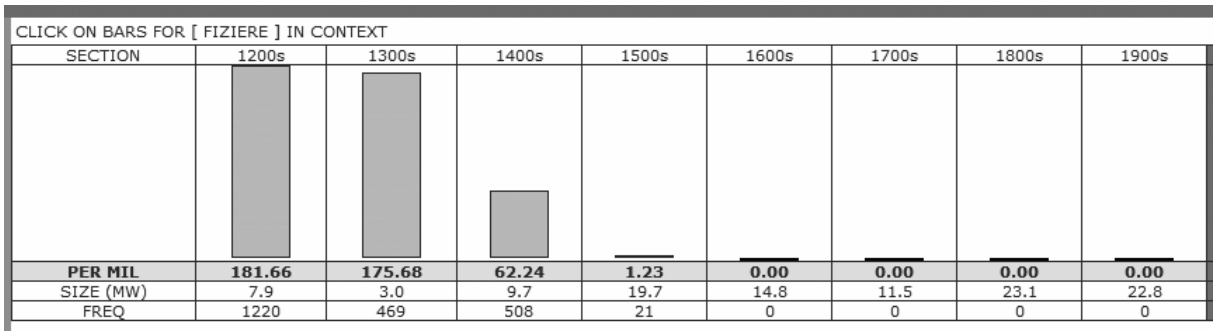
Einheit 10, Seite 219
Lösungen zu den Aufgaben

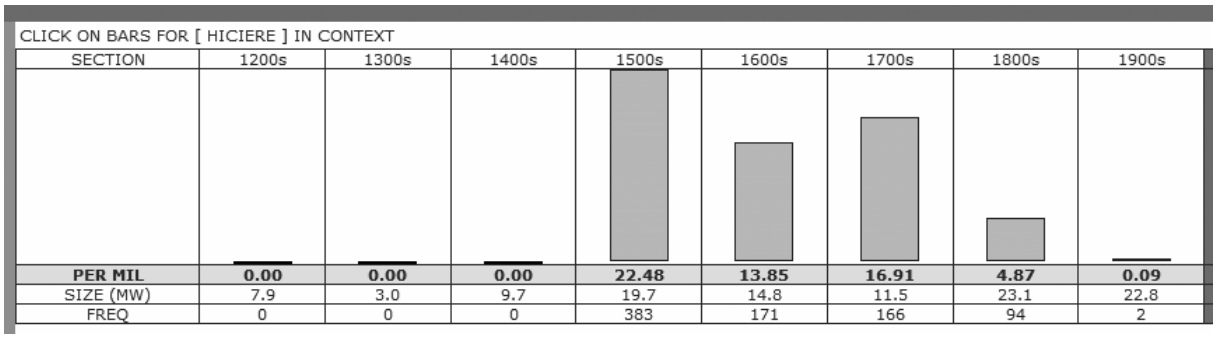
1.
Unter dem Beobachterparadoxon, das erstmals prägnant vom amerikanischen Pionier der Soziolinguistik William Labov im Zusammenhang mit einer Analyse zur Lautvariation in New York thematisiert wurde, versteht man, dass der (Sozio-)Linguistik beobachten will, wie sich seine InformantInnen sprachlich verhalten, wenn sie sich nicht beobachtet fühlen; in Labovs Worten: „the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.“ Dahinter steht die allgemein in der empirischen Sozialwissenschaft gemachte Erfahrung, dass die Situation der Beobachtung bzw. des Beobachtet-Werdens die Verhaltensweisen von Menschen gravierend beeinflusst; die bloße Anwesenheit von Forschern etwa kann dazu führen, dass die Angehörigen der untersuchten Gruppe ihr Verhalten ändern (in der Soziologie ist dies als Hawthorne-Effekt bekannt), und zwar häufig in der Form, dass sie sich der (vermuteten) Erwartungshaltung der beobachtenden Forscher anpassen, sich also so verhalten (wollen), wie sie meinen, dass die Forscher es von ihnen erwarten. Dies führt natürlich zu verzerrten und damit nicht mehr repräsentativen, ggf. sogar ‚falschen’ Ergebnissen. Um diesen Beeinflussungs- und Verzerrungseffekt so gering wie möglich zu halten, werden unterschiedliche Beobachtungs- und Befragungsstrategien vorgeschlagen: Labov selbst hat zur Umgehung einer Beeinflussung seiner InformantInnen das anonyme Schnell-Interview angewandt; auch wird vorgeschlagen, für Befragungen und Datenerhebungen wenn immer möglich solche BefragerInnen zu wählen, die derselben sozialen Gruppe bzw. denselben sozialen Netzen angehören wie die InformantInnen. Konkret bei linguistischen Feldforschungsvorhaben, bei denen es ja um sprachliche Attitüden oder sprachliche Produktion der InformantInnen geht, besteht eine oft angewandte Möglichkeit, das Beobachter-Paradoxon zu minimieren, darin, dass man von der sprachlichen Stoßrichtung der Datenerhebung ablenkt. Anstelle z.B. Gewährsleuten zu sagen, dass man sich für ihre Sprechweise, ihren Dialekt oder ihre Kommunikationsformen interessiert, kann der Linguist ein ethnographisches, biographisches oder historisches Forschungsanliegen vorspiegeln. Dies ist insofern wichtig, als gerade das Sprechen über (ihre) Sprache für viele Menschen sehr unangenehm ist und sie im Glauben, „schlecht“ zu sprechen, leicht zu unsicherem und keinesfalls authentischem kommunikativem Verhalten tendieren. Das Vorspiegeln eines anderen Forschungsinteresses als dem eigentlichen (linguistischen) kann allerdings mit ethischen Grundsätzen der empirischen Sozialforschung in Konflikt geraten, denen zufolge man den InformantInnen gegenüber aufrichtig sein soll.

