



Kollokationen

Lernziele des Moduls „Kollokationen“

Nach Bearbeitung des Moduls

- können Sie unterschiedliche Definitionsansätze von Kollokationen wiedergeben.
- sind Sie sich der terminologischen Problematik des Begriffs bewusst.
- haben Sie ein Grundverständnis für statistische Berechnungen und Begrifflichkeiten, insbesondere für Signifikanzmaße, erworben.
- sind Sie in der Lage, Ihrer Fragestellung bzw. Hypothese entsprechend aus verschiedenen Parametern zu wählen.
- haben Sie das Handwerkszeug und die Übung zur Interpretation von Kollokationsprofilen.

Einführung

Was ist eine Kollokation?

Der Begriff *Kollokation* geht auf J.R. Firth (1957) zurück, der damit charakteristische und **häufig wiederkehrende Wortkombinationen** bezeichnete. Er war der Ansicht, dass man anhand der Kollokatoren eines Wortes, dessen Bedeutung und Verwendungsweise veranschaulichen kann.

Ein Kollokator ist ein Element, also ein einzelnes Wort einer Wortverbindung bzw. Kollokation.

Vielzitiert ist in diesem Zusammenhang der Satz Firths „You shall know a word by the company it keeps“ (Firth 1957, in Evert 2009: 1213).

Dieser Ausspruch erinnert an Wittgenstein und seine Gebrauchstheorie der Bedeutung. (vgl. Lemnitzer/Zinsmeister 2010:30)

Beispiele für Kollokationen nach Firth wären „Kuh“ und „Milch“ oder „Tag“ und „Nacht“.

Im Vordergrund steht für ihn bei Kollokationen die **gegenseitige Erwartbarkeit** (*mutual expectancy*) der Wörter und somit auch der Einfluss der **Statistik**.

Die Definition Firths bleibt ein recht vager Ausgangspunkt. Heutzutage werden Kollokationen in der Computerlinguistik, der Phraseologie, der Fremdsprachendidaktik und nicht zuletzt der Korpuslinguistik daher verschieden ausgelegt.

Lesarten des Begriffs „Kollokation“

In der Phraseologie

versteht man unter einer Kollokation, basierend auf Hausmann (1989), eine **binäre Struktur**, bestehend aus den zwei Einheiten **Basis und Kollokator**. Es handelt sich um eine **phraseologische Kombination zweier Wörter** zwischen denen ein **hierarchisches Determiniertheitsverhältnis** besteht. Anschaulicher wird insbesondere der letzte Punkt durch Hausmanns folgende Erklärung: „Der Tisch braucht das Decken nicht, um Tisch zu sein. Der Kollokator *decken* aber braucht dringend *Basen*, um überhaupt etwas zu sein: *Dach decken*, *Stute decken*, *Unkosten decken* [...]“ (Hausmann 2004: 312).

Weitere Beispiele für Kollokationen nach Hausmann sind „Anker lichten“ oder „himmelweiter Unterschied“.

Zusammengefasst sind Kollokationen bei Hausmann demnach zweiteilige Wortverbindungen, zwischen denen ein Abhängigkeitsverhältnis und eine Rangordnung bestehen.

Kollokationen werden auch in Harald Burgers Dreibein-Modell der Phraseologie aufgegriffen, in welchem Phraseologismen in die drei Typen „Kollokationen“, „Idiome“ und „Teilidiome“ unterteilt sind. Es gilt zu beachten, dass Redewendungen bzw. Idiome von Kollokationen unterschieden werden, da die jeweilige Struktur verschieden ist. (vgl. Hausmann 2004: 313f.)

Auch in der Lexikographie und Fremdsprachendidaktik orientiert man sich am Kollokationsbegriff Hausmanns.

Seiner Meinung nach herrscht derzeit ein Terminologiekrieg um die Besetzung des Begriffs Kollokation, wobei das Nebeneinander der verschiedenen Definitionen „[...] eine historisch gewachsene Perversion des wissenschaftlichen Diskurses darstellt [...]“ (Hausmann 2004: 321).

In der Computerlinguistik

sind Kollokationen **lexikalisierte**, also festgelegte **Wortverbindungen mit semantischen oder syntaktischen Besonderheiten**. Sie sind **nicht kompositionell** und ihre Elemente sind **nicht modifizierbar** und **nicht substituierbar**.

