九州大学学術情報リポジトリ
Kyushu University Institutional Repository

検索事例に基づく言語情報コーパスの検証

恒川, 元行
九州大学言語文化部

https://doi.org/10.15017/5492

出版情報：言語文化論究. 言語情報特集号, pp.67-76, 1999-10-31. 九州大学言語文化部
バージョン：
権利関係：

検索事例に基づく言語情報コーパスの検証

恒 川 元 行

1．　本稿の目的

動詞 schreiben には、[postalisch] korrespondieren の意味で用いられ、通信相手が明示される場合、その通信相手が、例文(1a)のように３格で実現される場合（以下、schreiben + D）と、(1b)のように an＋４格で実現される場合（以下、schreiben an + A）の２通りがある。

(1) a. Du hast *deinen Eltern* lange nicht geschrieben. (DU)
　　b. Du hast *an deine Eltern* lange nicht geschrieben. (DU)

Latzel(1985)（以下、Latzel）は、1979 年時点で出力可能であった範囲で２つの現代ドイツ語コーパス、すなわち Mannheimer Corpus zur geschriebenen deutschen Sprache（以下、ＭＫ１）、Freiburger Corpus der gesprochenen deutschen Standardsprache（以下、ＦＫ）を検索し、そこから得られた 132 例に基づいて、上述の競合関係（３格か an＋４格か）について数量的、意味的、また統語的分析を行っている（第２章を参照）。

また、堀口・池内・恒川(1995)、池内・恒川・堀口(1996)、池内・恒川・堀口(1997)、堀口・池内・恒川(1997)、恒川・堀口・池内(1998)は、自作の現代ドイツ語コーパス（以下、ＨＩＴコーパス）[1]を利用し、そこから得られた事例に基づいて、おもに動詞の競合的用法を分析している。

このような研究に用いられるコンピュータ・コーパスは、分析に必要な事例を比較的容易に収集することが可能であり、ドイツ語を母語としない研究者にもドイツ語の実体に迫ることを可能にする可能性を秘めている。[2]しかし、それぞれの研究において引き出された結論が有効なものであるかどうかは、当該のコーパスがドイツ語の言語事実を正しく反映しているかどうかに依存している。

何がドイツ語の「正しい言語事実」であるのか、また「そもそも言語事実とは何なのか」（在間 1998）という問には容易には答えることができないとしても、あるコーパスの信頼性は、そこから得られたデータを別なコーパスのデータと比較することで、経験的に検証することが可能である。すなわち、当該のコーパスの信頼性は、そのデータが、別なコーパスを利用して得られたデータと量的・質的にほぼ対応しているならば、少なくともそのデータに関しては証明されたと考えることができる。

本稿の目的は、このような意味で、Latzel のデータをＭＫ１およびＨＩＴコーパスデータと比較し、コーパス検証のささやかな第一歩とすることである。ＭＫ１は、Latzel でも利用されているが、ＦＫ分を除いたデータとするため、今回別個に検索し改めて事例を収集した。

２． Latzel のデータと分析

Latzel は、得られた 132 例を、３格ないし an＋４格で実現される Adressat がどのような形式・内容のものかという観点から、以下のように集計している（Latzel は事例数は挙げず、パーセントで表示している)。

表１：Adressat の形式・内容

	schreiben + D	schreiben an + A
Personalpron.	75.35%	8.48%
Reziprokpron.	2.73%	0
Namen	6.85%	30.52%
Titel(mit Namen)	4.12%	25.42%
Institutionen	1.37%	20.33%
"Freund" "Vater" etc.	9.58%	11.86%
"Adresse"	0	3.39%
"Front"	0	0
	100.00%	100.00%

この集計から、Latzel は、１）schreiben＋D あるいは schreiben an＋A のうち一方のみが可能な場合、また２）両方が可能だがどちらか一方が優先的に選択される場合の２つのケースがあることを指摘し、おおよそ以下のようにまとめている。

２．１　一方のみが可能

２．１．１　schreiben an＋A のみが可能

ａ）Adressat が 'Adresse' の場合

Ich schrieb freundliche Briefe *an sechs Adressen*.

ｂ）Adressat が Institutionen の場合

Institutionen は通信内容の「読み手」ではありえず、「受取場所」にすぎない。本来の「読み手」は、その構成メンバーの誰かである。

Ich hatte *an die Sendung 'Zwischen Hamburg und Haiti'* geschrieben.

