



研究論文

Nuclear Fusion 誌のアブストラクトの統計的解析による 核融合科学分野の研究動向の調査の試み

Study of Research Trend in Nuclear Fusion Research by Statistical Analysis on Abstract of Nuclear Fusion

坂東隆宏, 笠田竜太¹⁾, 小南理恵²⁾

BANDO Takahiro, KASADA Ryuta¹⁾ and KOMINAMI Rie²⁾

量子科学技術研究開発機構, ¹⁾東北大学, ²⁾島根県立大学

(原稿受付: 2020年11月29日 / 原稿受理: 2021年1月7日)

研究者にとって、研究分野の動向を理解することは分野全体の研究方針や自身の研究の独自性を示す上で重要である。研究者は研究分野の動向を知る手段として、論文のアブストラクトの調査を日常的に行うが、日々出版される論文は増加傾向にあり、「要点」が書かれたアブストラクトの調査でさえも負担が大きくなっている。一方でアブストラクトの解析には自然言語処理を用いた統計的な解析が可能である。本研究では、アブストラクトの統計的な分析を核融合科学分野に適用し、研究動向の定量的な評価および可視化を試みた。あるキーワードがアブストラクトに出現する論文数をカウントする手法を採用し、核融合科学分野を代表する Nuclear Fusion 誌のアブストラクトを解析した。具体的には ITER における重要な制御対象である Edge Localized Mode (ELM) を取り上げ、近年 ITER と共に ELM 研究の関心が高まっていることを示す。これら具体例を通して、本手法により研究動向の可視化および考察が可能であることを明らかにする。

Keywords:

research trend, nuclear fusion research, abstract, text analytics, statistical analytics

1. はじめに

研究者にとって、研究分野の動向を理解することは分野全体の研究方針や自身の研究の独自性を示す上で重要である。研究者は研究分野の動向を知る手段として、論文のアブストラクトの調査を日常的に行うが、日々出版される論文は増加傾向にあり、「要点」が書かれたアブストラクトの調査でさえも負担が大きくなっている。

一方でアブストラクトの解析には自然言語処理を用いた統計的な解析が可能である[1]。本研究では、アブストラクトの統計的な分析を核融合科学分野に適用し、研究動向の定量的な評価を試みた結果を紹介する。統計的に解析することが可能となれば、(1)「研究動向の定量的な可視化」に加えて(2)「研究動向調査および研究のグランドデザイン作成の負担軽減」が可能になると思われる。

本研究における解析対象は1961年～2019年の Nuclear Fusion 誌 (NF 誌) のアブストラクトとした^a。現在の NF 誌の author guideline ではアブストラクトの長さを300語以内としており、この単語数の中で論文の内容が説明される。さて、NF 誌は核融合科学分野を代表する査読付き学術雑誌であることから、掲載論文のアブストラクトは分野全体の研究動向を反映していると考えられる。本論文では、

a 本研究では非商用目的である旨を伝えた上で NF 誌編集局からアブストラクトの解析の許可を得た。なお、IOP Publishing は Text and Data Mining (T&DM) Policy を Web 上に公開[2]している。

指定したキーワードがアブストラクトに出現する論文数をカウントし、該当する論文数の推移を「研究活動の活発さ」の変化として評価した。この「研究活動の活発さ」の年次推移を見ることで、研究動向の変化を考察した。

本稿の第2節では、アブストラクトの収集手法およびカウント手法の詳細を述べるとともに、キーワードがアブストラクトに出現する論文数が研究活動の活発さを知る指標として有効かどうかの検討を行う。第3節で研究動向の解析例を示す。ここでは、近年論文での出現回数が増加傾向にある研究キーワードとして、「ITER」における重要な制御対象の一つである「ELM」と、これの制御手法である「RMP」を取り上げる。第4節では本研究で検討した手法の意義とその限界について議論する。第5節で結びの言葉を述べる。

2. 研究手法

2.1 アブストラクトの収集手法

アブストラクトを含むデータベースは、NF 誌の公式 Web ページからダウンロードした html ファイルを用いて作成した。ダウンロードした html ファイルからタイトル、アブストラクトなどのデータを抽出し、データベース化し

