

Korpuspragmatische Analysen alpinistischer Literatur

Noah Bubenhofer

Mannheim

Joachim Scharloth

Tokyo

A view on linguistic data in the sense of corpus pragmatics leads to new insights in the domains of sociolinguistics, discourse linguistics and cultural studies. In corpus pragmatics, recurrent patterns of language use are interpreted as a result of social actions. Using modern corpus linguistic methods, these recurrent patterns of language use can be extracted in a data driven way from linguistic data. In our paper we show this techniques in the context of a corpus of alpine literature: The yearbooks of the Swiss Alpine Club (SAC). The study shows how "talking about mountains" has changed from 1864 to today.

1. Einleitung

1.1. Korpuspragmatik

Korpuslinguistische Methoden der Datenanalyse sind inzwischen nicht nur in den klassischen Bereichen wie Grammatik oder Lexikographie weit verbreitet, sondern bewähren sich auch als Mittel der quantitativ-empirisch interessierten Forschung in den Bereichen Soziolinguistik, Diskurslinguistik, Kulturanalyse oder bei pragmatischen Fragestellungen. Wir haben bereits dargelegt (Scharloth/Bubenhofer 2011), dass wir die Forschung als korpuspragmatisch verstehen, die wie folgt charakterisiert ist: Die Korpuspragmatik deutet signifikant häufig auftretende sprachliche Muster in Korpora als Ergebnis rekurrenter Sprachhandlungen der Autorinnen und Autoren der im Korpus enthaltenen Texte bzw. der sie autorisierenden Institutionen und Gruppen. Sie geht davon aus, dass sich pragmatische Informationen "im pragmatischen Mehrwert oder Gebrauchswert von Einheiten aller sprachlicher Strukturbereiche" (Feilke 2000, S. 78) zeichenhaft manifestiert. Damit werden pragmatische Spuren an der sprachlichen Oberfläche, in die sich ein Gebrauchswert eingeschrieben hat, als Muster sichtbar (vgl. zu dieser Rehabilitierung der Textoberfläche auch den Sammelband von Feilke/Linke 2009). Diese "Sprachgebrauchsmuster" (Bubenhofer 2009) werden damit als Ergebnis von sprachlich-sozialem Handeln gelesen und gedeutet.

Wir wollen im Folgenden anhand eines Korpus alpinistischer Literatur zeigen, welche Fragestellungen korpuspragmatische Methoden beantworten können. Bei der untersuchten 'alpinistischen Literatur' handelt es sich um die von 1864 bis 2009 erschienenen Publikationen des Schweizer Alpenclubs (dazu mehr weiter unten).

1.2. Korpuslinguistik

Attraktiv an dieser Orientierung an sprachlichen Mustern auf der Textoberfläche sind die methodischen Möglichkeiten, die sich dadurch ergeben. Dank der Verfügbarkeit von großen Korpora und der maschinellen Verarbeitung von

Textdaten wird es möglich, quantitativ zu arbeiten und Algorithmen zu entwickeln, um die Musterhaftigkeit in den Daten induktiv zu entdecken. Digitale Korpora sind somit nicht nur "Belegsammlungen oder Zettelkästen in elektronischer Form", sondern ermöglichen eine eigene "korpuslinguistische Perspektive" (Perkuhn/Belica 2006, S. 2).

Was ist mit dieser 'korpuslinguistischen Perspektive' gemeint? Die oben gemachten Ausführungen deuten es bereits an. Offensichtlich ist das quantitative Vorgehen: Nicht der Einzelbeleg interessiert, sondern die Musterhaftigkeit von Belegen, die als Ergebnis rekurrenter Sprachhandlungen gelesen werden können. Doch es kommt ein weiterer Aspekt hinzu: In Ergänzung zu den klassischen datenbasierten korpuslinguistischen Analysen, die der Überprüfung von Forschungshypothesen dienen ('corpus-based'-Paradigma), ist es fruchtbar, ein induktives Vorgehen der datengeleiteten Analyse ('corpus-driven'-Paradigma) hinzuzuziehen. Dieses Paradigma wird von Tognini-Bonelli (2001, S. 84ff.) vor dem Hintergrund der Arbeiten von Sinclair (1991) expliziert und im deutschen Sprachraum in Arbeiten von (Perkuhn u. a. 2005; Belica/Steyer 2006; Steyer 2004; Bubenhofer 2009) u.a. verbreitet. Statt eine Hypothese mit vorher festgelegten Analysekatogorien zu überprüfen, werden in einem Korpus sämtliche Zeichenkonfigurationen berechnet, die sich bei der Anwendung vorher festgelegter Algorithmen ergeben. Beispielsweise werden alle unterschiedlichen Kovorkommen von Wörtern berechnet. Im Vergleich mit einem Referenzkorpus können nun die Kovorkommen berechnet werden, die im Untersuchungskorpus besonders auffällig – also statistisch gesehen überzufällig häufig – sind. Diese Muster werden im Anschluss kategorisiert. Damit geraten häufig Evidenzen in den Fokus, die entweder quer zu den vorher existierenden Erwartungen stehen und die Grundlage für neue Hypothesen sind, oder im besten Fall sogar solche Evidenzen, die die Bildung neuer interpretativer linguistischer Analysekatogorien nahelegen.

Methodisch gesehen bieten sich für induktive Analysen die Berechnung von n -Grammen an: N -Gramme sind Einheiten einer durch n bestimmten Anzahl aufeinander folgender Wörter (Manning/Schütze 2002, S. 192ff.). Durch kombinatorische Zählverfahren lassen sich in einem Korpus leicht alle vorhandenen n -Gramme berechnen. Zudem können, wenn das Korpus vorgängig entsprechend annotiert wurde, neben den Wortformen die n -Gramme auch aus Wortart-Informationen oder einer Kombination von Wortform und Wortart bestehen. Solche "komplexen n -Gramme" (Scharloth/Bubenhofer 2011; Bubenhofer u. a. 2009) sind abstrakter als reine Wortformen- n -Gramme, da sie unterschiedliche Varianten des gleichen zugrundeliegenden morphosyntaktischen Musters zusammenfassen.

Neben der Berechnung von n -Grammen (und damit auch von Unigrammen, den für ein Korpus typischen Einzelllexemen) verwenden wir aber auch statistische Maße, etwa um die Textkomplexität auszudrücken, quantitative Analysen der verwendeten grammatischen Eigenschaften und bereits vorgängig aufgrund linguistischen Wissens geformte Kategorisierungen des verwendeten Vokabulars. Auf diese Messmethoden gehen wir im Folgenden genauer ein.

1.3. Untersuchungsgegenstand: Jahrbücher des Schweizer Alpenclubs

Wir wollen unsere theoretischen und methodischen Überlegungen an einem sehr spezifischen Korpus explizieren: Das Korpus 'Text + Berg' (Bubenhofer u. a. 2011) umfasst alle von 1864 bis 2009 erschienenen Ausgaben des Jahrbuchs bzw. der Zeitschrift des Schweizer Alpenclubs SAC.

Der Schweizer Alpenclub (SAC) wurde 1863, kurz nach dem British Alpine Club gegründet und entwickelte sich, ähnlich wie die anderen Alpenvereine, von einem kleinen Zirkel von Bergsteigern zu einem großen Dienstleistungsverein für breite Schichten von Bergbegeisterten. Von 1864 bis heute gibt es die kontinuierliche Publikationsreihe 'Jahrbuch des Schweizer Alpenclub', später 'Die Alpen', in der über die Tätigkeiten der Clubmitglieder und andere Themen im Bereich Berge, Bergsport, Natur und Kultur im alpinen Raum berichtet wird.¹ Die Publikationsreihe ist für diachrone Untersuchungen besonders interessant, da sie lückenlos über knapp 150 Jahre besteht und bezüglich der vorkommenden Themen und Textsorten – etwa Tourenberichten, biographischen Artikeln, philosophischen Essays, Protokollen und Finanzberichten des Vereins – relativ homogen ist.