Da schrieb Ute F. ihren Kummer *an 'Bild'*.

ｃ）Adressat が仲介者であり、本来の受け取り人ではない場合

Schreiben Sie *an mich,* ich werde den Brief weiterleiten.

ｄ）無冠詞の名前の場合で、主語と混同される恐れがある場合（電報文でも同様）

An Maria Theresia schrieb Josef: "Ihr Schwiegersohn Louis ... ist schlecht erzogen ..."
Brief *an Hanna* geschrieben.

ｅ）間接疑問文を伴い、schreiben が mitteilen でパラフレーズできない場合
Ich hatte *an den Rundfunk* geschrieben, ob die überhaupt an einem solchen Bericht interessiert seien.

２．１．２　**schreiben + D のみが可能**
ａ）相互代名詞の場合
Und geschrieben hatten wir *uns* nie.

ｂ）「読み手」である人物に、通信内容を示す名詞が続く場合
Damals haben wir *dir* nicht die volle Wahrheit geschrieben.

ｃ）schreiben von の場合
Du mußt *deinen Eltern* von dem Treffen schreiben.

２．２　一方がより一般的
２．２．１　**schreiben an + A の方が一般的**
ａ）Adressat が（冠詞・語尾変化のない）名前の場合（特に、４格目的語を伴わない場合）
Annegret schrieb *an Julian* in der Stadt.

ｂ）Adressat が肩書き・称号（＝役職者）の場合
　役職者はしばしば宛先でしかなく、通信は別人によって処理される。
So schrieb der Jesuit Horrion *an den Fürstbischof Dietrich von Fürstenberg*.

ｃ）Adressat が、文域の最後尾に置かれる場合
＊Adressat が、情報構造上、通信手段・内容よりも重要である場合
Weil ich weiß, daß Sie zuverlässig sind, habe ich *an Sie* geschrieben.

＊枠外配置、Isolierung の場合
Du brauchst nicht viel zu schreiben ... Einen Brief, einen kleinen. *An ein unbekanntes Mädchen*.

＊関係文が後続する場合
Er schrieb einen Brief *an General Washington*, dem er sich tief verbunden fühlte.

d）schreiben が、sich mit einer Bitte, Aufforderung, Frage an jn. wenden でパラフレーズが可能な場合（下記 2.2.2 c と比較）。
Ich schrieb *an Haußmann,* er möge sich vom Juristischen her um die Sache kümmern.

２．２．２ schreiben + D の方が一般的

a）Adressat が、強勢のない代名詞である場合
Seine Verlobte schrieb *ihm* eine Postkarte.

b）文の中域で、Adressat が通信内容に先行する場合
Der vorgedruckte Absender sagte mir, daß da ein Dr. Erich Werner *der Schwester Dorothea* einen Brief geschrieben hatte.

c）schreiben が mitteilen でパラフレーズ可能な場合（上記 2.2.1 d と比較）。
Schreiben Sie *uns,* was Ihrer Meinung nach in der kathorischen Kirche anders werden müßte.

３．データの比較

以下では、Latzel のデータ（schreiben + D、schreiben an + A に分けて再掲）、ＭＫ１のデータ、ＨＩＴコーパスのデータを比較可能なように並べて提示した。

３．１ Adressat の実現形式

表２－１

	Latzel	MK1	HIT
schreiben + D	73(55.3%)	61(60.4%)	107(64.5%)
schreiben an + A	59(44.7%)	40(39.6%)	59(35.5%)
	132	101	166

ＭＫ１は約 15.3 メガバイト、ＨＩＴコーパスは約 19.5 メガバイトの規模である。コーパス毎の事例数を比較しやすくするため、一定規模（10 メガバイト）あたりの事例数を出したのが、以下の表２－２である。得られた事例の割合は、ＨＩＴコーパスの方が大きいことがわかる。