た。html ファイルのパースとしては Python 言語のライブラリである BeautifulSoup を用いている。NF 誌の掲載記事にはいわゆる会議報告や書評などが含まれているが、これらは前処理において極力除去した。

2.2 キーワードのカウント手法

先に述べたように、本解析では、指定したキーワードがアブストラクトに出現する論文数をカウントする。指定したキーワードが1つのアブストラクトに複数回出現した場合でも「1」としてカウントする^b。複数回の出現を「1」とカウントしたのは、キーワードの出現回数そのものではなく、キーワードがアブストラクトに出現した論文数自体をカウントすることが研究動向の把握のための指標として適切であると考えたからである。また、単一の単語の出現に加えて、指定した二つのキーワードが同時に出現する「共起」[4]もカウントした。共起をカウントすることで、二つのキーワードが同じ文脈で使われている論文数を評価できると考えた。

指定したキーワードが出現する論文数については、その年の出版論文数で規格化した値を考察する必要もある。図1にNF誌の出版論文数の年次推移を示す。1961年から1978年ごろまでは緩やかに増加し、その後2000年までは多少の増減を示しながらも横ばい傾向が続いている。そして、2000年から2019年までは再び増加傾向^cにある。キーワードがアブストラクトに出現した論文数を用いて研究動向を評価する場合、カウントした論文数の増加だけで判断してしまうと年当たりの出版論文数の増加のトレンドと区別がつかなくなる。そこで、本研究ではカウントした論文数をその年の出版論文数で規格化した値も評価した。

2.3 キーワードがアブストラクトに出現する論文数が研究活動の指標として有効かどうかの検討

核融合科学分野においては「装置の運転期間では装置に関する研究活動が活発」と推定される。この推定を真として、「キーワードがアブストラクトに出現する論文数は研究活動の活発さの指標」とできることを示す。ここでは、すでにシャットダウンした装置名を指定して、その装置名が出現した論文数の年次推移と運転期間の相関をみる。シャットダウンした装置の一例としてはTFTRを取り上げ

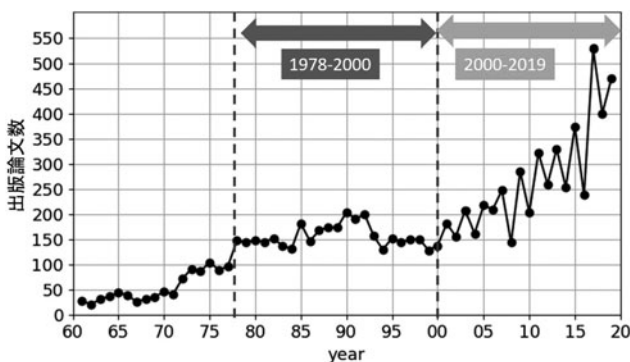


図1 NF誌の出版論文数の年次推移。2019年までの累積出版論文数は9700本程度。

る。TFTRは、1982年～1997年まで運転した米国の装置[6]であり、重水素と三重水素を用いたいわゆるDTプラズマ実験を行った数少ない磁場閉じ込めプラズマ実験装置である。図2には、アブストラクトで「TFTR」が出現した論文数およびその出版論文数に対する割合を示した。「TFTR」がアブストラクトに出現した論文数およびその割合は運転期間と正の相関があることがわかる。このことから、キーワードがアブストラクトに出現する論文数が研究活動の指標として有効であると判断した。

さらに、図2について論文数の時間変化をより詳細にみると、運転期間中に徐々に論文数が増加し、シャットダウン後は4年程度で減少している。運転期間中に徐々に論文数が増加しシャットダウン後に徐々に減少するという傾向はJT-60U, ORMAK, TEXTORでも確認された。