Das Gesamtkorpus umfasst 196 Bände, was einer Menge von 35,8 Mio. laufenden Wortformen entspricht. Die Publikation ist teilweise parallel in französischer Sprache erschienen. Das ganze Korpus steht in einer XML-Version zur Verfügung, in der Artikelgrenzen und Metainformationen erfasst und zudem Wortarten- und Lemma-Informationen annotiert sind.² Die Texte wurden mittels OCR-Software maschinell erkannt, was jedoch kein fehlerfreier Prozess ist. Die Texte wurden vollständig mit halbautomatischen Methoden und teilweise zusätzlich manuell korrigiert, trotzdem ist mit Erkennungsfehlern zu rechnen. Sie sollten bei den großen Textmengen jedoch nicht ins Gewicht fallen.

Für die vorliegende Untersuchung sollen zwei Zeiträume im Fokus stehen: Einerseits die Zeit von 1960 bis 1979, andererseits von 1990 bis 2009. Die Auswahl dieser Perioden ist relativ zufällig, orientiert sich aber an einer früheren Studie (Bubenhofer/Schröter 2010), die die Perioden 1880 bis 1899 und 1930 bis 1949 in den Blick nahm. Damit sind die beiden gewählten Perioden eine gute zeitliche Ergänzung zu den bereits gemachten Analysen.

Die Hypothesen, die anhand des Materials geprüft werden sollen, sind:

1. Lassen sich mit korpuslinguistischen Methoden Sprachgebrauchsmuster identifizieren, die als Ergebnis von sprachlich-sozialem Handeln gelesen werden können?
2. Wir wollen zwei Zeitperioden genauer in den Blick nehmen: 1960 bis 1979 und 1990 bis 2009. Lassen sich über die verwendeten Methoden zeitlich spezifische Sprachgebrauchsmuster, von denen auf unterschiedliches 'Sprechen über Berge' geschlossen werden kann, feststellen?

¹ Die Beiträge zum „Jahrbuch des Schweizer Alpenclub“ und den „Alpen“ betreffen nicht nur die Alpen, sondern auch andere Gebirge.

² Das Korpus wurde mit Hilfe des „TreeTaggers“ (Schmid 1994) und den standardmäßig verfügbaren deutschen Trainingsdaten, die auf dem STTS-Tagset (Schiller u. a. 1995) beruhen, annotiert.

Unser Untersuchungskorpus umfasst nur die deutschen Artikel, die aber teilweise auch anderssprachige Sätze enthalten. Damit ergeben sich die in Tabelle 1 aufgeführten Größenordnungen (im Vergleich dazu sind auch die Untersuchungskorpora der früheren Studie genannt). Wie die Tabelle zeigt, werden gewisse Analysen auch am Gesamtkorpus, das alle deutschsprachigen Artikel von 1864 bis 2009 enthält, vollzogen.

Tabelle 1: Umfang der Korpora in der Übersicht

Name	Periode	Laufende Wortformen ³	Artikel
Korpus A ⁴	1880 bis 1899	3.394.471	1021
Korpus B ⁴	1930 bis 1949	2.795.149	1124
Korpus C	1960 bis 1979	2.607.881	870
Korpus D	1990 bis 2009	3.840.933	3891
Korpus komplett	1864 bis 2009	20.138.876	7385

2. Analysen

2.1. Formale Aspekte der Texte

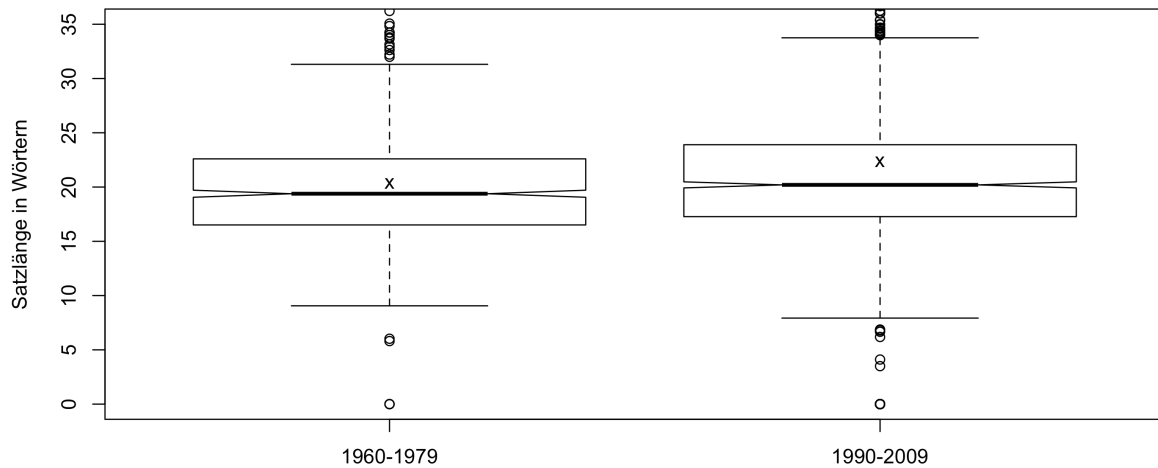
Zunächst werden Analysen zu den formalen Aspekten der Daten vorgenommen. Dabei beziehen wir uns auf das Gesamtkorpus von 1864 bis 2009.

Die durchschnittliche Satzlänge beträgt über alle Jahre hinweg 29.6 Wörter. In den beiden im Fokus stehenden Perioden der 1960er/70er und 1990er/2000er-Jahre liegt die Satzlänge einiges tiefer: Im älteren Zeitraum sind es 20.3 Wörter und im jüngeren Zeitraum 22.4 Wörter. Während die Satzlänge zunächst von 1864 bis in die 1930er-Jahre tendenziell abnimmt, steigt sie in neuerer Zeit wieder leicht an.

³ Die für Korpus A und B genannten Zahlen der laufenden Wortformen differieren leicht von den in Bubenhofer/Schröter (2010) genannten (A dort: 3.463.496; B dort: 2.915.548), da dort die (wenigen) nicht-deutschen Artikel nicht ausgesondert wurden. Um die Korpora aber mit den Korpora C und D vergleichen zu können, werden hier die Zahlen genannt, die sich ergeben, wenn die Korpora auf die gleiche Weise zusammengestellt werden wie die Korpora C und D.

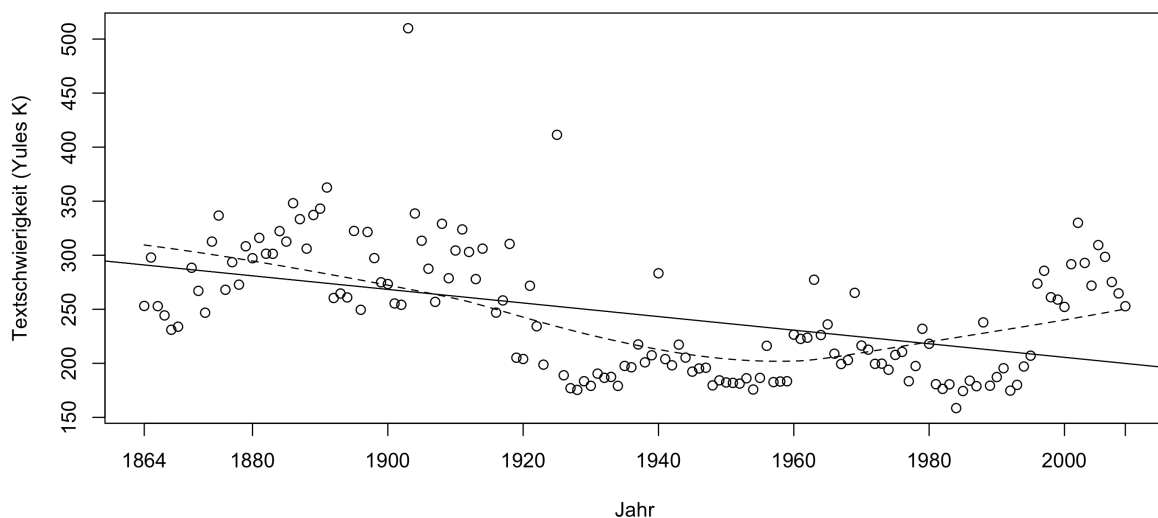
⁴ Korpus A und B wurden in Bubenhofer/Schröter (2010) verwendet.