表２－２

	MK1	HIT
10MB あたりの事例数	66	85

3．2　Adressat の形式・内容

表３－１：schreiben + D の場合（カッコ内は事例数、以下同じ）

	Latzel	MK1	HIT
Personalpron.	75.35%	67.2%(41)	57.9%(62)
Interrogativpron.	0	0	1.9%(2)
Relativpron.	0	0	0.9%(1)
Reflexivpron.	0	0	0.9%(1)
Reziprokpron.	2.73%	3.3%(2)	3.7%(4)
Indefinitpron.	0	1.6%(1)	1.9%(2)
Demonstrativpron.	0	0	0.9%(1)
Namen	6.85%	8.2%(5)	9.3%(10)
Titel(mit Namen)	4.12%	4.9%(3)	8.4%(9)
Institutionen	1.37%	6.6%(4)	0.9%(1)
"Freund" "Vater" etc.	9.58%	8.2%(5)	13.1%(14)
"Adresse"	0	0	0
"Front"	0	0	0
	(73)	(61)	(107)

表３－２：schreiben an + A の場合

	Latzel	MK1	HIT
Personalpron.	8.48%	5.0%(2)	3.4%(2)
Interrogativpron.	0	0	1.7%(1)
Relativpron.	0	2.5%(1)	6.8%(4)
Reflexivpron.	0	0	0
Reziprokpron.	0	0	0
Indefinitpron.	0	0	0
Demonstrativpron.	0	0	0
Namen	30.52%	35.0%(14)	25.4%(15)
Titel(mit Namen)	25.42	32.5%(13)	16.9%(10)
Institutionen	20.33	17.5%(7)	10.2%(6)
"Freund" "Vater" etc.	11.86%	5.0%(2)	32.2%(19)
"Adresse"	3.39%	2.5%(1)	1.7%(1)
"Front"	0	0	1.7%(1)[3]
	(59)	(40)	(59)

３．３　文の中域での語順

以下の数値は、それぞれの事例中、文の中域において Adressat（＝D／an ＋ A）と４格目的語（＝A）が現れている場合、その語順を調査したものである。D／an ＋ A →A は「Adressat → ４格目的語」の順を、 A→D／an＋Aは「４格目的語 → Adressat」の順を示す。

表４－１ ：schreiben＋D の場合

	Latzel	MK1	HIT
D→A	100.00%	86.4%(19)	96.0%(24)
A→D	0.00%	13.6%(3)	0.4%(1)

ここでの A→D の語順は、４例とも４格目的語が先行文脈を受ける代名詞 es の場合である。この４例を除けば、Latzel の集計と同じく、すべて D→A の語順である。

表４－２ ：schreiben an＋A の場合

	Latzel	MK1	HIT
an+A→A	29.41%	25.0%(4)	15.8%(3)
A→an+A	70.59%	75.0%(12)	84.2%(16)

４．考察とまとめ

第２章で紹介した Latzel の分析は、コーパスの量的・質的信頼性を前提として行われている。ここでは、第３章のデータを考え合わせた場合、この点についてどのように考えられるかを、簡単に検証してみたい。

まず、量的側面から見た場合、第３章に示した数値には、Latzel、ＭＫ１、ＨＩＴコーパスにおける個々それぞれの違いにも関わらず、schreiben＋D と schreiben an＋A の競合関係における、共通する一定の傾向を確認することができる。以下の表は、表３－１および表３－２の各項目を、代名詞類と名詞類にまとめて再集計したものである。

表５－１ ：schreiben＋D の場合（表３－１参照）

	Latzel	MK1	HIT
代名詞	78.68%	72.1%(44)	68.2%(73)
名詞	21.92%	27.9%(17)	31.8%(34)
	(73)	(61)	(107)

表５－２：schreiben an + A の場合（表３－２参照）

	Latzel	MK1	HIT
代名詞	8.48%	7.5%(3)	11.9%(7)
名詞	91.52%	92.5%(37)	88.1%(52)
	(59)	(40)	(59)

この表では、表３－１、表３－２において確認される分布の傾向が、さらに明確になっている。すなわち、いずれのコーパスについても、その分布に schreiben + D における代名詞への集中（特に人称代名詞：75.35%／67.2%／57.9%）、および schreiben an + A における具体的な名詞類への集中が認められる。また、表４－１、表４－２に挙げた語順に関するデータも、共通する一定の傾向を示している。このような数値を見る限り、これらのデータには、schreiben + D および schreiben an + A の競合関係における、ある種の「言語事実」が反映されていると考えなければならない。この意味で、これらのデータ、またコーパスには、十分に信頼性を認めることができると思われる。[4)]