なお、ほかの磁場閉じ込め装置の近年の動向については、参考までに付録1に示した。

3. 研究動向の解析例

本節ではITERに関連した研究キーワードについて研究動向の解析例を示す。ITER計画の成功は核融合炉の実現に向けて重要なマイルストーンであり、研究分野全体としても関心が高いと思われる。実際、図3(a)で示すように、アブストラクトで「ITER」が出現する論文数は年々増加しており、近年では年当たりの出版論文数の20%～30%でITERに言及していることがわかる。ITERに関連する研究テーマとしては様々なものがあるが、ここではEdge Localized Mode (ELM) [7]に着目する。ELMはプラズマ周辺で励起する磁気流体力学的モードであるが、ELMに起因する間欠的な熱流束がダイバータに過渡的に高熱負荷を与えることが知られている。この熱負荷がダイバータの損傷を引き起こすことから、ELM制御がITERにおいて必要不可欠とされている。図3(b)ではアブストラクトに「ITER」と「ELM」が共起した論文数の年次推移を示した。ここでは、「ITER」に言及する論文数とともにその数が増大しており、ITERと同様に関心が高まっている様子が確認できる。「ELM」は関心が高まっている研究キーワード

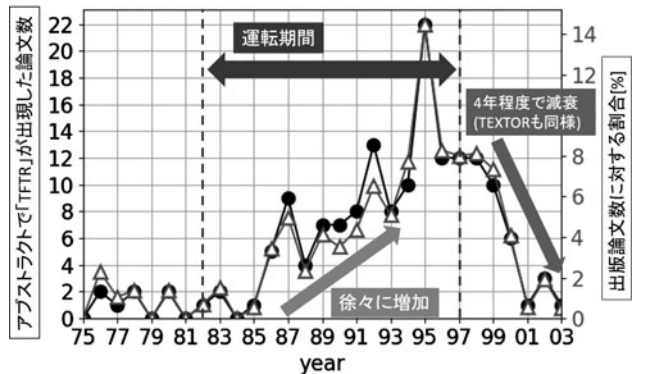


図2 「TFTR」がアブストラクトに出現した論文数(○)と年当たりの出版論文数に対する割合(△)の年次推移。

b 論文数のカウント方法について関心がある方は、文献[3]などを参照いただきたい。

c 2010年から2018年まで増加する様子は全分野的な傾向と一致している。文献[5]の図表4-1-1を参照いただきたい。

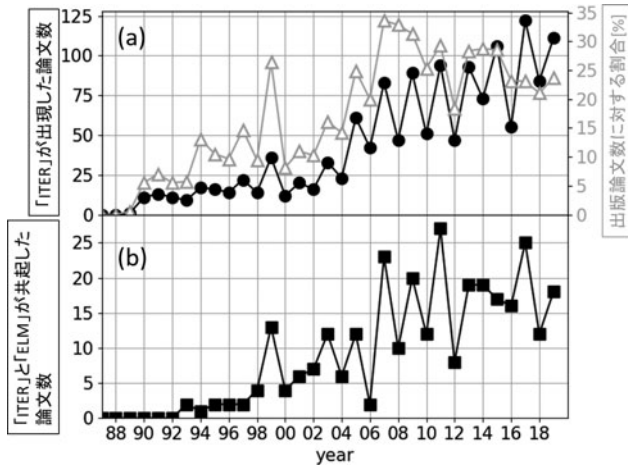


図3 (a)「ITER」がアブストラクトに出現した論文数 (○) と年当たりの出版論文数に対する割合 (△) の年次推移. (b) 「ITER」と「ELM」がアブストラクトで共起した論文数の年次推移.

であることが可視化できたが、では、一体どの装置での研究が活発に行われているだろうか。図4では、アブストラクトでELMと共起する装置名の論文数を積み上げ面グラフで示した。ELMと共起する回数が多いほど、より積極的にELM研究を行っていると考えられる。図4からわかることとして、まず、DIII-DおよびJETでは2000年以降長期的にELM研究を積極的に行っていることが推察できる。2008年にシャットダウンしたJT-60U[8]でもELM研究が積極的に行われていたことが読み取れる。一方で、近年ELM研究が活性化している装置としてはEASTがあり、2011年ごろから増加傾向にあることが読み取れる。「EAST」をアブストラクトに含む論文数は2011年ごろから増えており(付録1の図A-1を参照)、先の解析結果と合わせてEASTではその初期から積極的にELM研究を行っていることが推察される。