Nicht kompositionell bedeutet, dass sich die Bedeutung der Kollokation nicht aus ihren einzelnen Elementen ableiten lässt. Nicht modifizierbar meint, dass ein Kollokator nicht abgewandelt werden kann, indem beispielsweise statt der Singularform („Mittel zum Zweck“) die Pluralform verwendet wird („Mittel zu Zwecken“). Nicht substituierbar heißt, dass ein Element („zur Sache kommen“) nicht durch ein anderes, wenn auch bedeutungsgleiches Element ausgetauscht werden kann („zur Sache gelangen“).

Gemäß dieser Definition, können Kollokationen somit beispielsweise im Bereich der maschinellen Übersetzung nicht Wort für Wort in die Zielsprache übertragen werden und stellen für die maschinelle Sprachverarbeitung im Allgemeinen eine Herausforderung dar. Neuerdings wird der Terminus *Kollokation* in der Computerlinguistik jedoch durch den weniger vorbelasteten Begriff *multiword expression* abgelöst.

Computerlinguistische Kollokationen sind beispielsweise „Mittel zum Zweck“ oder „zur Sache kommen“.

In der Korpuslinguistik

werden Kollokationen zumeist in der Tradition Firths als **überzufällig häufig gemeinsames Vorkommen zweier Wörter** ausgelegt.

Kann man also mit statistischen Methoden nachweisen, dass „[...] Paare von Wörtern signifikant häufiger miteinander vorkommen, als

dies auf Grund einer zufälligen Verteilung von Wörtern in Texten zu erwarten wäre [...]“ (Lemnitzer/Zinsmeister 2010: 15f.), so bezeichnet man diese, unabhängig von deren Wohlgeformtheit, als Kollokationen.

Um sie von der phraseologischen Definition klar abzugrenzen, erweitern einige Wissenschaftler wie Evert (2009) oder Biemann/Bordag/Quasthoff (2003) die Begriffsbezeichnung und unterscheiden zwischen **empirischen** und theoretischen bzw. **statistischen** und linguistischen **Kollokationen**.

Beispiele für solche empirischen oder statistischen Kollokationen wären demnach Wortpaare wie „und er“ oder „welchen Einfluss“.

Einen anderen Weg aus dem definitorischen Dilemma geht man beispielsweise am Institut für Deutsche Sprache. Dort werden lediglich wohlgeformte Kookkurrenzen, die der Basis-Kollokator-Dichotomie gemäß Hausmann entsprechen, als Kollokationen bezeichnet.

Was ist denn eine Kookkurrenz?

Für Scherer beispielsweise sind Kookkurrenzen **überdurchschnittlich häufig benachbarte Wörter** und ein Synonym zu Kollokationen.

Perkuhn/Keibel/Kupietz (2012) bevorzugen den Begriff *Kookkurrenz* aufgrund der Problematik der unterschiedlichen Auffassungen des Kollokationsbegriffs nach Firth und Hausmann. Dabei betonen sie jedoch, dass es nicht um das **bloße Kovorkommen** zweier Wörter geht, wie die wörtliche Bedeutung des Terminus nahelegt, sondern dass die Rekurrenz und die statistische Bewertung wesentliche Aspekte der Verbindung darstellen.

Am IDS wiederum sagt man: „Wir definieren ‚Kookkurrenz‘ als eine Kohäsionsqualität, die durch mathematisch-statistische Berechnungen ermittelt wird [...]“ (Steyer 2004: 96). Lassen sich die statistisch auffälligen Wortverbindungen darüber hinaus in eine hierarchische Beziehung setzen und im Sinne Hausmanns interpretieren, qualifizieren sie sich außerdem als Kollokationen.

Die kontextuelle Verbindung von „Wasser“ und „Verbrauch“ wäre gemäß der am IDS vertretenen Auffassung also eine Kookkurrenz, da mittels mathematisch-statistischer Berechnungen ein starker und auffälliger Zusammenhalt der Wörter nachgewiesen werden konnte. Das Wortpaar „sauber“ und „Wasser“ entspricht zusätzlich noch der Definition von Hausmann und stellt somit eine Kollokation dar.