Abbildung 1: Durchschnittliche Satz­längen in zwei Perioden im Vergleich



Wenn die Satz­länge tendenziell zurückgeht, wie verläuft dazu die Entwicklung des Wortreichtums? Mit Yules K (Yule 1944; Miranda-García/Calle-Martín 2005) steht ein Maß zur Verfügung, das den 'lexikalischen Reichtum' eines Textes messen kann. Je mehr Wortwiederholungen es gibt, desto kleiner ist der Wert und – so die Hypothese – desto einfacher ist der Text zu verstehen. Die Grafik zeigt, dass, ähnlich wie die durchschnittliche Satz­länge, die Textschwierigkeit tendenziell sinkt, wobei ab 1960 wieder ein Anstieg des lexikalischen Reichtums festzustellen ist.

Abbildung 2: Durchschnittliche Textschwierigkeit pro Jahr (Yules K)

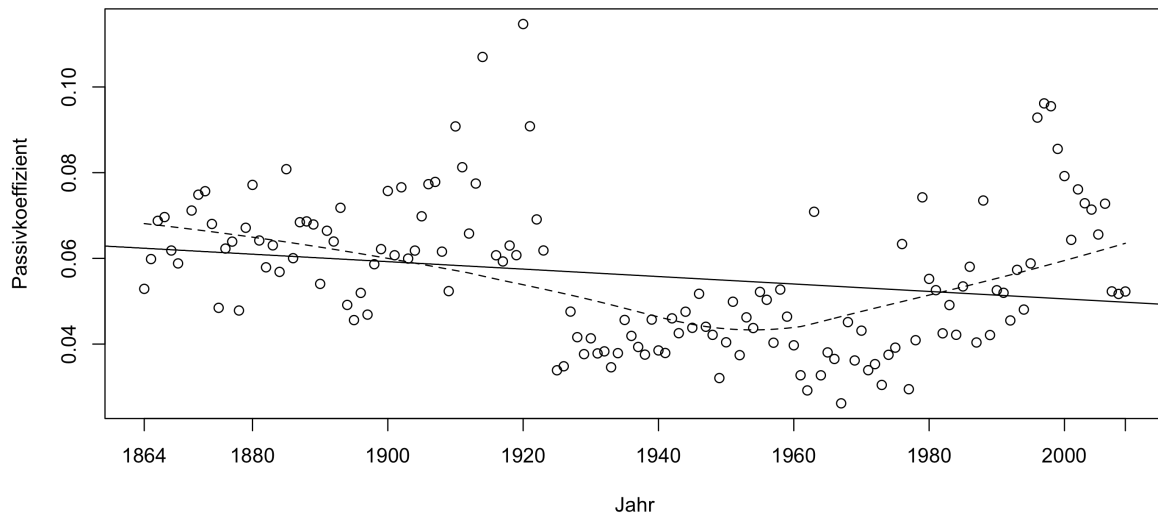


Die Analysen der komplexen n-Gramme weiter unten werden zeigen, dass in neueren Texten vermehrt Passivkonstruktionen vorkommen und die älteren von einem narrativen Charakter geprägt und eine persönliche Erzählperspektive widerspiegeln. Hinweise für diese Hypothese ergeben sich auch aus zwei weiteren Maßen: Dem Passivkoeffizienten und der relativen Anzahl an Intensivierern im Text.

Der Passivkoeffizient beruht auf einer maschinellen Analyse, die die automatische Wortarten-Annotation des TreeTagger (Schmid 1994) weiter auswertet. Wie folgende Grafik zeigt, ist auch hier tendenziell eine Abnahme der

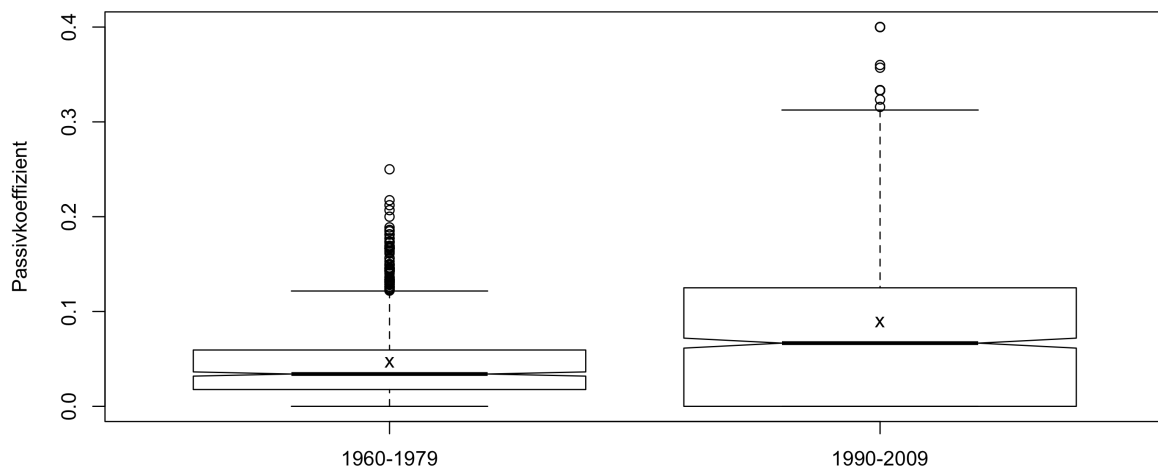
Passivkonstruktionen feststellbar, wobei ab den 1955er-Jahren ein Anstieg zu beobachten ist.

Abbildung 3: Durchschnittlicher Passivkoeffizient pro Jahr



Vergleicht man die beiden Zeitperioden 1960 bis 1979 und 1990 bis 2009 miteinander, ergibt sich ein hochsignifikanter Unterschied der Passivkoeffizienten; in den neuen Daten sind mehr Passivkonstruktionen vorhanden ($t = -13.2002$, $df = 1997.302$, $p < 2.2e-16$).

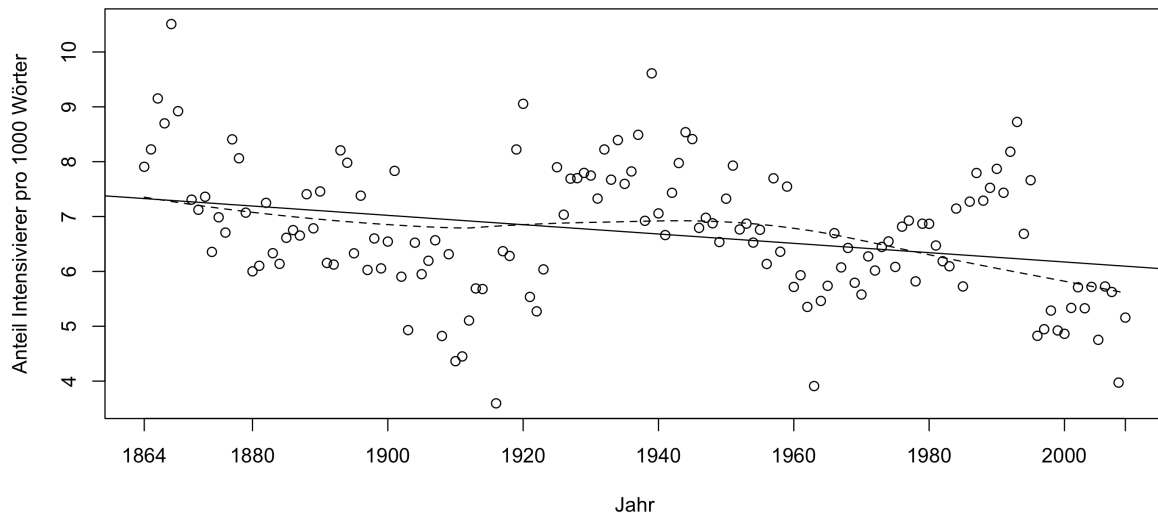
Abbildung 4: Durchschnittlicher Passivkoeffizient in zwei Perioden im Vergleich