以上に対し、質的側面から見た場合、コーパスの信頼性はそれほど明白ではない。質的というのは、当該のコーパスがある用法の全体像を解明するために必要十分な事例を含んでいるかどうか、という側面である。Latzel は第２章で紹介した schreiben の分析において、ＭＫ１とＦＫからの事例に、その他からの事例を加えて競合関係を論じているが、このことは、Latzel が得られた事例数（132 例）を必ずしも十分ではないと考えたことを示している。

さらには、必要な事例が収集事例中に含まれていても、確実な結論を引き出すには事例数が少なすぎるという場合もある。Latzel の分析には、このような少ない事例に基づいて引き出されたと思われる結論が含まれており、以下具体的に２点を例示しておきたい。

その第１点は、Institutionen の項目に関してである。Latzel は、Institutionen は schreiben an + A にのみ可能としており（2.1.1. b 参照）、「注５」の中で、schreiben + D に現れる唯一の事例 Werk（以下の例文(2)）は、作者（Uwe Johnson）の文体の問題であり、例外と見なすべきものであると述べている。

(2) ¥mk1¥mk1abz.txt 302:[...] , zwei Wochen lang hatte Achim faul gelegen mit den Beinen auf der Balustrade und *dem Werk* Briefe geschrieben wegen möglicher Verbesserungen an der Maschine , die ihm nach diesem Sommer tiefer in den Gliedern saß als das westdeutsche Fabrikat .

schreiben + D の場合の Institutionen の例は、表３－１、表３－２から明らかなように、schreiben an + A の場合に比べ少ないのは事実である（ＭＫ１、ＨＩＴコーパスの合計で、前者が５例、後者が１３例）。しかし、上述のような結論を維持するため、当該の Werk 以外の４例についても、同様に例外として扱うことができるかどうかには疑問が残る。以下、

ＭＫ１からの１例を例示する。

(3) ¥mk1¥mk1aeq.txt 259:er schrieb *dem britischen Außenministerium* , daß die Stämme dort sich gern unter den Schutz Großbritanniens stellen wollten .

第２点は、間接話法文を伴う場合、schreiben が mitteilen の意味であれば schreiben + D が、また sich mit einer Bitte, Aufforderung, Frage an jn. wenden の意味であれば schreiben an + A が優先的に選択されるという指摘（2.2.1 d、2.2.2. c 参照）である。この点に関しても、すべてが明快であるわけではない。なぜならば、mitteilen にパラフレーズ可能な例は比較的数が多いのに対し、後者に該当する事例数は限られているからである。表６－１および表６－２は、間接話法文を伴う schreiben が上述のいずれの意味に解されるかを調べたものである。一番左の列は、Latzel 自身が「注８」に挙げている数値である。

表６－１ ：schreiben + D の場合

	Latzel	MK1	HIT
Mitteilung	90.00%(9)	100.0%(11)	83.3%(15)
Aufforderung	10.00%(1)	0	16.7%(3)
Anfrage	0	0	0
	100.00%(10)	100.0%(11)	100.0%(18)

表６－２ ：schreiben an + A の場合

	Latzel	MK1	HIT
Mitteilung	37.50%(3)	80.0%(4)	75.0%(3)
Aufforderung	50.00%(4)	20.0%(1)	25.0%(1)
Anfrage	12.50%(1)	0	0
	100.00%(8)	100.0%(5)	100.0%(4)

表６－２から明らかなように、Latzel は上述のような結論を、mitteilen の意味と解される事例の３例に対し、後者の場合に該当する５例に基づいて引き出していると思われる。３例対５例がどれほどの有効性を持つのか疑わしいだけでなく、ＭＫ１、ＨＩＴコーパスのデータでは逆に前者の例が多く、これによる支持も得られない。

ここに取り上げた２例は、いずれも事例数が限られているため、それに基づいた結論が疑われることになった。このようなマイナーな事例について確実なことを主張するためには、さらに事例が必要である。しかし、マイナーな用法であればあるほど、たとえコーパスの規模を拡大し収集事例を増やしたとしても、当該の用法自体の事例が増えることは期待しにくいと思われる。逆に言えば、このことは、そもそも質的な分析を行おうとする場

合、コーパスから得られる事例だけに頼ることが正しいのかという疑問にもつながっている。ある用法の「全体像」は、冒頭に述べたドイツ語の「実体」という概念同様仮定のものでしかなく、検索事例を土台にしつつ、最終的には理論的にしか究明できないのかもしれない。この点から考えれば、Latzel が分析に際し、コーパス以外からの事例を加えて議論を展開したことには、合理的な意味があったと見ることができる。