Und dann wäre da auch noch die Kolligation...

Der Begriff *Kolligation* findet sich nicht allzu oft in korpuslinguistischer Sekundärliteratur, wird dann aber zumindest einheitlich gebraucht. Im Gegensatz zur lexikalischen oder inhaltlichen Ebene der Kollokation, steht bei der Kolligation die **innere Struktur**, also die **grammatikalische Ebene** im Vordergrund. Das Augenmerk liegt dabei auf der **grammatischen Beziehung zwischen Wortpaaren**, wie sie etwa zwischen Verben und Adverbien besteht.

Diese, ein Syntagma bildenden Einheiten, erinnern an Kollokationen nach Hausmann. Lemnitzer/Zinsmeister führen als Beispiel für eine Kolligation „Antrag stellen“ an und bezeichnen im Gegensatz dazu die frequente Wortverbindung „und er“ als Kollokation.

Ein Beispiel für eine Kolligation ist „Antrag stellen“.

vgl. Biemann/Bordag/Quasthoff 2003; Evert 2009: 1212-1214; Hausmann 2004: 312-320

vgl. Lemnitzer/Zinsmeister 2010: 15f., 30f.; Perkuhn/Keibel/Kupietz 2012: 112f.;

Scherer 2006: 46; Steiner 2004: 38f.; Steyer 2004: 96-99;



Terminologische Festlegung

Durch die vorhergehenden Ausführungen sollte klar geworden sein, dass es sowohl in Bezug auf den Begriff *Kollokation*, wie auch auf den Begriff *Kookkurrenz* eine gewisse Problematik gibt, während von *Kolligationen* eher selten die Rede ist.

Im Fokus der Korpuslinguistik steht das **Kovorkommen von Wörtern**. Von Interesse ist dabei aber nicht die bloße Kookkurrenz, sondern die Frage, ob eine Art Anziehungskraft zwischen den Wörtern besteht und deren gemeinsames Auftreten **überzufällig häufig**, wenn nicht gar **signifikant** ist. Dabei gilt es zu beachten, dass die Wörter nicht unmittelbar nebeneinander stehen müssen. Üblicherweise werden **Textfenster** von fünf Wörtern rechts und links vom Bezugswort als Analysegrundlage herangezogen.

Darüber hinaus gibt es aber auch Ansätze, die das signifikant gemeinsame Vorkommen zweier Wörter in einem Textausschnitt (z.B. innerhalb eines Satzteils oder Satzes) untersuchen, was sich vor allem bei Sprachen mit vergleichsweise freier Wortstellung anbietet. Bei einer anderen Herangehensweise, werden nur die Wörter als Kookkurrenzen aufgefasst, die in direkter oder zum Teil auch indirekter syntaktischer Beziehung stehen. Es werden also nur Kolligationen analysiert. (vgl. Evert 2009: 1222f.)

Dieser statistische Mehrwert gegenüber einer Kookkurrenz soll in der Tradition Firths nachfolgend als **Kollokation** bezeichnet werden. Zwar besteht somit eine Ambiguität zur Begrifflichkeit Hausmanns, doch die Aspekte der Wohlgeformtheit und der hierarchischen Determiniertheit bleiben unberücksichtigt.

Zusammenfassend heißt das: Wenn zwei Begriffe mittels mathematisch-statistischer Verfahren als auffällig dargestellt werden, handelt es sich um eine Kollokation, egal ob die beiden Wörter eine grammatikalisch korrekte oder inhaltlich sinnvolle Einheit bilden oder nicht.

Wissenswerte Begriffe

Kollokationsanalyse

Bei einer Kollokationsanalyse – in der Sekundärliteratur auch Kookkurrenzanalyse genannt – werden **mithilfe statistischer Verfahren** die **auffälligsten Kollokatoren** oder auch Kookkurrenzpartner eines **gewählten Bezugswortes** ermittelt. Hierfür müssen einige **Parameter** sowie ein **Signifikanzmaß** festgelegt werden.

Diese bilden die Grundlage der Berechnungen und können unterschiedliche Betrachtungsweisen des Korpus widerspiegeln. Sie sollten daher an die Frage- bzw. Zielstellung angepasst werden.