Intensivierer (Biedermann 1969; Bierwisch 1987; Os 1989) wie *völlig*, *möglichst*, *kaum* etc. können als Indikatoren für subjektive Einschätzungen bzw. den Grad der Rigorosität, mit der sie gegeben werden, verstanden werden. Besonders interessant sind hier die Intensivierer des “absoluten” und des “extrem hohen Intensivierungsbereichs” (Os 1989, S. 134ff.). Zum absoluten Intensivierungsbereich zählen nach Os Ausdrücke wie *absolut*, *gänzlich*, *komplett* etc., zum extrem hohen Bereich Ausdrücke wie *höchst*, *äußerst* oder *zutiefst*.

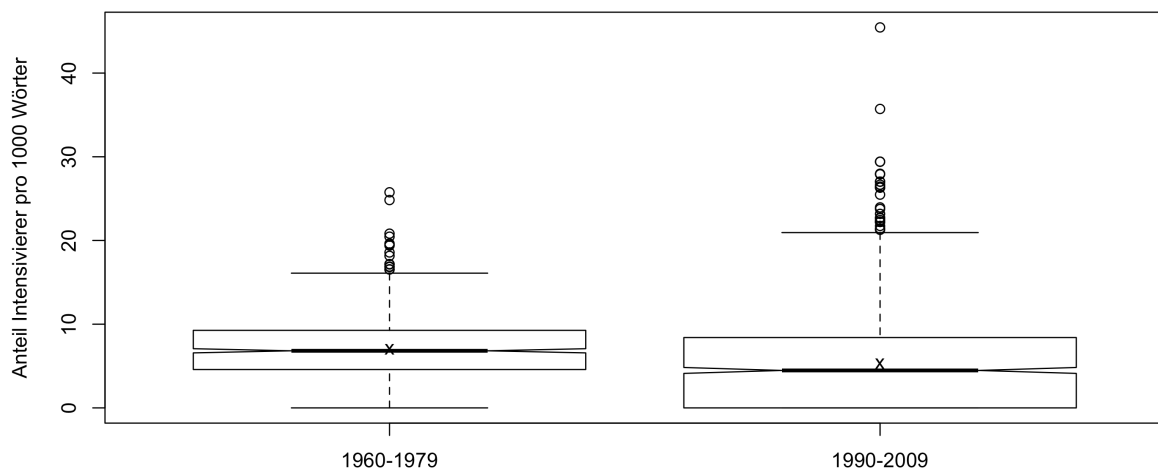
Obwohl die Streuung der verwendeten Intensivierer hoch ist, ist doch eine leichte Abnahme über das Gesamtkorpus zu beobachten:

Abbildung 5: Anteil von Intensivierern im Text



Vergleicht man nur die beiden Perioden 1960 bis 1979 und 1990 bis 2009 miteinander, gibt es in der älteren Periode signifikant mehr Intensivierer als in der jüngeren Periode ($t = 8.9323$, $df = 2258.657$, $p < 2.2e-16$):

Abbildung 6: Durchschnittliche Verwendung von Intensivierern in zwei Perioden im Vergleich



2.2. Schlagwörter

Der einfachste Zugang zu den typischen Lexemen eines Teilkorpus im Vergleich zu den anderen Korpora ist eine Schlagwortanalyse. Damit werden die Lexeme berechnet, die im Untersuchungskorpus statistisch signifikant häufiger als im Referenzkorpus (bestehend aus den restlichen Teilkorpora) auftreten.

In Abbildung 7 sind die Lexeme ersichtlich, die für das ältere Korpus (1960 bis 1979), in Abbildung 8 die Lexeme, die für das jüngere Korpus (1990 bis 2009) typisch sind. Für alle Lexeme gilt, dass der Frequenzunterschied zum jeweiligen Referenzkorpus hoch signifikant ist (LLR, $df=1$, $p < 0.001$). Die Größe des Lexems drückt die statistische Signifikanz aus: Je größer, desto signifikanter ist der Frequenzunterschied.

Schlüsselwörter einen sehr viel professionelleren Alpenclub mit einem umfangreichen Dienstleistungsangebot zu repräsentieren, wo auch das alpine Betätigungsfeld ausdifferenziert wurde und mit dem Wettkampfbereich ein neues Element Eingang findet.

Vergleiche mit den Teilkorpora A und B der Zeitperioden 1880 bis 1899 bzw. 1930 bis 1949 (vgl. Bubenhofer/Schröter 2010) zeigen folgende Parallele: Ähnlich wie im Korpus C (1960 bis 1979) sind in den Texten der 1930er/40er viele emotionale Ausdrücke und Indikatoren für Narration zu finden, wobei dort der Wortschatz der Gefühle und des Seelenlebens noch ausgeprägter ist (*Seele, Erlebnis, Herz, Erleben, Freude, Gefühl, Sehnsucht, Liebe, Glück, Einsamkeit, Mut, Empfinden, Gemüt* etc.). Weiter scheint dort in Lexemen wie *Leben, Schönheit, Wunder, Geheimnis, Menschlein, Romantik, Erfüllung, Unendlichkeit, Gott* und *Glaube* ein religiös-romantisches Verhältnis des Menschen zur Natur anzuklingen. Und auch hier betonen Personenbezeichnungen wie *Kamerad, Freund, Bergkamerad* und *Gefährte* die soziale Verbundenheit und Gemeinschaft der Alpinisten.

2.3. Komplexe n-Gramme

Komplexe n-Gramme sind eine erweiterte Form von n-Grammen, die nicht nur auf einer Folge von Wortformen bestehen, sondern auch aus einer Kombination von Wortformen und Wortart-Informationen bestehen können.⁵ Während ein n-Gramm beispielsweise als Wortformenkette *so verbringen wir* definiert ist, werden bei der Berechnung von komplexen n-Grammen die Wortarten mit einbezogen, so dass eine Reihe von ähnlichen Wortformen-n-Grammen abstrakter als „so – finites Verb – Personalpronomen“ gefasst werden.⁶

Für die Teilkorpora wurden alle komplexen n-Gramme berechnet, die jeweils für das Teilkorpus im Vergleich zu den anderen Teilkorpora typisch sind. Die große Menge an signifikanten n-Grammen wurde anschließend manuell gesichtet und kategorisiert.

Zusammenfassend lässt sich sagen, dass für die Zeit von 1960 bis 1979 Muster typisch sind, die für einen besonderen narrativen Stil stehen, der eine sehr persönliche Erzählperspektive wiedergibt. Denn die Muster weisen viele Personalpronomen und Personennamen, sowie typisches Wortmaterial von Erzählungen wie Konnektoren auf. Diese Beobachtung deckt sich ebenfalls mit den Spezifika von Korpus B (1930 bis 1949), wo ebenfalls auf ähnliche Weise Geschichten erzählt werden (Bubenhofer/Schröter 2010).

Im Gegensatz dazu stehen die Muster des jüngeren Korpus von 1990 bis 2009: Hier herrscht ein deskriptiver Stil vor und es sind mehr Passiv-Konstruktionen zu beobachten als im älteren Korpus.

Im Folgenden werden eine Reihe der auffälligen Muster für das ältere Korpus dargestellt. Aus Platzgründen beschränken wir uns beim jüngeren Korpus auf wenige Beispiele.