ELM制御の重要性については先に述べた通りであるが、実際にELMを制御する手法としては、Resonant Magnetic

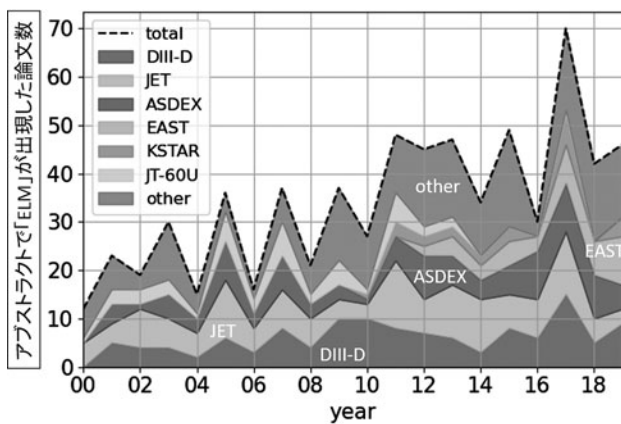


図4 「ELM」と各装置名(DIII-D, JET, ASDEX, EAST, KSTAR, JT-60U)が共起した論文数の積み上げ面グラフ.

d この割合については次の通りに書ける。すなわち、キーワードXがアブストラクトに出現する論文数をC(X)とし、キーワードXとキーワードYがアブストラクトで共起する論文数をC(X & Y)とすると、 $C(ELM \& RMP)/C(RMP)$ として書ける。

e 磁場閉じ込めプラズマでは特に水素の同位体による閉じ込め改善効果[12]のこと。

Perturbation (RMP) [9]による制御手法が提案されている。この背景のもと、「ELM」と「RMP」が共起することが予想されるが、これについても図5で実証された。まず、図5(a)から、「RMP」がアブストラクトに出現する論文数は増加傾向にあることがわかる。そして図5(b)にみられるように、「RMP」がアブストラクトに出現した論文数のうち「ELM」と共起した論文数の割合^dは30%~60%であり、RMPはELM研究の文脈で語られていることが読み取れる。

ここまで、ITERに関連する研究キーワードについてその年次推移をみてきたが、新たな物理現象の実験観測から論文数が増加した研究キーワードもある。たとえば、ASDEXにおけるプラズマコアから発せられたと思われるIon Cyclotron Emission (ICE)の観測[10]を発端として「ICE」をアブストラクトに含む論文数の増加が観測されている。他にも、近年の機械学習に関する技術革新[11]により、機械学習に関するキーワードを含む論文数の増加が観測された。なお、ここまで年次的に論文数が増加したキーワードを紹介してきたが、キーワードを含む論文数が増加した後に減少した例も確認されており、その様子を付録2に示した。

本節では論文数の年次推移を見てきたが、ある年に着目しデータを分割表に示すことで、個々の装置がどの研究キーワードの論文を特に執筆しているか検討することができることがわかった。2019年のデータを用いてELM, RMP, isotope effect^eについてまとめた様子を表1に示す。2019年にELM, RMP, isotope effectがアブストラクトに出現した論文数は、列名に記載した通りそれぞれ46, 27および5である。これら研究トピック(研究キーワード)で共起した

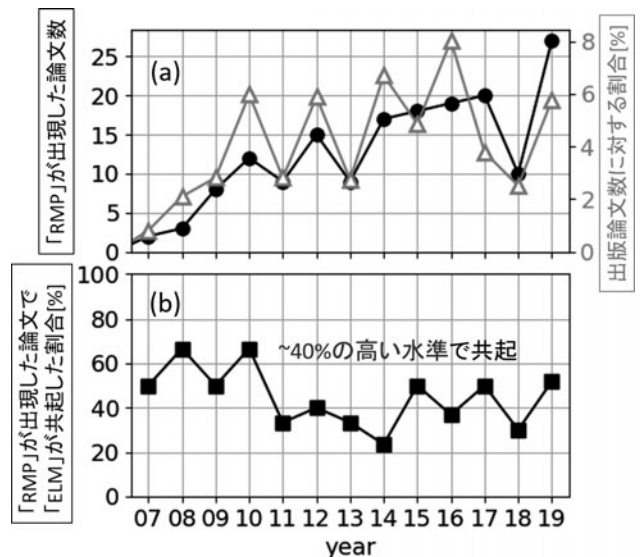


図5 (a)「RMP」がアブストラクトに出現した論文数 (○) と年当たりの出版論文数に対する割合 (△) の年次推移. (b) 「RMP」がアブストラクトに出現した論文数に対する「RMP」と「ELM」が共起した論文数の割合の年次推移.