Als Ergebnis einer Kollokationsanalyse wird das Kollokationsprofil eines Bezugswortes angezeigt.

[IMAGE „Beispielhaftes Ergebnis einer Kollokationsanalyse“]

Kollokationsprofil

Das Kollokationsprofil eines Wortes ist die **Gesamtheit seiner relevanten Kollokationen**. Während eine einzelne Kollokation einen Teilbereich davon abbildet, wie ein Wort üblicherweise gebraucht wird, stellt ein Kollokationsprofil „[...]“ viele, wenn nicht sogar alle wichtigen Aspekte oder Nuancen der Verwendung eines Wortes [dar]. Wir können uns damit einen Überblick über alle wichtigen Diskurse verschaffen, in denen sich die vielleicht unterschiedlichen Bedeutungen eines Wortes konstituiert haben“ (Perkuhn/Keibel/Kupietz 2012: 127).

Kollokationen erster/zweiter/dritter Ordnung

Grundsätzlich liegt Kollokationen eine binäre Struktur zugrunde: Berechnet man auf Grundlage eines Bezugswortes die signifikantesten Kollokationen, so wird als Resultat üblicherweise eine Liste mit den signifikantesten Kollokatoren erster Ordnung ausgegeben. Ansätze wie bspw. das Clustering erweitern diese Struktur, indem in einem zweiten Schritt die signifikantesten Kollokatoren eines Kollokatoren erster Ordnung berechnet werden. Auf diese Weise erhält man die Kollokatoren zweiter Ordnung. Man kann dieses Vorgehen auch auf Kollokatoren dritter Ordnung und darüber hinaus anwenden. Die erzielten Ergebnisse bilden dann sozusagen ein Kollokationsnetz und können auch entsprechend visualisiert werden:

[IMAGE]



© www.ClipProject.info

Anwendungsmöglichkeiten

Wozu kann die Berechnung von Kollokationen dienen?

- Als Hinweis für die Anordnung bzw. Hierarchie von Lesarten eines sprachlichen Zeichens in einem Wörterbuchartikel.
 - *Welches ist die gängigste Bedeutung des Lexems?*
- Zur Identifikation, Verifizierung und Beschreibung typischer Wortverbindungen und Wortverbindungphänomene (z.B. Idiome)
 - *Welche Kollokationen werden verwendet?*
 - *Welche sind besonders gebräuchliche Kollokationen und welche finden sich kaum oder gar nicht?*
 - *In welcher Bedeutung und in welchem Kontext werden sie gebraucht?*
- Gängige grammatische Gebrauchsmuster können ermittelt werden
- Syntaktische Bindungen können aufgefunden gemacht werden
 - z.B. typische präpositionale Anschlüsse*
- Die Bedeutungsfelder eines Wortes werden anschaulich
- Thematische Felder können ausgemacht werden
- Die Verwendungsweisen von bestimmten Begriffen im Diskurs können deutlich gemacht werden
- Kollokationen können als Widerspiegelung von Traditionen des Formulierens bzw. Gebrauchskonventionen untersucht werden

Kollokationsanalysen sind damit auch für die Lexikographie und Lexikologie, die Fremdsprachendidaktik und die Übersetzungswissenschaft von besonderer Bedeutung.

Lemnitzer/Zinsmeister 2010: 144f.; Steyer 2004: 87-94, 103

Praxisbezug

Anwendungsbeispiel aus der Lexikographie und Fremdsprachendidaktik

Vor dem Hintergrund, dass usuelle Wortverbindungen einen essentiellen Bestandteil unserer Sprache darstellen, für Nichtmuttersprachler aber oft eine große Hürde sind, entstand am IDS 2008 das Projekt *SprichWort. Eine Internetplattform für das Sprachenlernen*. 2011 ging es in das Projekt *Usuelle Wortverbindungen* über und seine Zielgruppe bilden insbesondere Fremdsprachenlerner und –lehrer sowie Entwickler von Lehr- und Lernmaterialien.