⁵ Vgl. das ähnliche Konzept der 'Collostructions' von Stefanowitsch und Gries (2003).

⁶ Zur Berechnung der komplexen n-Gramme wurde ein von der Forschergruppe semtracks entwickeltes Programm „cwb-n-grams“ verwendet, das eine Erweiterung zur Open Corpus Workbench ist (Evert/Team 2010).

2.3.1. 1960er und 1970er-Jahre

Auffällig für das ältere Teilkorpus sind Muster, die in ihren Realisierungen einen narrativen Stil erzeugen. Das Muster „Adverb – finites Verb – Personalpronomen“, das im Korpus hochfrequent ist (5383 Mal im älteren Teilkorpus) und im älteren Korpus hoch signifikant häufiger vorkommt als im neueren (LLR 1667, $p < 0.00001$), steht für Realisierungen wie *so verbringen wir*, *so sehnten wir*, *dann stiegen wir* etc. Das Muster konzentriert sich zudem nicht auf wenige Autoren, sondern kommt bei 489 unterschiedlichen Autoren und in 758 unterschiedlichen Artikeln vor – gestreut über alle 20 Jahre des Teilkorpus.

Für eine ganze Reihe ähnlicher Muster ist zudem gemeinsam, dass alle Personalpronomen enthalten. Die folgenden Tabellen stellen diese Muster mit den jeweiligen Belegen dar.

Die folgenden Muster enthalten neben Personalpronomen Adverbien oder Konjunktionen:

ADV VVFIN PPER ⁷	Frequenzen absolut: 5383 (2688); LLR: 1666.761182; $p < 0$
	So verbringen wir trotz der scharfen Kälte eine letzte Nacht auf dem Gletscher auf 5000 m Höhe. (Band 1964, "Der Pico Unis West und der Tocllaraju", J.-J. Fatton) Dann stiegen wir mit der silbernen Kabine in lautloser Fahrt über die verträumten Weiler Winkelmatten und Blatten, zur Zwischenstation Furi, dann nach dem Umsteigen über Hermettje nach Schwarzsee an den Fuss des Matterhorns hinauf, das da so nahe steht, dass man ehrfürchtig-staunend es mit der Hand zu berühren glaubt. (Band 1962, "Über den Anderson- zum Lauteraargrat", Richard Knecht) Damals umschwirrten uns Fragen, Erwartungen, Hoffnungen (Band 1979, "Die Via Lepontina: eine Route in den Tessiner Alpen", Claudio Abächerli)
	Unterschiedliche Autoren: 489; Artikel: 758; Jahre: 20
./\$, ADV dass/KOUS PPER	Frequenzen absolut: 439 (69); LLR: 421.061016; $p < 0$
	Ich war ziemlich müde vom Aufstieg, so dass wir bis zur Gianettihütte (oder Capanna Badile, 2534 m) fünf Stunden benötigten. (Band 1973, "Ein unbeabsichtigter Rekord", W. Kirstein) Das Terrain war wenig schwierig, so dass ich etwas unaufmerksam kletterte und einen aus dem Riss herausragenden grossen Stein berührte. (Band 1962, "Auf der Spannorthütte", W. Müller-Hill)
ADV VMFIN PPER	Frequenzen absolut: 840 (429); LLR: 246.851027; $p < 0$
	Zuerst mussten wir uns zwar auch hier über eine fast zwei Meter hohe Wächte hinunterschwindeln, dann stob der Schnee gegen das Gesicht, und wir sahen nichts mehr, bis wir sanft in einer Mulde landeten. (Band 1976, "Die Reise zum Gran Sasso d'Italia", Alfred Graber) Doch inzwischen dürften sie im Spital Visp in guter Obhut sein. (Band 1971, "Bergkameradschaft", Ernst Bucher) Dort musste ich einen Wagen nehmen, der mich in richtigem Fieberzustand nach Martigny brachte. (Band 1963, "Eine Bergsteigerkarriere vor 100 Jahren: G. Ad. Koella, 1822-1905", Louis Seylaz) Jetzt müssen wir uns aber beeilen. (Band 1972, "Rosenlauistock-

⁷ Vergleiche zu den Abkürzungen der Wortarten das STTS-Tagset (Schiller u. a. 1995), das auch übers Web einsehbar ist: <http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html> (27. Oktober 2011).

Südwestwand", Andreas Flückiger)
Unterschiedliche Autoren: 304; Artikel: 415; Jahre: 20

./\$. Als/KOUS PPER Frequenzen absolut: 398 (198); LLR: 122.157474; p < 0

Ich kam mir sehr schlau vor und fühlte mich meiner Sache sicher. **Als ich** am Samstagabend die Argentierehütte betrat, war ich beides nicht mehr, dafür aber nass, pudelnass wie noch nie in meinem Leben. (Band 1970, "Die Aiguille Verte", Pierre Vittoz)

Der nun aufgegangene Mond leuchtete uns, während wir über den Gletscher abstiegen. **Als wir** aber auf die riesige Moräne kamen, wurde sein Licht ungenügend, und mehr kugelnd und oft fallend, ohne aber den geringsten Schaden davonzutragen, erreichten wir eine Stunde vor Mitternacht das Basislager. (Band 1960, "Nevado Panta", Geny Steiger)

Unterschiedliche Autoren: 176; Artikel: 231; Jahre: 20

Es gibt jedoch auch Muster, die zwei Personalpronomen in Kombination mit einem finiten Verb enthalten und damit meistens die Verwendung von reflexiven Verben ausdrücken, was eine logische Folge der häufigen Verwendung von Personalpronomen ist.

PPER VVFIN PPER Frequenzen absolut: 1913 (858); LLR: 693.553991; p < 0

Wir amüsieren uns immer sehr bei diesem Anlass. (Band 1964, "Erstbesteigung des Pumo Ri", Ueli Hürlemann)

Es blieben mir noch zehn weitere Tage, um Japan kennenzulernen, dieses farbenreiche, poetische Inselland, weit ab von unserer westlichen Zivilisation, aber voll von Schätzen einer alten Kultur. (Band 1965, "Eine Besteigung des Fujiyama", W. K. Rieben)

Wir trösteten uns mit dem Gedanken, dass wir uns ja auf einer uralten Karawanenstrasse fortbewegten, die einst Europa mit Asien verband. (Band 1972, "Erstbegehung des Ararat-Nordostpfeilers", Arnost Cernik)

Unterschiedliche Autoren: 370; Artikel: 537; Jahre: 20

Ebenso auffällig sind Muster mit Partizipien in Nebensätzen, an die ein Hauptsatz anschließt, was ebenso typische Muster narrativer Texte sind.

VVPP ./\$, VVFIN PPER Frequenzen absolut: 450 (236); LLR: 146.060641; p < 0

Auf dem flachen Gletscher **angelangt, genossen wir** eine lange Rast im milden Sonnenschein. (Band 1966, "Weisse Ostern in den Ötztaler Alpen", Georges Perrin)

Die Ski **geschultert, steigen wir** zu Tale. (Band 1967, "Haute-Route", Lisa Schellenberg-Gensetter)

Mit unseren überschüssigen Kleidern **ausgerüstet, schickte ich** sie unter der Führung von Jean-Antoine Richtung Gipfel. (Band 1965, "Die Besteigung des Cotopaxi", Edward Whymper)

Mit einer Zeltplane **zugedeckt, versuchen wir**, uns an den Berg zu kleben. (Band 1975, "Pilier d'Angle: Bonatti/Gobbi-Führe", Michel Vaucher)

Natürlich finden sich auch narrative Muster ohne Personalpronomen, die aber den für narrative Strukturen typischen Konnektor *und* oder Präpositionen (hier: *nach*) enthalten.

und ADJD VVFIN Frequenzen absolut: 435 (176); LLR: 180.838479; p < 0

Mein Urlaub war ohnehin zu Ende, **und wehmütig nahm** ich Abschied für lange Zeit. (Band 1972, "Die Haute Route - damals", Kurt Wolf)