表1 ELM, RMP, isotope effectがアブストラクトに出現した論文数および各キーワードと装置名がアブストラクトで共起した論文数に関する分割表(2019年のデータ)。各値の意味は表中の凡例を参照。

kwdと装置名が共起した論文数(A)		kwd<kwd>が出現した論文数(B)>			2019年のデータ		
装置名	kwd	ELM<46>	RMP<27>	isotope effect<5>			
DIII-D		9 [9/46 ~ 19.5%]	5 [5/27 ~ 18.5%]	0 [0/5 ~ 0.0%]			
EAST		10 [10/46 ~ 21.7%]	3 [3/27 ~ 11.1%]	0 [0/5 ~ 0.0%]			
LHD		0 [0/46 ~ 0.0%]	3 [3/27 ~ 11.1%]	2 [2/5 ~ 40.0%]			

(A)/(B) × 100

割合が高いのは、ELMではEAST・DIII-Dであり、RMPではDIII-Dであり、isotope effectではLHDであることがわかる。LHDは近年重水素プラズマ実験を開始し、これまでの軽水素プラズマ実験の知見を活かして積極的にisotope effectの研究に取り組んでいる様子が推察される。

4. 本手法の意義とその限界

本研究では、研究分野の動向を統計的な解析から取得することを念頭に、指定したキーワードがアブストラクトに出現する論文数をカウントする手法をNF誌のアブストラクトに対して用い、研究動向の可視化を試みた。第2節、第3節で示したように、本手法により研究動向の可視化および可視化結果の無理のない考察が可能であることを示した。当分野の現役の研究者であれば何が流行しているのかは日々の研究活動の中で「なんとなく」把握することはできていたが、定量的に可視化する手法はなかったため、本研究が提示した手法の意義は大きいと考える。また、定量的な可視化を通じて、個々の研究トピックの論文数に寄与している装置の比較検討が可能であることを示した(図4や表1)。こうした比較検討を行うことで、研究の独自性を再確認し今後の研究方針を検討することもできると考える。

なお、論文中の文字列を解析する場合、選択肢の一つとしてアブストラクトに加えて本文も解析対象とすることが考えられる。ただし、論文の本文を解析対象とする場合、論文の背景説明などで論文の要点とは関連性の低い単語が出現する可能性がある。この観点から、論文の要点を取得するという本研究の文脈ではアブストラクトのみを解析することが重要であると考えられる。

ここで、本手法にはいくつか限界があることを指摘しておかねばならない。まず、指定したキーワードをカウントするという特性上、複数の名前がある同一の概念(energetic-particle-driven wall modeとoff-axis-fishbone-like modeが一例[13])や特徴などが定まっておらず名称に表記ゆれのある概念はカウントすることができないこと

があげられる。また、キーワードを用いて傾向を見るという特性上、解析者が「あたり」をつける必要があり、これは、「流行しているものの解析者が感知していないキーワード」の指定が難しいことを意味する^f。

5. 結語

本稿では、指定したキーワードがアブストラクトに出現する論文数をカウントする手法を用いて研究動向を可視化および考察する方法を示し、その意義と限界について述べてきた。当たり前ではあるが、より広い視点で科学技術研究の様相を知るためには本手法だけではそもそも不十分なことは注意せねばならない^g。たとえば、単に出版された論文数だけではなく、その研究に従事する研究者数などほかの情報を加えることで、より詳細な分析ができるようになる。また、個々の研究者は「研究のインパクト」にも関心があるだろうが、そのインパクトを測るためには引用分析[15]など複数の手法を併用する必要がある。本研究で提示された手法が核融合科学分野の科学技術研究の理解の一助になることを願う。