„Wortverbindungen verkörpern [...] ‚geronnene kulturelle Erfahrungen‘ einer Sprachgemeinschaft (z.B. Sprichwörter), und es ist deshalb wichtig, sie zu sammeln, zu archivieren und in ihrem gegenwärtigen Gebrauch auch für die Nachwelt zu dokumentieren. Wenn wir die ‚Logik der Wortverbindungen‘ verstehen, können wir sehr viel über die Strukturen der Sprache und des Denkens erfahren. Deshalb kann die Erforschung von Wortverbindungen auch zur Theorienbildung in der Linguistik selbst wertvolle Beiträge leisten“ (Quelle).

Für gewöhnlich greifen Lexikographen bei dem Versuch, übliche Wortverbindungen und Verwendungsbeispiele aufzulisten, auf ihre Erfahrung, Intuition und Sprachkompetenz zurück. Zudem werden Versuchspersonen befragt oder psychologische Assoziationstests durchgeführt.

Am Institut für Deutsche Sprache wollte man sich hingegen die Computertechnologie und den Bestand an Korpora zunutze machen. Mit statistischen Methoden wurden authentische Sprachdaten analysiert und linguistisch interpretiert.

Basierend auf dem Deutschen Referenzkorpus entstand so ein erstes, empirisch untermauertes und online dokumentiertes Kerninventar der aktuell gebräuchlichen Sprichwörter des Deutschen. Die 300 Wörterbuchartikel beinhalten eine Validierung des Sprichwortstatus, die Bedeutung, in der das Sprichwort verwendet wird, seine typischen Formvarianten und lexikalischen Ersetzungen, Beispiele für die textuelle Einbettung des Sprichwortes sowie seine typischen Gebrauchsbesonderheiten. Zudem sind Sprichwort-Äquivalente für andere Sprachen verlinkt. Das Sprichwörterbuch ist zum einen über die SprichWort-Plattform verfügbar und wurde zum anderen in das Online-Wortschatz-Informationssystem Deutsch (OWID) des Instituts für Deutsche Sprache integriert.

<http://www.owid.de/index.jsp>

<http://www.sprichwort-plattform.org/>

<http://www1.ids-mannheim.de/lexik/uwv.html>

<http://www1.ids-mannheim.de/lexik/uwv/uwv.html>

<http://www.owid.de/wb/sprw/start.html>

<http://www1.ids-mannheim.de/lexik/sprichwort/projekt.html>

Anwendungsbeispiel aus der Diskurslinguistik

Angelehnt an die Kognitive Linguistik sowie die Stereotypen- und Diskursforschung unternimmt Friedemann Vogel in einer Studie den Versuch, mittels korpuslinguistischer Verfahren das öffentliche Medienimage der *Türkei* sowie von *Türken* im deutschsprachigen Raum zu untersuchen.

„Im Fokus liegen dabei rekurrente Sprachmuster, die Hinweise geben auf medienvermittelte ‚Prototypen der‘ bzw. ‚des‘ >Türken< / ‚der‘ >Türkei<“ (Vogel 2010: 11).

Um diesen auf den Grund zu gehen, macht er zunächst sämtliche Lexeme ausfindig, welche die Zeichenfolge *türk* enthalten. In einem nächsten Schritt führt er unter anderem eine Kookkurrenzanalyse des Lexems *Türkei* mit einer Fenstergröße von -5/+5 durch.

Er stellt dabei fest, dass die hitzige Debatte um den EU-Beitritt der Türkei ein ganz zentrales Thema seiner Korpusdaten darstellt.

Des Weiteren lassen sich die 100 signifikantesten Kookkurrenzpartner von *Türkei*, wie zum Beispiel *Irak*, *Nato*, *Frauen*, *Erdbeben*, *Reise* oder *abgeschoben* „[...] den Kontexten [...] >Integration und Assimilation<, >Militärische Bündnispartner und Intervention< [...], >Menschenrechts- bzw. Demokratiedefizit< [...], >problematische Geschlechterverhältnisse< [...], >Religion<, [...] >Terrorismus, Radikalität, Verschlossenheit und orientale Fremde<; ferner >Naturkatastrophen< [...], >Urlaub [...] sowie schließlich >Asyl- und Abschiebeprozesse< [...] [zuordnen]“ (Vogel 2010: 20).