本稿では、比較した３つのコーパスのいずれについても、量的には信頼性を示す数値が得られたこと、しかし当該の競合的用法の質的な分析を行うためには、得られた事例が必ずしも必要十分とは言えない面があることを述べた。今後は、Latzel を再検討し、ＭＫ１の 101 例、ＨＩＴコーパスの 166 例（合計すれば、Latzel のちょうど２倍の事例数となる）と対比し、また各種の先行研究文献にもあたることで、データおよびコーパスの質的側面について考察を深めていきたい。

注

１）フィクション、ノンフィクション、新聞・雑誌の３種類のテキストから構成された現代ドイツ語・書き言葉のコーパス（1994～95 年作成)。規模約 19.5 メガバイト。具体的な構成内容については、池内・恒川・堀口(1997)巻末のリストを参照。

２）在間(1998) は、コンピュータ・コーパスを利用した研究について、「文字通りの意味で、ドイツ語全般を分析対象にすることは不可能ですが、実感的には、現在、ドイツ語全般を分析対象にすると言っても差し支えのないほどの段階にあると思います」と述べている。

３）これは、以下の例である。

¥hit¥r¥ehe.bun 398:Aktion >Deutsche Mädel schreiben *an die Front*<.

これは、an + A が通信の「読み手」ではなく、本来「受け取り手／受け取り場所」（すなわち「方向」）を表示するものであるという、Latzel の「注７」における指摘を支持する事例である。

４）ここに、量的信頼性を示すデータを、もう一つ提示しておきたい（恒川 1998)。以下は、「信用する・あてにする」の意味での動詞 vertrauen の事例数と、それぞれ 10 メガバイトあたりの事例数を示したものである。コーパス規模に応じて事例の総数はさまざまであるが、10 メガバイトあたりの数値は Spiegel コーパス (36/1995～21/1997 の 82 号分；福教大・堀口里志氏作成）を除き、極めて似ている（Spiegel の場合も、事例数が他よりも多いのであり、コーパスの信頼性に問題があるわけではない)。このことから、一定規模のコーパスには、ほぼ同じ割合で事例を期待できることがわかる。

	MK1	HIT	Spiegel	Mann	全体
コーパス規模	15.3	19.5	13	26	73.8MB
事例数	50	64	59	89	262
10MB あたり	33	33	45	34	36

参考文献

池内宣夫・恒川元行・堀口里志：beginnen の一用法における４格と mit＋３格の競合について。「大分大学工学部研究報告」第 33 号、1996 年。

池内宣夫・恒川元行・堀口里志：anrufen における４格と副詞規定の交替について。「大分大学工学部研究報告」第 35 号、1997 年。

Duden. Deutsches Universalwörterbuch. 3., neu bearbeitete und erweiterte Auflage. Mannheim, 1996. (= DU)

堀口里志・池内宣夫・恒川元行：「不定詞句＋helfen」構文における zu の有無について。日本独文学会西日本支部「西日本ドイツ文学」第 7 号、1995 年。

堀口里志・池内宣夫・恒川元行：数量名詞と定動詞の数の一致について。「福岡教育大学紀要」第 46 号第 1 分冊、1997 年。

Latzel, Sigbert: *jm. schreiben - an jn. schreiben*. in: Zielsprache Deutsch. Jg. 16. Heft 3. 1985.

恒川元行：コーパスを用いた語法論的結合分析。日本独文学会 1998 年度春季研究発表会（中央大学）・シンポジウム「コーパスを用いた新しいドイツ語研究の方向性」における口頭発表原稿（未刊行）。1998 年。

恒川元行・堀口里志・池内宣夫：テキストデータベース・コーパスの構築と利用の試み。九州大学言語文化部「独仏文学研究」第 45 号、1995 年。

恒川元行・堀口里志・池内宣夫：vertrauen における３格補足語と auf＋４格補足語の競合について。九州大学言語文化部「独仏文学研究」第 48 号、1998 年。

在間進：コーパスを用いた新しいドイツ語研究の方向性。日本独文学会 1998 年度春季研究発表会（中央大学）・シンポジウム「コーパスを用いた新しいドイツ語研究の方向性」における口頭発表原稿（未刊行）。1998 年。