謝辞

本稿はFusion2030研究会における発表を元にしております。研究会にて発表の機会をくださった坂本隆一博士、森芳孝博士に御礼申し上げます。また、アブストラクトの解析の許可をくださったNF誌編集局に感謝いたします。加えて、ELM研究について御教示いただいた大山直幸博士、ITB研究について御教示いただいた本多充博士に御礼申し上げます。最後に、核融合科学分野の研究動向について日ごろから議論いただいている金史良博士、隅田脩平博士に感謝いたします。

参考文献

- [1] T. L. Griffiths *et al.*, "Finding scientific topics" *Proc. National academy of Sciences* 101, suppl 1 5228 (2004).
- [2] IOP Publishing, Text and Data Mining (T&DM) Policy. <https://iopscience.iop.org/info/page/text-and-data-mining> (2020/12/22 閲覧).
- [3] 中村優文: 情報の科学と技術 70, 315 (2020).
- [4] 小林雄一郎: ことばのデータサイエンス (朝倉書店, 2019).
- [5] 文部科学省: 科学技術指標2020 (2020).
- [6] Wikipedia, List of fusion experiments. https://en.wikipedia.org/wiki/List_of_fusion_experiments#Tokamak (2020/11/15 閲覧).
- [7] 鎌田 裕: プラズマ・核融合学会誌 82, 566 (2006).
- [8] N. Oyama and the JT-60 Team, *Nucl. Fusion* 49, 104007 (2009).
- [9] 武智 学: プラズマ・核融合学会誌 88, 162 (2012).
- [10] R. Ochoukov, *Rev. Sci. Instrum* 89, 10J101 (2018).
- [11] 浜口智志: プラズマ・核融合学会誌 95, 535 (2019).

f 力業を使ってよいのなら解決する方法はある。すなわち、全出現単語の年次推移をみて、ある特徴的な年次推移のパターンを持つ単語だけ抽出することで流行しているキーワードを推定することである。ただし、本解析において全出現単語数は1万7千語程度あり、現実的な時間で評価するには検索する単語の候補を削減する必要がある。

g 他の手法については例えば文献[14]で紹介されている。

[12] H. Yamada *et al.*, Phys. Rev. Lett. **123**, 185001 (2019).
 [13] M. Okabayashi *et al.*, Phys. Plasma **18**, 056112 (2011).
 [14] 鳥谷真佐子：情報管理 **57**, 841 (2015).
 [15] 孫媛：情報の科学と技術 **70**, 255 (2020).
 [16] Y. Koide *et al.*, Phys. Rev. Lett. **72** 3662 (1994).
 [17] JT-60SA Research Unit. "JT-60SA Research Plan Version 4.0" (2018).

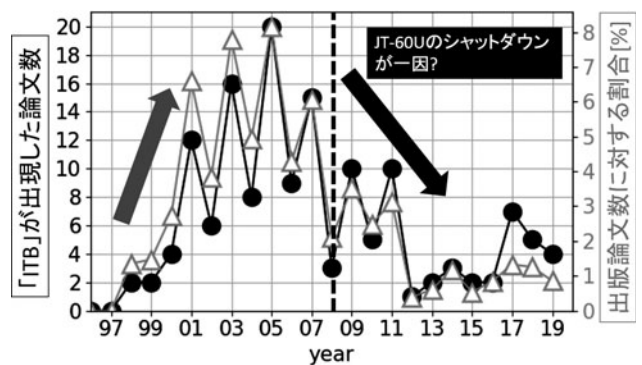
付録.1 中・大型装置名がアブストラクトに出現する論文数の年次推移

図A-1に磁場閉じ込めプラズマ実験装置のうち中・大型装置 DIII-D, JET, LHD, EAST, KSTAR, Wendelstein-7X, JT-60U の各装置をアブストラクトに含む論文数の年次変化を示した。図A-1(a)から読み取れるように各装置名を含む論文数は増加傾向にあることがわかる。増加量に着目すると、EASTおよびWendelstein-7Xの増加量が2015年以降にかけて急激に増加していることがわかる。両装置について計測装置や加熱装置が充実したことで論文数が伸びている可能性がある。出版論文数に対する割合(図A-1(b))に着目すると、DIII-DおよびJETは長年10%~15%と他の装置と比べて多くの関連論文を出版していることが読み取れる。JT-60Uはシャットダウンした装置ではあるが、2009年以前はDIII-D・JETと同等の論文数があるので、参考までに示してある。