Kühl **und gelassen greift** die Dämmerung um sich, und aus den Tälern steigt das Dunkel langsam gegen die Gipfel. (Band 1973, "Erinnerung an einen Seilgefährten", K. Scheu)

Grau **und verwaschen wölbte** sich der Himmel über uns. (Band 1971, "Lauchernalp", Alfred Schmocke)
Einzelnen muss jeder Tritt und Griff sorgfältig abgeleuchtet werden, **und vorsichtig schieben** wir uns Meter um Meter höher. (Band 1964, "Meije-Überschreitung", Wolfgang Stefa)

Unterschiedliche Autoren: 208; Artikel: 260; Jahre: 20

./\$. Nach/APPR ADJA NN
Frequenzen absolut: 259 (98); LLR: 130.614103; p < 0

Das Unwetter greift langsam, aber unerbittlich auf den ganzen Wüstenraum über. **Nach vierstündigem Marsch** von der Ostlandstrasse aus haben wir eine Höckerlandschaft gewonnen, die annähernd fünf Kilometer östlich des Eilifsvatn liegt und aus lössartigem Lehm besteht. (Band 1965, "Dettifoss und Selfoss in Island", Hugo Nünlist)

Da sich der Patient passiv verhält, ist dies allein schon ein schönes Stück Arbeit. **Nach verschiedenen Wehlauten** ist es endlich soweit, und Werner stapft tapfer voran, wie weiland Christophorus. (Band 1971, "Bergkameradschaft", Ernst Bucher)

Bald haben wir das Skidepot wieder erreicht. **Nach kurzer Beratung** beschliessen wir, ohne Seil abzufahren; (Band 1969, "Eine Haute-Route für Feinschmecker", Christian Flämi)

Unterschiedliche Autoren: 160; Artikel: 185; Jahre: 19

Die narrativen Strukturen sind teilweise durch emotionale Ausrufe geprägt, wie dieses Muster von Adjektiv-Nomen-Ellipsen zeigt:

Elliptische Ausrufe

ADJA NN !/\$.
Frequenzen absolut: 709 (468); LLR: 120.542663; p < 0

War das ein **wertvoller Fund!** (Band 1978, "Dammagrat", Tony Müller)
Ein **wunderbares Wandergebiet!** (Band 1966, "Wandern in Mittelgebirgen", Walter Zeugin)
Aber auch hier ein **schiefgestelltes Kreuz!** (Band 1961, "Almageller Eindrücke", Immanuel Limbac)
Eine **wunderbare Welt!** (Band 1968, "Das Binntal", Jakob Marzohl)
Ein **faszinierender Anblick!** (Band 1970, "Die Mythen", Willy auf der Maur-Pfy)
Welch **andere Welt!** (Band 1972, "Vom Biwakieren", Henri Furrer)
Was für **verlorene Zeit!** (Band 1974, "Die Südwand des Torre Trieste (Civetta, Dolomiten)", Maurice Cochran)

Unterschiedliche Autoren: 260; Artikel: 367; Jahre: 20

Eine ähnliche Funktion kann das folgende Muster haben, bei dem Konnektoren am Satzanfang stehen:

./\$. ADV ,/\$,
Frequenzen absolut: 414 (184); LLR: 151.434942; p < 0

Endlich gab es etwas zu trinken. **Da**, ein Knall, ein Feuerstrahl schoss durch das Zelt! (Band 1965, "Deutsche Himalaya-Expedition 1965 auf den Gangapurna", Günter Hauser)

Dann lässt man nämlich die Finger davon, wenn man nicht zu den ganz guten unter den Alpinisten zählt. **Nun**, für den Normalweg reicht es allemal, und dieser bewegt sich über lange Strecken auf schwach ansteigenden Schneehängen. (Band 1977, "Ein Viertausender für jedermann (Breithorn)", Karl-Wilhelm Spech)

Durch ein System offener Verschneidungen und vier Seillängen hoher Platten überwinden wir die letzten Schwierigkeiten. **Jetzt**, da sich die Wand unter unseren Füßen befindet, sind wir glücklich, dass wir sie bezwingen konnten. (Band 1970, "Die Westflanke der Petites Jorasses", Raymond Monnera)

Unterschiedliche Autoren: 202; Artikel: 278; Jahre: 20

Viele Artikel weisen eine persönliche Note auf, da sie als Erzählungen aus erster Hand, als Erlebnisberichte, inszeniert sind. Daher auch die häufige Verwendung von Personalpronomen, hier in Verbindung mit Eigennamen, was fast immer Vornamen sind:

NE und PPER	Frequenzen absolut: 344 (104); LLR: 196.531700; $p < 0$
	Während eines mehrtägigen Aufenthaltes im folgenden Winter auf den Black Tusk Meadows, dem Mount Garibaldi gegenüber, reifte in Ray und mir der Entschluss, im kommenden Frühjahr eine Besteigung mit Ski zu versuchen. (Band 1972, "Skitour durch Sloweniens Alpen", Albert Allgaier) Gregorio und ich haben vor, am nächsten Tag noch den Nebengipfel zu erobern. (Band 1973, "Cerro Morado - mein erster Fünftausender", H. Eigenmann) Am 27. Januar 1962 verlassen wir, G. Lenser, H. Rützel, E. Forrer und ich , unsere Heimat, um bei strömendem Regen Richtung Genua zuzusteuern. (Band 1964, "Erstbesteigung des Pumo Ri", Ueli Hürleman)
Unterschiedliche Autoren: 141; Artikel: 173; Jahre: 20	

2.3.2. 1990er und 2000er-Jahre

Während im älteren Korpus Erzählungen in der ersten Person vorherrschen, sind im jüngeren Korpus Berichte typisch, bei denen in indirekter Rede oder in dritter Person berichtet wird:

VVFIN NE NE	Frequenzen absolut: 1392 (415); LLR: 358.730056; $p < 0$
	In der Schweiz hingegen sei es durch die hiesige Plaisirkletterkultur in den Hintergrund gedrängt worden, erklärt Emanuel Wassermann , Bergführer und Präsident des Verwaltungsrats von Bergpunkt. (Band 2009, "Klettern wie die Erstbegeher. Clean Climbing", Caroline Fink) Es folgte eine heftige Diskussion, und das Militärdepartement verbot den Anlass, erläutert Marius Robyr . (Band 2008, "Vom Glücksgefühl in Verbier. Die Patrouille des Glaciers", Ruth Oehli) Nun begann Jürg Blum sich für die Einflüsse der Alphaltung auf Gesundheit, Wachstum und Leistung der Tiere zu interessieren, insbesondere auf die Milch- und Fleischleistung. (Band 2007, "40 Jahre ETH-Forschung am Albulapass", Werner Eugster und Hans Leuenberger)
Unterschiedliche Autoren: 298; Artikel: 754; Jahre: 20	

Hinter den Mustern „finites Hilfsverb – *die* – Nomen“, „Partizip – finites Hilfsverb – Satzende“ bzw. „Nomen – Partizip – finites Hilfsverb“ (und vielen weiteren ähnlichen Mustern) verbergen sich entweder Passivkonstruktionen oder aber Perfekt-Konstruktionen:

VAFIN die/ART NN	Frequenzen absolut: 4308 (2188); LLR: 308.201684; $p < 0$
	So wurden die Berge südlich der Vispertäler Ende des 18. Jahrhunderts als schweizerisches Grönland und die scheusslichste Wildnis der Schweiz beschrieben. (Band 2006, "Die vielen Gesichter der Dent d'Hérens. Einer der schönsten Viertausender", Françoise Funk-Salami) Trotzdem hat die Expedition ohne Hochträger und Sauerstoff – mit Ausnahme eines Mitglieds – den Gipfel erreicht. (Band 2003, "Kangchen... wie bitte? «Unkonventionelle» 8000er-Expedition", Brian Duthiers)
Unterschiedliche Autoren: 629; Artikel: 1726; Jahre: 20	

VVPP VAINF ./\$.	Frequenzen absolut: 3522 (1233); LLR: 693.346469; $p < 0$
	Aber eben, auch als Rückweg darf diese Variante nicht benützt werden .

(Band 1994, "Begegnung mit Stein- und Eiszeit", Veronika Meyer)
Helm, Anseilgurt und Klettersteig waren vorhanden, gutes Schuhwerk und Schwindelfreiheit hingegen mussten **mitgebracht werden**. (Band 2002, "Klettersteig am Eggstock. Über drei Gipfel", Felix Ortlieb)
Im Gebiet ist kein Auto nötig, für tageweise Ausflüge kann direkt vor Ort ein Mietauto **gebucht werden**. (Band 2008, "SAC Clubhüttenverzeichnis 2008 (inkl. AACB, AACBs, AACZ, CAAG)", Redaktion)

Unterschiedliche Autoren: 570; Artikel: 1494; Jahre: 20

Auffällig sind zudem Muster, die den Service-Aspekt der neueren Hefte widerspiegeln, so Konstruktionen der Art *sich melden bei [Personennamen]* oder *in der Schweiz*-Konstruktionen im Zusammenhang von Informationen wie *Anbieter/Hersteller in der Schweiz* etc.

3. Fazit

Wenn die verschiedenen Beobachtungen oben zusammengefasst werden, wird deutlich, dass sich das 'Sprechen über Berge' in den SAC-Jahrbüchern von 1880 bis 2009 stark verändert hat. Um die bisher abstrakten Analysen etwas zu veranschaulichen, seien an dieser Stelle Ausschnitte aus Beispieltextrn der vier Phasen zitiert:

1) Wir hielten uns möglichst hoch auf der nördlichen, nach Süden fallenden Seite von Vaplona, gewannen das kleine Tobel, das gegen den Punkt 2547 m ansteigt, waten durch dasselbe hinaus und erreichten den genannten Punkt um 8 Uhr 45 Min. Hier sahen wir etwa 200 m unter uns den Schottensee mit Schnee und Eis bedeckt. *Schottenseefurke* wäre vielleicht, der *Wildseefurke* (2515 m) entsprechend, der passendste Name für diese Scharte zwischen den Punkten 2647 m und 2650 m. Wir blieben nicht lange hier, sondern stiegen bald wieder links auf, theils über Schnee, theils über Verrucanotrümmer, und betraten den Gipfelpunkt 2650 m um 9 Uhr 15 Min. Es ging ein schwacher Windzug, und das Thermometer zeigte auf -2 C.

Jahrbuch 1888-1889: Eine Sectionsfahrt auf den Piz Sol (J. J. Schiesser)

2) [N]ochmal wird angegriffen. Und endlich gelingt es mir, für die linke Hand ganz oben einen guten Griff zu schaffen. Die Rechte bohrt sich mit dem Eisbeil in der ersehnten Rampe ein Loch, der rechte Fuss steigt nochmal nach, und ich sehe über den Rand, indessen der Körper schwer nach aussen hängt. Sich ganz auf die rechte Hand verlassend, fährt die linke blitzschnell weit über den Rand in den Firn, ein Ruck, und ich liege verschnaufend auf dem Bauch, während die Füsse in der Luft baumeln.

Die Alpen 1932: Piz Bernina-Nordostflanke (Karl Schneider)

3) Und da der Tag gut ausgefüllt war, beschliessen wir zu biwakieren. Schon summt der Kocher, und alles wäre in Ordnung, wenn sich nicht über dem Brenva-Becken schwere, schwarze Wolken entleeren würden. Wir sind auf der Höhe der Poire. Der ganze Mont-Blanc-Gipfel steckt in den Wolken, während die Sicht nach unten gut ist. Da erleuchtet ein Blitz den ganzen Berg. Es ist Nacht geworden, aber eine Stunde lang zuckt ein Blitz nach dem andern. Mit einer Zeltplane zugedeckt, versuchen wir, uns an den Berg zu kleben. Bei jedem Blitzstrahl halten wir einen Augenblick den Atem an. Leise beginnt es zu schneien. Hie und da sehen wir Lampen in der Wand. Eine Seilschaft klettert in den Ausläufern der Major-Route trotz schlechtem Wetter und Lawinengefahr. Wir wissen, was wir an ihrer Stelle tun würden! Für uns gibt es keinen anderen Ausstieg als den Gipfel des Pilier. Dann werden wir weiter sehen.

SAC-Jahrbuch 1975: Pilier d'Angle: Bonatti/Gobbi-Führe (Michel Vaucher)

4) Bei P. 2508 den Weg direkt nordwärts einschlagen, der Hirschplatten quert und den Tällistock umrundet. Achtung, die Tunnels von Hirschplatten können Anfang Saison noch mit Schnee gefüllt sein. In diesem Fall muss man absteigen, denn sie können nicht umgangen werden. Rechts ist der Muttgletscher zu bewundern, der die Rhone speist. Weiter auf dem Höhenweg bis in den Furkapass. Bis Anfang des 20. Jahrhunderts war die Furka der längste

befahrbare Pass der Schweiz. Vom Pass können Sie, wenn die Füße nicht mehr tragen, ins Postauto steigen und gemütlich nach Andermatt hinunterfahren.

Die Alpen 2006: Zu Fuss um den Pizzo Rotondo. Gebirgige Wasserlandschaft (Vincent Theler)

Sowohl in den 1930er/40er als auch in den 1960er/70er werden Geschichten aus sehr persönlichen Erzählperspektiven heraus erzählt. Allerdings scheinen die Texte der 1930er/40er noch emotionaler zu sein, was man auch an der Verwendung von Intensivierern ablesen kann (vgl. Abbildung 5): In den 1930ern/40ern werden besonders häufig Intensivierer verwendet, in den 1960er/70er werden sie etwas seltener.

Dies steht im Kontrast sowohl zu den Daten aus jüngster Zeit (1990 bis 2009) als auch zu den ältesten analysierten Daten (1880er/90er). Erstere widerspiegeln den Professionalisierungsschub der SAC-Publikation, wo heute, wie in vielen journalistischen Erzeugnissen, distanziertere Berichte ohne Involvierung des Erzähler-Ichs und mit systematischen Info-Zusammenstellungen als Serviceangebot für die Leserschaft Verbreitung finden. Damit einher geht eine häufigere Verwendung von Passivformen (Abbildung 4).

Die ältesten untersuchten Texte lesen sich dagegen wie Laborberichte, in denen von wissenschaftlichen Experimenten, nämlich dem Besteigen von Bergen, die Rede ist (Bubenhofer/Schröter 2010). Auch hier sind Passivformen frequenter als in den 1930er bis 1970er-Jahren (Abbildung 3). Sowohl diese ältesten Texte als auch die ganz neuen (hauptsächlich ab 2000) weisen zudem einen höheren K-Wert nach Yule auf, der den lexikalischen Reichtum ausdrückt (Abbildung 2). Dies widerspricht auf den ersten Blick den anderen Beobachtungen, wo für die Zeit der 1930er bis 1970er-Jahre ein emotionales Erzählen postuliert wurde. Dieses Erzählen geht demnach nicht unbedingt mit großer lexikalischer Abwechslung einher; viel eher ist es wohl so, dass die Texte in den 1880/1890er und den 2000er-Jahren heterogener sind und daher ein reicheres Vokabular zustande kommt. Das gilt ganz besonders für die jüngsten Texte, wo die Konzeption der Hefte sich im Vergleich zu den Jahrgängen bis in die 1980er-Jahre stark verändert hat und auch ein ganz neuer Mix von Textsorten und Texttypen vorhanden ist.