2.
Hinsichtlich der sprachhistorischen Entwicklung zeigt sich, dass die Form des Subjunktivs Futur *fiziere* in den Texten ab dem 16. Jahrhundert von der Form *hiciere* ersetzt wird; dies wird besonders anschaulich, wenn man die Verteilung beider Formen im *Corpus del español* von Mark Davies gegenüberstellt, wie auf den nachfolgenden beiden Abbildungen geschehen:



SECTION	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
PER MIL	181.66	175.68	62.24	1.23	0.00	0.00	0.00	0.00
SIZE (MW)	7.9	3.0	9.7	19.7	14.8	11.5	23.1	22.8
FREQ	1220	469	508	21	0	0	0	0

Abb. 1: *fiziere* im *Corpus del español* von Mark Davies (www.corpusdelespanol.org).

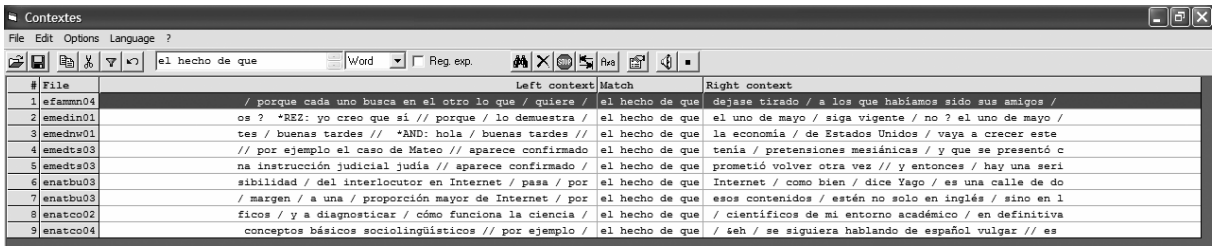


SECTION	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s
PER MIL	0.00	0.00	0.00	22.48	13.85	16.91	4.87	0.09
SIZE (MW)	7.9	3.0	9.7	19.7	14.8	11.5	23.1	22.8
FREQ	0	0	0	383	171	166	94	2

Abb. 2: *hiziere* im *Corpus del español* von Mark Davies (www.corpusdelespanol.org).