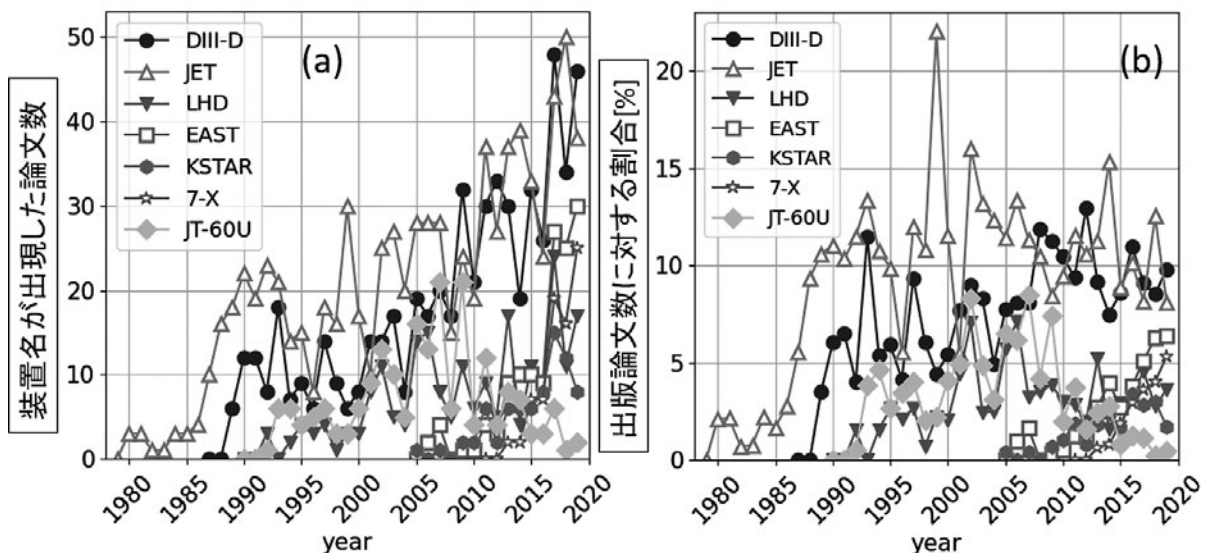
付録.2 キーワードを含む論文数が増加した後に減少した例

図A-2に Internal Transport Barrier (ITB) がアブストラクトに出現した論文数の年次推移を示した。ITBはトーラ

スプラズマにおける閉じ込め改善モードの一つであり、JT-60Uで報告[16]されて以来多くの装置で研究が活発に進められた経緯がある。この様子は図A-2で2007年まで論文数が増加していることで読み取れる。しかしながら一方で、2008年以降は論文数が減少している様子が読み取れる。この原因としてはいくつか考えられるが、ITB研究を積極的に行っていたJT-60Uが2008年にシャットダウンしたことが一因と思われる。実際、JT-60Uの運転期間においては、「ITB」がアブストラクトに出現した論文のうち20%~30%は「JT-60U」と共起しており、その寄与は大きかった。一方で、ITB研究についてはJT-60SAのresearch plan[17]でも言及されており、今後の研究で論文数が再度増加に転じる可能性がある。



図A-2 「ITB」がアブストラクトに出現した論文数(○)と年当たりの出版論文数に対する割合(△)の年次推移。JT-60Uがシャットダウンした2008年に縦線を引いた。



図A-1 (a)各装置名(DIII-D, JET, LHD, EAST, KSTAR, Wendelstein-7X, JT-60U)がアブストラクトに出現した論文数の年次推移。(b)年当たりの出版論文数に対する割合の年次推移。