In einem weiteren Kontext könnte man diese Befunde als Zeichen dafür interpretieren, dass sich die Motivation, sich in den Bergen zu bewegen, verändert hat. Während im ausgehenden 19. Jahrhundert der wissenschaftliche Entdeckergeist vorherrschte, mit dem in sachlicher und systematischer Weise die Bergwelt entdeckt wurde, steht ab den 1930er-Jahren der Mensch im Vordergrund: Es ist seine innere Welt, die in Reaktion zum äußeren Erleben im Vordergrund steht.⁸ Kameradschaft und Gemeinschaft scheinen besonders wichtige Werte zu sein, die als wichtige Tugenden des Bergsteigens konstruiert werden. Dieses Wir-Gefühl in Kombination mit der Verherrlichung der Natur in religiös-romantischer Weise könnte besonders während der

⁸ Günther (1998, S. 13, 161), die sich in ihrer Studie des „bürgerlichen Alpinismus (1870-1930)“ stark auf die Schriften des Deutschen und Österreichischen Alpenvereins (DÖAV) stützt, spricht dementsprechend davon, dass „der Entdeckeralpinismus [um die Jahrhundertwende] vom Schwierigkeitsalpinismus abgelöst wird“.

nationalsozialistischen Zeit als Form des Eskapismus – aber natürlich auch der Geistigen Landesverteidigung (Jorio 2008) – gedeutet werden.

In neuster Zeit ab den 1990er-Jahren zeigt sich ein utilitaristischer Zugang zu den Bergen, bei dem diese als Objekte der Freizeitgestaltung und des sportlichen Wettkampfs betrachtet werden und daher Tipps und Vorschläge gefragt sind, wie die Freizeit am besten gestaltet werden kann. Vergleichbar damit beobachtet StremLOW (1998, S. 278f., 265), in "Magazinen einzelner Aktivsportarten" des späten 20. Jahrhunderts, dass die "Landschaft der Alpen" dort "zur Steigerung des Ich-Erlebnisses funktionalisiert" werde. Dass die "postmoderne Freizeitgesellschaft" die Alpen als "Sportgerät zur Auslösung von körperlichen Erlebnissen" auffasst, bemerkt auch Bätzing (2003, S. 19).

Diese Hypothesen decken sich teilweise mit anderen Forschungsergebnissen und Einschätzungen, wie oben verwiesen. Aus sprachwissenschaftlicher Sicht ist es jedoch interessant, die Hypothesen linguistisch fassen zu können: Die Veränderungen des sozialen Handelns findet einen Niederschlag in den untersuchten Texten und die Texte konstituieren gleichzeitig Deutungsmuster, mit denen soziales Handeln bewertet wird. Gleichwohl ist klar, dass die gewonnenen Hypothesen wiederum durch weitere Analysen validiert werden müssen. Methodisch gesehen bietet es sich deshalb auch an, das bisher eher datengeleitete in der Folge um ein hypothesengeleitetes Vorgehen zu ergänzen.

Methodisch hoffen wir gezeigt zu haben, dass ein datengeleiteter, korpuslinguistischer Zugang linguistische Phänomene zutage fördert, die pragmatisch gedeutet werden können. Natürlich sind nicht alle gefundenen Muster pragmatisch motiviert, sondern häufig gibt es formale Gründe für typische Phänomene, z.B. veränderte Textsorten und thematische Verschiebungen. So spielen in den heutigen Heften des Alpenclubs Berichte über Besteigungen eine weit geringere Rolle als früher, was zwangsweise auch den Einsatz linguistischer Mittel verändert. Trotzdem lassen sich auch solche offensichtlichen, formalen Veränderungen teilweise als Produkt veränderter Denk- und Handlungsweisen lesen.

Bibliographie

- Bätzing, W. (2003): Die Alpen: Geschichte und Zukunft einer europäischen Kulturlandschaft.
- Belica, C./Steyer, K. (2006): Korpusanalytische Zugänge zu sprachlichem Usus. In: AUC (Acta Universitatis Carolinae), Germanistica Pragensia XX.
- Biedermann, R. (1969): Die deutschen Gradadverbien. Heidelberg.
- Bierwisch, M. (1987): Semantik der Graduierung. Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven. In: Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven. (Studia Grammatica). Berlin, 91-286.
- Bubenhof, N. (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. (Sprache und Wissen). Berlin, New York.
- Bubenhof, N./Dussa, T./Ebling, S. et al. (2009): „So etwas wie eine Botschaft.“ Korpuslinguistische Analysen der Bundestagswahl 2009. In: Sprachreport 4, 2–10.
- Bubenhof, N./Schröter, J. (2010): Die Alpen. Sprachgebrauchsgeschichte – Korpuslinguistik – Kulturanalyse. In: Wohin steuert die Historische Sprachwissenschaft? Tagung vom September 2009 in Debrecen, Ungarn.
- Bubenhof, N./Volk, M./Althaus, A. et al. (Hrsg.) (2011): Text+Berg-Korpus (Release 145).

Evert, S./The OCWB Development Team (2010): The IMS Open Corpus Workbench (CWB) CQP Query Language Tutorial.

Feilke, H. (2000): Die pragmatische Wende in der Textlinguistik. In: Brinker, K. (Hrsg.): Text- und Gesprächslinguistik/Linguistics of Text and Conversation. (Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science). Berlin/New York, 64–82.

Feilke, H./Linke, A. (Hrsg.) (2009): Oberfläche und Performanz. Untersuchungen zur Sprache als dynamische Gestalt. Berlin, New York.

Günther, D. (1998): Alpine Quergänge: Kulturgeschichte des bürgerlichen Alpinismus (1870 -1930).

Jorio, M. (2008): Geistiges Landesverteidigung. In: Historisches Lexikon der Schweiz (HLS).

Manning, Ch. D./Schütze, H. (2002): Foundations of Statistical Natural Language Processing. 5. Aufl. Cambridge, Massachusetts.

Miranda-García, A./Calle-Martín, J. (2005): Yule's Characteristic K Revisited. In: Language Resources and Evaluation 39, 4, 287-294.

Os, Ch. v. (1989): Aspekte der Intensivierung im Deutschen. (Studien zur deutschen Grammatik). Tübingen.

Perkuhn, R./Belica, C. (2006): Korpuslinguistik – Das unbekannte Wesen. Oder Mythen über Korpora und Korpuslinguistik. In: Sprachreport 22, 1, 2–8.

Perkuhn, R./Belica, C./al-Wadi, D. et al. (2005): Korpustechnologie am Institut für Deutsche Sprache. In: Schwitalla, Johannes/Wegstein, Werner (Hrsg.): Korpuslinguistik deutsch: synchron – diachron – kontrastiv. Würzburger Kolloquium 2003. Tübingen, 57–70.

Scharloth, J./Bubenhofer, N. (2011): Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse. In: Felder, E./Müller, M./Vogel, F. (Hrsg.): Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen von Texten und Gesprächen. Berlin, New York.

Schiller, A./Teufel, S./Thielen, C. (1995): Guidelines für das Tagging deutscher Textcorpora mit STTS. Stuttgart.

Schmid, H. (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees.

Sinclair, J. (1991): Corpus, Concordance, Collocation. Oxford.

Stefanowitsch, A./Gries, S. Th. (2003): Collostructions: Investigating the Interaction of Words and Constructions. In: International Journal of Corpus Linguistics 8, 2, 209–243.

Steyer, K. (2004): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: Steyer, K. (Hrsg.): Wortverbindungen – mehr oder weniger fest. (Institut für Deutsche Sprache. Jahrbuch 2003). Berlin, New York, 87–116.

Stremlow, M. (1998): Die Alpen aus der Untersicht: von der Verheissung der nahen Fremde zur Sportarena: Kontinuität und Wandel von Alpenbildern seit 1700.

Tognini-Bonelli, E. (2001): Corpus Linguistics at Work. (Studies in Corpus linguistics). Amsterdam.

Yule, G. U. (1944): The Statistical Study of Literary Vocabulary. Cambridge.