In Verbindung mit weiteren Berechnungen kommt Vogel letztlich zu dem Fazit, dass man in Deutschland lebenden Türken zwar durchaus positive Eigenschaften zuschreibt, diese aber durch negative Attribute überzeichnet werden. Insbesondere junge Türken werden des Öfteren als gefährlich und auf verschiedenen Ebenen als Bedrohung wahrgenommen.

Die Türkei gilt als in sich widersprüchliches Land, wobei auch positive Aspekte zu finden sind, da es sich um einen wirtschaftlichen und militärischen Partner sowie ein beliebtes Urlaubsziel handelt.

Vogel räumt schließlich noch zwei wichtige Punkte ein: Zum einen ist es korpuslinguistisch kaum möglich, Ironie in den Daten auszumachen und zum anderen würde eine kontrastive Analyse der Ergebnisse mit einem Referenzkorpus großes Potenzial bergen.

vgl. Vogel 2010:

Es lässt sich also festhalten, dass Kookkurrenzanalysen ein produktives wie nützliches Verfahren für die Diskurslinguistik darstellen und auch dem Gebiet der Linguistischen Imageanalyse dienlich sein können.

Ein bisschen Statistik



Auch das noch!

Statistische Berechnungen spielen eine wesentliche Rolle bei Kollokationsanalysen. Um die Analyseparameter einstellen und die Ergebnisse richtig interpretieren zu können, sollte man gewisse Grundkenntnisse in diesem Bereich erwerben. In den bisherigen Aussagen über Kollokationen fielen bereits die mathematisch-statistischen Begriffe *überzufällig häufig* und *signifikant*. Zunächst soll geklärt werden, was genau man darunter versteht:

Überzufällig häufig

Verteilt man die Wörter in einem Korpus ganz beliebig, so können ein Wort *a* und ein Wort *b* **x-mal zufällig** gemeinsam vorkommen. Ist eine Wortverbindung zufällig, so sind die beiden Wörter voneinander unabhängig. Man nennt dies auch die **Nullhypothese**. Kommen die beiden Wörter aber **häufiger** gemeinsam vor **als x**, so ist dieses Vorkommen **überzufällig häufig**.

*In natürlichen Sprachen ist auszuschließen, dass Wörter zufällig aufeinandertreffen, da sie syntaktischen, semantischen und lexikalischen Einschränkungen unterliegen. Daher ist die sogenannte Nullhypothese, bei der man davon ausgeht, dass Wort *a* und Wort *b* voneinander unabhängig sind und keinerlei Attraktion zwischen ihnen besteht, umstritten.*

Denken Sie zurück an Ihren Mathematikunterricht: Zu Beginn wird die Wahrscheinlichkeitsrechnung oftmals anhand von Würfeln erklärt. Wir wollen das einfachere Beispiel einer Münze heranziehen: Wirft man eine Münze zehnmal, so könnte man erwarten, dass fünfmal Kopf und fünfmal Zahl erscheint. Zeigt die Münze sechs- oder siebenmal Kopf, so kann das immer noch Zufall sein. Trotzdem **übersteigt** das Ergebnis unsere **Erwartung einer gleichmäßigen Verteilung** und ist damit überzufällig häufig.

Signifikant

Bleiben wir bei erwähntem Beispiel: Wirft man eine Münze zehnmal und sieht neun- oder gar zehnmal Kopf, so ist es mehr als unwahrscheinlich, dass es sich hierbei noch um Zufall handelt. Die **Abweichung** von dem, was wir erwartet hätten, ist **so stark**, dass sie als signifikant bezeichnet werden kann.

Bei der Interpretation von überzufällig häufig auftretenden Kollokationen kann man Aussagen über das zugrundeliegende Korpus ableiten. Sind Wortverbindungen jedoch signifikant, so kann man allgemein gültige Schlussfolgerungen ziehen. Da es sich um statistische Berechnungen auf der Grundlage von Wahrscheinlichkeiten handelt, sind diese aber trotz allem fehlbar.

Bei der Berechnung von überzufälliger Häufigkeit und Signifikanz geht es also darum zu ermitteln, wie häufig ein Wort bzw. eine Wortverbindung vorkommt und wie oft wir das erwartet hätten. Für eine Kollokationsanalyse sind daher folgende Frequenzen relevant: