18F18302）の助成を受けた。

参考文献

- Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge University Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823. R package version 1.1-19. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Dalmajjer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior research methods*, 46(4), 913-921.
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62(3), 284-301.
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175-197.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. R package version 3.0-1. Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, 88(5), 375-407.
- Miwa, K., Libben, G., Dijkstra, T., & Baayen, H. (2014). The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *The Quarterly Journal of Experimental Psychology*, 67(1), 79-113.
- Nakayama, M. (2002). The cognate status effect in lexical processing by Chinese-Japanese bilinguals. *Psychologia*, 45(3), 184-192.
- Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese-English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes*, 27(10), 1563-1583.
- Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, 68, 315-332.
- Tamaoka, K., Makioka, S., Sanders, S. & Verdonchot, R.G. (2017). www.kanjidatabase.com: A new interactive online database for psychological and linguistic research on Japanese kanji and their compound words. *Psychological Research*, 81, 696-708.
- Xun, E., Rao, G., Xiao, X., & Zang, J. (2016). The construction of the BCC Corpus in the age of Big Data. *Corpus Linguistics*, 3(1), 93-118.
- Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, 97(2), 259-268.
- 于劭贇・玉岡賀津雄 (2015). 「日韓中同形二字漢字語の品詞性ウェブ検索エンジン」『ことばの科学』29, 43-61.
- 熊可欣・玉岡賀津雄・マンスブリッジ マイケル パトリック (2016). 「2言語間の非選択的活性化は統語情報の処理においても起こるかー日中同形同義漢語動詞の受動態の処理を例にー」『認知科学』, 23(4), 395-410.