Diesem Korpus zufolge fällt die Ablösung der Form *fiziere* durch *hiciere* auf die Zeit zwischen dem 15. und dem 16. Jahrhundert. Diese Veränderung kann auf einen tatsächlich vollzogenen Lautwandel *f-* > *h-* zurückzuführen sein (s. Abschn. 12.11), aber auch auf veränderte Schreibgewohnheiten der Schreiber, die ab dem 16. Jh. wenig stark etymologisierend, also unter Berücksichtigung der lateinischen Basis in der Graphie schrieben. Auffällig ist weiterhin, dass die Formen *fiziere* / *hiciere* insgesamt – von einem leichten Anstieg im 18. Jh. abgesehen – kontinuierlich in ihrer Frequenz zurückgehen: die Frequenz von *hiciere* im 16. Jh. beträgt nur mehr ein Drittel der Frequenz von *fiziere* im vorhergehenden Jahrhundert, das dort ebenfalls schon sehr viel weniger häufig vorkam als in den Korpustexten des 14. Jh. Die Futur-Formen des Subjunktivs werden also im Laufe der Sprachgeschichte obsolet und sind im heutigen Spanisch so gut wie verschwunden.
Bei der Arbeit mit den beiden Korpora *Corpus del español* und CORDE zeigt sich, dass beide den Zugang zu einer so großen Datenmenge ermöglichen, dass beide Datenbanken in die Kategorie der Megakorpora eingeordnet werden können; das CORDE ist mit 250 Mio. Wörtern in etwa doppelt so umfangreich wie das Corpus del español. Darüberhinaus ist die Bandbreite der integrierten Textsorten bei CORDE deutlich größer als beim *Corpus del español*, das weniger kurze vereinzelte Texte als vielmehr umfangreiche Texte – mit ihren jeweiligen, der Texttradition und der Autoren- bzw. Schreiberhand zuzuschreibenden Eigenheiten – bevorzugt. Insofern ist die Datenbasis des CORDE repräsentativer, auch bietet sie eine bessere Auswahloption bei den Textsorten und ihrer geographischen Herkunft, was eine differenziertere Erstellung des Arbeitskorpus erlaubt. Das *Corpus del español* bietet demgegenüber eine sehr bequem zu bedienende Abfragemaske, die hinsichtlich der diachronischen Streuung der Daten gute Auswahlmöglichkeiten gibt und die Ergebnisse der Suche in kürzester Zeit in anschaulicher Form – wie etwa in den o.g. Balkengraphiken – visualisiert. Bei den Textsorten ist dagegen nur eine Grobauswahl möglich, und auch dies nur bei den jüngeren Epochen, da sich in weiter zurückliegenden Jahrhunderten die Auswahl der Texte, die in das Korpus integriert wurden, mitunter als eher unausgeglichen erweist.

3.
Die Modusverteilung nach der Struktur *el hecho de que* ist in der Praxis sehr komplex: gemäß normativer Grammatik steht das Verb des dadurch eingeleiteten Nebensatzes zumeist im Subjunktiv; wird *el hecho de que* jedoch mit einer Präposition angeschlossen (*vuelve por el hecho de que llueve*), steht in der Regel Subjunktiv. Daneben erwähnen die Grammatiken auch Fälle von scheinbar freier Modusvariation. Dazu kommt, dass *el hecho de que* eine Struktur darstellt, nach der häufig *dequeísmo* eintritt, d.h. dass die Sprecher das – normativ geforderte – *de* vor dem Junktor *que* weglassen. In variationslinguistischen Untersuchungen wurde vermutet, dass der *dequeísmo* bei *el hecho (de) que*, der von der normativen Grammatik als ‚Fehler' betrachtet wird, pragmatisch gesteuert ist und eine Modalisierungsstrategie darstellt. Da hier der Modus im untergeordneten Satz mit der Variation zwischen voller *de-que*-Struktur und *dequeísmo*-Struktur zu korrelieren scheint, ist die Modusvariation nach *el hecho (de) que* sehr ausgeprägt. Dies lässt sich mit den im Buch erwähnten Korpora des gesprochenen Spanisch auch bestätigen.
Allerdings ist eine solche Suche mit den gängigen Konkordanzprogrammen und den derzeit verfügbaren Korpora mit gewissen Problemen und großem Aufwand verbunden: zwar lässt sich die Struktur *el hecho (de) que*, die außer eventuellem *dequeísmo* keine formale Variation zeigt, sehr leicht als (Mehr-)Wortsuche in den Korpora auffinden. Wenn man etwa mit dem Konkordanzprogramm *Contextes* die spanischen Teile des C-ORAL-ROM-Korpus (Cresti / Moneglia ed. 2005) durchsucht, erhält man folgende Treffer als KWIC-Liste:



#	File	Left context	Match	Right context
1	efamnn04	/ porque cada uno busca en el otro lo que / quiere /	el hecho de que	dejase tirado / a los que habíamos sido sus amigos /
2	emedin01	os ? *REZ: yo creo que sí // porque / lo demuestra /	el hecho de que	el uno de mayo / siga vigente / no ? el uno de mayo /
3	emednv01	tes / buenas tardes // *AND: hola / buenas tardes //	el hecho de que	la economía / de Estados Unidos / vaya a crecer este
4	emedts03	// por ejemplo el caso de Mateo // aparece confirmado	el hecho de que	tenía / pretensiones mesiánicas / y que se presentó c
5	emedts03	na instrucción judicial judía // aparece confirmado /	el hecho de que	prometió volver otra vez // y entonces / hay una seri
6	enatbu03	sibilidad / del interlocutor en Internet / pasa / por	el hecho de que	Internet / como bien / dice Yago / es una calle de do
7	enatbu03	/ margen / a una / proporción mayor de Internet / por	el hecho de que	esos contenidos / estén no solo en inglés / sino en l
8	enatco02	ficos / y a diagnosticar / cómo funciona la ciencia /	el hecho de que	/ científicos de mi entorno académico / en definitiva
9	enatco04	conceptos básicos sociolingüísticos // por ejemplo /	el hecho de que	/ &eh / se siguiera hablando de español vulgar // es

Abb. 3: *el hecho de que* im Korpus C-ORAL-ROM

Allerdings muss die Modusform des untergeordneten Satzes dann manuell gesucht und bestimmt werden. Eine kombinierte Suche der lexikalischen Form *el hecho de que* und Subjunktiv-/Indikativ-Formen in dem ihr nachfolgenden engeren Kotext wäre zwar möglich, würde aber voraussetzen, dass das Korpus syntaktisch annotiert ist, zu den Verbformen also morphologische Metainformationen gegeben werden, was bei den momentan verfügbaren Korpora des Spanischen nicht der Fall ist. Theoretisch wäre auch eine Kookkurrenz-/Kollokationssuche nach *el hecho de que* + Subjunktiv- bzw. Indikativformen in seiner Umgebung möglich, dann müssten allerdings alle in dieser Position möglichen Indikativ- und Subjunktiv-Formen aller möglichen Verben als lexikalische Elemente eingegeben werden – eine unrealistisches Vorhaben. Damit zeigt sich, dass die computergestützte Korpusanalyse mittels Konkordanzerstellung bei flektierten Elementen, die damit eine große Zahl an Formen aufweisen, sehr schnell an ihre Grenzen stößt, wenn keine morphologisch-syntaktischen Annotationen vorliegen: in diesem Fall kommt man um viel „Handarbeit" nicht herum.

4.
Unter einem offenen Korpus versteht man eine Sammlung von Sprachdaten, die so angelegt ist, dass jederzeit neue Texte / Transkriptionen / Daten hinzugefügt werden können. Von den im Buch vorgestellten Korpora des Spanischen sind z.B. das CREA und das CORDE der Real Academia Española offene Korpora; sie wachsen mit der Verfügbarkeit neuer Texte und bieten sich zur Untersuchung aller möglichen Fragen und Themen an. Im Gegensatz dazu sind geschlossene Korpora solche, bei denen der Umfang der erhobenen und integrierten

Sprachdaten von vornherein festgelegt wird und die zumeist für spezifische Forschungsvorhaben und spezifische Fragestellungen erstellt werden. So sind die *Habla-culta*- und PRESEEA-Korpora geschlossene Korpora: Art und Umfang der Sprachdaten, die dafür erhoben werden, wurden zu Beginn dieser Projekte festgelegt, und sind sie einmal vollständig erhoben, werden keine neuen Sprachdaten hinzugefügt. Geschlossene Korpora versuchen daher, Repräsentativität vor allem über ein Gleichgewicht bestimmter soziolinguistischer und textsortenspezifisch-textueller Kriterien herzustellen, was ein genaues Einhalten dieser einmal festgelegten Kriterien verlangt. Offene Korpora zielen dagegen häufig darauf ab, Repräsentativität über die (möglichst große) Quantität der Sprachdaten zu erlangen.