

Wie öffentlich ist die Hand?

Über Sinn und Unsinn eines Signifikanztests in der Korpuslinguistik

Sven STAFFELDT

Abstract:

How *public* is the *hand*? On the sense and nonsense of the significance test in corpus linguistics

In this article it will be shown that the use of a special statistical method for testing the significance of the co-occurrence of the type *öffentlich+Hand* (the Chi square test) does not make sense in a very large corpus. That means that one main test for measuring the significance of a collocation cannot be applied under standard conditions.

Keywords: corpus linguistics, statistical methods, significance tests, collocation, phraseology

1. Die Frage

Obwohl qualitative und quantitative Analyseverfahren nicht unversöhnlich sind, sondern ganz im Gegenteil Hand in Hand gehen können und sollten, wenn es um die Absicherung und Interpretation von Analyseergebnissen geht, trennt beide doch häufig ein tiefer methodisch-ideologischer Graben. Aus qualitativer Sicht droht bei quantitativen Verfahren die Gefahr, sich nicht mit den interessanten (und relevanten) Gegenständen zu befassen. Aus quantitativer Sicht sind die Ergebnisse qualitativer Studien intersubjektiv häufig nicht genügend nachvollziehbar, um als objektiv (oder wahr) eingestuft werden zu können.

„Fillmore [...] hat die unterschiedlichen Erkenntnisinteressen und die damit einhergehenden methodischen Divergenzen in der viel zitierten Gegenüberstellung zweier karrierter Figuren, des ›Armchair-Linguisten‹ (der, in seinem Schreibstuhl sitzend, introspektiv arbeitet) und des ›Korpuslinguisten‹ (der Daten befragt), auf den Punkt gebracht:

These two don't speak to each other often, but when they do, the corpus linguist says to the armchair linguist, ›Why should I think that what you tell me is true?‹ and the armchair linguist says to the corpus linguist, ›Why should I think that what you tell me is interesting?‹ (Fillmore 1992:35)“ (Spitzmüller/Warnke 2011:26)

Forschungspraktisch ideal ist es natürlich, wenn in den einzelnen Analysen quantitative und qualitative Verfahren zusammengeführt werden. Im Grunde brauchen beide auch einander: Qualitative Analysen nützen nichts, wenn sie nicht eine irgendwie relevante Anzahl von Fällen erfassen, in denen sie Geltung haben: Inhalte ohne Vorkommen sind spekulativ. Und quantitative Analysen würden ohne qualitative zu überhaupt keinen relevanten Ergebnissen kommen, weil man immer noch den Schritt von den Daten zur Interpretation der Daten machen muss, um Forschungsfragen

beantworten zu können: Zahlen ohne Interpretationen sind blind. Es bedarf hier des qualitativ arbeitenden Linguisten, der bei der Beschäftigung mit seiner eigenen Sprache natürlich auch Kompetenz (oder Intuition) mitbringt, die er nutzen kann. So hebt etwa Steyer (2002:234) hervor, „dass linguistische Kompetenz in [...] keiner Phase durch den Computer zu ersetzen ist. Die Entscheidung, wie mit welcher Korpusinformation umzugehen ist und welche Schlussfolgerungen zu ziehen sind, bleibt dem Menschen vorbehalten.“

Im Folgenden wird gezeigt, dass die Ermittlung der statistischen Signifikanz von ausgewählten Kollokationen in einem großen Korpus mittels einer einzigen Analysemethode, die allerdings für diesen Fall als besonders geeignet angesehen wird (nämlich mittels des Chi-Quadrat-Tests), mitunter zu fragwürdigen Ergebnissen führen kann.

Nehmen Sie an, Sie hätten im Archiv der geschriebenen Sprache des IDS über COSMAS mittels der Suchanfrage &Hand¹ über 650.000 Treffer² erzielt, von diesen an 10 gleichmäßig verteilten Stellen jeweils 200 Belege markiert und diese 2.000 Belege schließlich exportiert. Nun haben Sie davon – aus Zeitgründen doch ‚nur‘ – die ersten 1.500 Belege daraufhin durchgesehen, welche Verwendungen von *Hand* auftauchen, insbesondere auch, welche phraseologischen. Und ein Ergebnis ist: In 92 Fällen stehen *öffentlich* und *Hand* direkt aufeinander folgend im selben Satz. Man könnte dies als Kookkurrenz bezeichnen. Die Frage ist nun: Ist diese Kookkurrenz eigentlich auch statistisch signifikant (und damit eine Kollokation i. S. v. Bubenhofer 2009:122)?

Nun haben Sie häufiger gelesen und in Vorträgen gehört, dass Belegkorpora (oder Belegsammlungen) zwar auch Korpora, aber für genau diese Frage nur eingeschränkt brauchbar sind. Das liegt hier natürlich daran, dass wir es mit einem gewissermaßen verzerrten Korpus zu tun haben, in dem *Hand* ja schon vorgegeben ist, es also höchstens überprüft werden kann, ob das linksseitige Vorkommen von *öffentlich* irgendwie auffällig ist gegenüber dem linksseitigen Vorkommen anderer Adjektive zu *Hand*. Nicht aber, ob das rechtsseitige Vorkommen von *Hand* bei *öffentlich* auffällig ist. Von dem Zusammenauftreten dieser beiden sprachlichen Einheiten können Sie mit diesem Belegkorpus also eigentlich nur eine Seite untersuchen. Hier aber immerhin ergibt sich: die Kookkurrenz *öffentliche Hand* ist innerhalb der festgestellten Phraseologismen (N=983) mit 92 Treffern gegenüber anderen Phraseologismen die zweithäufigste phraseologische Verwendung, wobei der erste Platz von einer Einheit mit 94 Treffern gehalten wird.³

Sie müssen also ein anderes Korpus als die Belegsammlung heranziehen und haben sich entschieden, das Archiv der geschriebenen Sprache des IDS dafür zu benutzen. Sie möchten wissen: Ist die Type-Kookkurrenz *öffentliche Hand* im Archiv der geschriebenen Sprache des IDS statistisch signifikant?

2. Das Vorgehen

In linguistischen Lehrbüchern mit einem etwas größeren Statistikteil werden verschiedene Signifikanztests (so etwa der Chi-Quadrat-Test, der Mann-Whitney-Test bzw. U-Test, der Wilcoxon-Test oder der t-Test) erwähnt und beschrieben. Von dem in jedem Lehrbuch anzutreffenden Chi-Quadrat-Test heißt es dabei:

¹ Ausformuliert: Gib mir alle Stellen aus, in denen eine Form von *Hand* auftritt.

² Diese Zahl erhöht sich natürlich ständig, weil das Schriftspracharchiv beständig anwächst. So waren es 2009 noch etwas über 260.000 Treffer bei dieser Anfrage. Das erklärt auch die in Staffeldt (2011:195) genannte Zahl zu dieser Suchanfrage. Zum Zeitpunkt der Veröffentlichung war der Aufsatz nämlich bereits zwei Jahre alt.

³ Die weitere (qualitative) Analyse gestaltet sich so, dass man für *öffentliche Hand* ein eigenes Belegkorpus erstellt und dieses dann für die Analyse heranzieht. So kann man mittels &öffentlich /+w2 &Hand derzeit (März 2012) 23.569 Treffer erzielen. Ausformuliert: Liefere mir alle Stellen, in denen auf eine Form von *öffentlich* im Höchstabstand von zwei Wortformen eine Form von *Hand* folgt. Der Abstand ist so gewählt, um herauszubekommen, ob es auch Vorkommen gibt, in denen zwischen *öffentlich* und *Hand* noch eine weitere Einheit steht, was aber nicht der Fall ist. Von diesen analysiert man die ersten 100 aus dem Jahr 2000 näher und alle Ergebnisse beziehen sich auf (haben Gültigkeit in Bezug auf) zu nächst einmal nur dieses Belegkorpus.

„Immer, wenn Sie das Auftreten von **Häufigkeitsunterschieden** von einzelnen Merkmalen oder auch Merkmalskombinationen analysieren wollen, sind Sie bei der Chi-Quadrat-Familie gut aufgehoben. [...] Man untersucht mit diesen Tests, ob eine empirisch beobachtete Häufigkeitsverteilung nur zufällig oder systematisch von einer theoretisch erwarteten Häufigkeitsverteilung (unter der H_0 [d. i. die Nullhypothese; d. Verf.]) abweicht. [...] Solche Tests werden anhand sogenannter *Kontingenztafeln* oder auch *Kreuztabellen* berechnet.“ (Meindl 2011:162)

„Der Chi-Quadrat-Test wird verwendet, um herauszufinden, ob zwei Variablen in Beziehung zueinander stehen [...] [Er] wird sehr oft gebraucht, wenn wir mit Häufigkeitsdaten zu tun haben, z.B. wenn wir gezählt haben, wie oft etwas vorkommt [...]“ (Albert/Koster 2002:118)

„Sie untersuchen nun die Fragestellung, ob die beiden Konstruktionen entsprechend auch gleich häufig sind oder nicht. [...] Derartige Fragestellungen werden im Allgemeinen mit Tests aus der Familie von Chi-Quadrat Tests behandelt, wohl eine der wichtigsten Testarten überhaupt.“ (Gries 2008:156f.)

Es scheint so zu sein, dass ein Chi-Quadrat-Test für unsere Fragestellung der richtige (oder zumindest nicht der falsche) Test ist (vgl. auch Hauser/Humpert 2009:174). Das bestätigt auch Bubenhofer. Der Chi-Quadrat-Test kann „verwendet werden, um im Vergleich zu anderen Kollokationen im selben Korpus oder zu selben Kollokationen in anderen Korpora eine Rangordnung zu erzeugen“ (Bubenhofer 2009:138), solange die beobachteten Frequenzen einen Wert von 5 nicht unterschreiten.

Das Vierfelderschema (die Kontingenztafel) sieht dann so aus (Erklärung folgt direkt danach):

	S	-S	Total
P	A	B	A+B
-P	C	D	C+D
Total	A+C	B+D	N

Tabelle 1: Vierfelderschema („-“ = Negationsoperator ‘nicht’) (vgl. Bubenhofer 2009:136)

„Um die Signifikanz einer Zwei-Wort-Kollokation zu testen (z. B. *runder Tisch*), werden die beiden Wörter den Variablen P und S zugewiesen. Feld A enthält die Anzahl der Fälle, bei denen Wort P auf S trifft (*runder + Tisch*), Feld B die Fälle in denen P auftritt, jedoch nicht gleichzeitig S (*runder + X* [aber nicht *Tisch*]), Feld C die Fälle, in denen zwar S nicht aber P auftritt (*X* [aber nicht *runder*] + *Tisch*) und Feld D, in denen weder P noch S auftreten.“ (Bubenhofer 2009:136)

Das ist genau das, was wir brauchen, und unsere Fragestellung scheint ein Paradefall für die Anwendung des Chi-Quadrat-Tests zu sein. Wie funktioniert dieser Test nun?

Der Test läuft über den Vergleich von beobachteten vs. erwarteten Häufigkeiten. Man muss also zunächst einmal die Anzahl der beobachteten Fälle ermitteln und daraus dann die Anzahl der erwarteten errechnen. Die Fälle A, B und C sind die Anzahlen der Vorkommen der Kookkurrenz (A) oder der Teile der Kookkurrenz jeweils allein (B und C), D ist die Anzahl aller Wortformen abzüglich A, B und C in einem Korpus. Auf Grundlage dieser Werte wird ein zweites Schema erstellt, das jeweils die errechnete erwartete Anzahl der Vorkommen enthält. Diese ergibt sich für die einzelnen Zellen aus der Summe der Reihe multipliziert mit der Summe der Spalte geteilt durch die Gesamtsumme N. Für die Zelle A ist die Formel zur Errechnung des erwarteten Werts also: ((A+B) mal (A+C)) durch N).

Für jede einzelne Zelle wird nun der Chi-Quadrat-Test durchgeführt. Die Berechnung der einzelnen Chi-Quadrat-Werte geht so vor sich, dass man vom beobachteten Wert den erwarteten abzieht, diese Differenz ins Quadrat setzt und dies schließlich noch einmal durch den erwarteten Wert teilt. Die Summe dieser Tests ist der Wert des Chi-Quadrat-Tests insgesamt. Diesen Wert überprüft man schließlich anhand einer Tabelle, in der Schwellenwerte verzeichnet sind, daraufhin, ob es sich

bei dem mit diesem Wert errechneten Unterschied um einen (auf verschiedenen Niveaus) signifikanten handelt oder nicht.

3. Die Ergebnisse

Über die folgenden Suchketten erhält man die Anzahl der Treffer für die Fälle:

- A (*öffentlich* und *Hand* direkt aufeinander folgend): `&öffentlich /+w1 &Hand`,
- B (*öffentlich* aber *–Hand* direkt aufeinander folgend): `&öffentlich %+w1 &Hand` und
- C (*–öffentlich* aber *Hand* direkt aufeinander folgend): `&Hand %-w1 &öffentlich`.

Den Fall D (weder *öffentlich* noch *Hand*, also: *–öffentlich* und *–Hand* aufeinander folgend) kann man nicht über COSMAS suchen, weil man über COSMAS ja kein totales Nicht-Vorkommen suchen kann. Man muss ihn errechnen, indem man von der Gesamtsumme der Wortformen im benutzten Archiv die Summe der Fälle A, B und C abzieht. Laut IDS hat das Archiv der geschriebenen Sprache derzeit über zwei Milliarden Wortformen, nämlich 2.687.216.673.⁴ Es ergibt sich also für den letzten Fall:

- $D = (2.687.216.673 - A - B - C)$.

Über diese Suchen erhält man die Kontingenztafel für die beobachteten Werte, in der jetzt nur noch die Reihen- und Spaltensummen mit angegeben werden:

beobachtet	Hand	–Hand	total
öffentlich	23538	587726	611264
–öffentlich	634208	2685971201	2686605409
total	657746	2686558927	2687216673

Abbildung 1: Kontingenztafel für die Kookkurrenz *öffentlich+Hand* im Archiv der geschriebenen Sprache

Wie im dritten Abschnitt dargestellt, kann man nun für jeden Fall die Erwartung berechnen. Es ergibt sich das folgende Vierfelderschema:

erwartet	Hand	–Hand	total
öffentlich	150	611114	611264
–öffentlich	657596	2685947813	2686605409
total	657746	2686558927	2687216673

Abbildung 2: Errechnete Erwartungen für die Kookkurrenz *öffentlich+Hand*

Mit den beobachteten und erwarteten Werten kann nun für jede Zelle der Chi-Quadrat-Wert errechnet werden. Die Summe dieser Werte ergibt: 3.657.809,78. Diesen Wert muss man in einer Tabelle daraufhin überprüfen, welchen kritischen Wert er überschreitet, auf welchem Signifikanzniveau hier also Signifikanz vorliegt, wenn eine solche überhaupt vorliegt. Eine solche Tabelle findet man schnell in Statistiklehrbüchern oder anderen Nachschlagewerken (etwa in Albert/Koster 2002:179). Für die Überprüfung benötigt man nur die sog. Freiheitsgrade (engl. *degrees of freedom*, abgekürzt: df), „die sich wie folgt berechnen lassen: $(\text{Reihenanzahl} - 1) \times (\text{Spaltenanzahl} - 1)$. Bei einer

⁴ Vgl. <http://www.ids-mannheim.de/cosmas2/projekt/referenz/archive.html>, letzter Aufruf am 12. 03. 2012.

2×2-Tabelle beträgt df also 1“ (Bubenhofers 2009:137). Nun kann man in der entsprechenden Tabelle nachschauen, welche Schwelle der Wert überschreitet. Für df 1 sind die Werte:

0,05	3,841458821
0,01	6,634896601
0,001	10,82756617
0,0001	15,13670523

Abbildung 3: Signifikanzniveaus von Chi-Quadrat bei df 1

Der errechnete Wert übersteigt das höchste Signifikanzniveau, nämlich: 0,0001. Das bedeutet: Die Kookkurrenz von *öffentlich+Hand* ist höchst signifikant. Und das heißt: Die Nullhypothese N_0 (Die Aufeinanderfolge von *öffentlich+Hand* ist zufällig.) kann mit einer Fehlerwahrscheinlichkeit von 0,01 % verworfen werden. Ein schönes Ergebnis, oder?

4. Das Unbehagen

Etwas seltsam ist vielleicht die Höhe des Chi-Quadrat-Werts. Zwar ist die Intuition bestätigt worden, dass es sich bei *öffentlich+Hand* um eine statistisch auffällige Kookkurrenz (also eine Kollokation i. S. v. Bubenhofers) handelt, aber dass das Ergebnis so hoch ausfällt, stimmt doch ein wenig misstrauisch. Um zu überprüfen, wie aussagekräftig dieser Test ist, habe ich ihn noch für die folgenden Einheiten vor *Hand* durchgeführt: *groß*, *fein*, *leuchtend*, *alt*, *kriminell* und *laut*.⁵ Das Ergebnis ist erstaunlich: In fast allen Fällen liegt (höchste!) Signifikanz vor. Höchst signifikant ist demnach etwa die Kookkurrenz *fein+Hand*. Der Wert ist zwar sehr viel kleiner als der für *öffentlich+Hand* (die Differenz von beobachtet vs. erwartet ist im Vergleich ja auch sehr viel geringer), aber auch er überschreitet die 0,0001-Schwelle:

beobachtet	Hand	-Hand	total
fein	80	153339	153419
-fein	657666	2686405588	2687063254
total	657746	2686558927	2687216673
erwartet	Hand	-Hand	total
fein	38	153381	153419
-fein	657708	2686405546	2687063254
total	657746	2686558927	2687216673
	Chi-Quadrat	47,9963329	höchst signifikant
	df	1	

Abbildung 4: Chi-Quadrat-Test für *fein+Hand*

⁵ Wer das selbst ohne viel Berechnungsaufwand nachprüfen möchte, kann dafür übrigens ein von Bubenhofers im Netz bereitgestelltes Excel-Beispiel benutzen (zu finden auf der sehr hilfreichen Korpuslinguistik-Seite von Bubenhofers (<http://www.bubenhofers.com/korpuslinguistik/kurs/index.php?id=uebersicht.html>) unter: <http://www.bubenhofers.com/korpuslinguistik/kurs/Dokumente/VorlageSignifikanz.xls>).

Glücklicherweise ist die Kookkurenz von *leuchtend+Hand* statistisch nicht auffällig. Die beobachtete Fallzahl ist extrem gering und die Differenz zum erwarteten Wert ebenfalls. Das heißt: Man erwartet eine niedrige Anzahl des gemeinsamen Auftretens und beobachtet auch tatsächlich eine solche. Bei derart niedrigen Werten ist allerdings die Verwendbarkeit des Chi-Quadrat-Tests gefährdet. Allgemein wird gesagt, dass der Chi-Quadrat-Test spätestens ab einem Wert von 5 nicht mehr anwendbar ist.

beobachtet	Hand	-Hand	total
leuchtend	12	30015	30027
-leuchtend	657734	2686528912	2687186646
total	657746	2686558927	2687216673
erwartet	Hand	-Hand	total
leuchtend	7	30020	30027
-leuchtend	657739	2686528907	2687186646
total	657746	2686558927	2687216673
	Chi-Quadrat	2,94314998	nicht signifikant
	df	1	

Abbildung 5: Chi-Quadrat-Test für *leuchtend+Hand*

Dass allerdings *kriminell+Hand* nur knapp die erste Schwelle überschreitet, ist im Vergleich dagegen wegen der sehr geringen Trefferzahl für den Fall A (mittels &kriminell /+w1 &Hand) nicht überraschend.

beobachtet	Hand	-Hand	total
kriminell	19	48218	48237
-kriminell	657727	2686510709	2687168436
total	657746	2686558927	2687216673
erwartet	Hand	-Hand	total
kriminell	12	48225	48237
-kriminell	657734	2686510702	2687168436
total	657746	2686558927	2687216673
	Chi-Quadrat	4,38339592	signifikant
	df	1	

Abbildung 6: Chi-Quadrat-Test für *kriminell+Hand*

Es ist allerdings erstaunlich, dass bei diesen Werten auch hier überhaupt Signifikanz vorliegt. Immerhin erzielt man über &kriminell 48.237 Treffer und nur in 19 Fällen folgt danach eine Form von *Hand*. Zudem ist die Abweichung vom erwarteten Wert (12) denkbar gering.

Bei den folgenden drei Fällen ergibt sich eine, wenn man so will, negative Signifikanz: Die Erwartung ist höher als die Beobachtung. Das ist im Fall von *laut+Hand* (übrigens fast alles Fehlbelege) auch erwartbar. Aber auch im Falle von *groß+Hand*?

beobachtet	Hand	-Hand	total
groß	331	4684743	4685074
-groß	657415	2681874184	2682531599
total	657746	2686558927	2687216673
erwartet	Hand	-Hand	total
groß	1147	4683927	4685074
-groß	656599	2681875000	2682531599
total	657746	2686558927	2687216673
	Chi-Quadrat	581,454093	höchst signifikant
	df	1	

Abbildung 7: Chi-Quadrat-Test für *groß+Hand*

Dasselbe soll für den Fall *alt+Hand* gelten:

beobachtet	Hand	-Hand	total
alt	60	2625093	2625153
-alt	657686	2683933834	2684591520
total	657746	2686558927	2687216673
erwartet	Hand	-Hand	total
alt	643	2624510	2625153
-alt	657103	2683934417	2684591520
total	657746	2686558927	2687216673
	Chi-Quadrat	528,803301	höchst signifikant
	df	1	

Abbildung 8: Chi-Quadrat-Test für *alt+Hand*

Ebenso höchst negativ signifikant ist – und hier ist das Ergebnis endlich wieder einmal plausibel – der Fall bei *laut+Hand*:

beobachtet	Hand	-Hand	total
laut	18	983139	983157
-laut	657728	2685575788	2686233516
total	657746	2686558927	2687216673
erwartet	Hand	-Hand	total
laut	241	982916	983157
-laut	657505	2685576011	2686233516
total	657746	2686558927	2687216673
	Chi-Quadrat	206,118087	höchst signifikant
	df	1	

Abbildung 9: Chi-Quadrat-Test für *laut+Hand*

5. Zusammenfassung und Ausblick

Die Ergebnisse des Signifikanztests sind unbefriedigend, weil selbst bei kleinsten Abweichungen von beobachtet vs. erwartet Signifikanz vorliegen kann. Zudem sind die errechneten Werte zwar sehr hoch, unterscheiden sich aber dennoch so stark, dass zwar ein überaus deutlicher Unterschied festzustellen ist. Dieser lässt sich aber statistisch über den Chi-Quadrat-Test nicht mehr als Signifikanzprüfung einholen. So sind etwa *öffentlich+Hand* und *fein+Hand* zwar beide höchst signifikant, aber die Differenz der beiden Chi-Quadrat-Werte ist erheblich.

Das einzige, was man sinnvollerweise vielleicht machen kann, ist ein Ranking aufzustellen. Bubenhofer hat dies im Grunde bereits selbst gesehen, wenn er sagt, man könne mit dem Chi-Quadrat-Test eine Rangfolge ermitteln (s. o.).

	Wert	pos/neg	Signifikanz
öffentlich	3657809,78	positiv	höchst signifikant
fein	47,9963329	positiv	höchst signifikant
kriminell	4,38339592	positiv	signifikant
groß	581,454093	negativ	höchst signifikant
alt	528,803301	negativ	höchst signifikant
laut	206,118087	negativ	höchst signifikant
leuchtend	2,94314998	positiv	nicht signifikant

Abbildung 10: Ranking der Signifikanzwerte bei den geprüften Fällen von ADJ+*Hand*
(„positiv“ = beobachtet > erwartet, „negativ“ = beobachtet < erwartet)

Ein solches Ranking aber – etwa zwischen höchst signifikanten Ergebnissen – sollte eigentlich unnötig sein. Höchst signifikant sollte höchst signifikant sein.

Ob diese Ergebnisse nun plausibel (oder aussagekräftig) sind, insbesondere was die negativen Fälle anbetrifft, sei dahingestellt. Eines aber dürfte klar sein: Die alleinige Anwendung des Chi-Quadrat-Tests ist, auf die Menge der Ausgabewerte {nicht signifikant, signifikant, hoch signifikant,

höchst signifikant} bezogen, hier (also für die Prüfung von Kookkurrenzen auf ihre statistische Auffälligkeit) nicht sinnvoll. Denn selbst größte Unterschiede werden nivelliert und selbst kleinste Abweichungen sind höchst signifikant. Man muss die einzelnen errechneten Werte für die positiven Fälle mindestens noch miteinander vergleichen, um Aussagen machen zu können, wie stark die Bindekraft der beiden Teile der Kookkurrenz nun eigentlich gegenüber anderen Kookkurrenzen ist. Dies könnte man auch über die Berechnung des Phi-Wertes (vgl. Bubenhofer 2009:138) tun:

	Wert	pos/neg	Signifikanz	Phi
öffentlich	3657809,78	positiv	höchst signifikant	0,036894295
fein	47,9963329	positiv	höchst signifikant	0,000133645
kriminell	4,38339592	positiv	signifikant	0,000040388
leuchtend	2,94314998	positiv	nicht signifikant	0,000033094

Abbildung 11: Ranking der Chi-Quadrat-Signifikanzwerte in den positiven Fällen incl. Phi-Wert

Dieser drückt die Assoziationsstärke der Kollokation in einem standardisierten Maß aus und wird errechnet über die Wurzel aus (Chi-Quadrat geteilt durch die Gesamtanzahl). Auch hier ist aber m. E. nur ein Vergleich sinnvoll, denn der absolute Wert für *öffentlich+Hand* ist wiederum so niedrig, dass es fraglich ist, ob man überhaupt eine Kollokation annehmen kann. Aber das ist ein neues Kapitel, das in diesem Aufsatz nicht mehr aufgeschlagen wird. Es ging mir hier nur darum zu zeigen, dass die Anwendung eines bestimmten Signifikanztests in einem bestimmten Fall – und sei es der Paradefall der Anwendung dieses Tests – durchaus in die Irre führen kann. Zurückkommend auf den Untertitel sollte also *eines* nicht als unbestimmter Artikel, sondern als Zahladjektiv verstanden werden. Und *in der Korpuslinguistik* ist nicht als allgemein gedacht, sondern nur als auf diesen Fall bezogen. In anderen korpuslinguistischen Zusammenhängen mag eine einzelne Signifikanzprüfung über den Chi-Quadrat-Test sinnvoll sein. Bei Kollokationen aber nicht.

Literaturverzeichnis

- ALBERT, Ruth/KOSTER, Cor J. (2002): *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch*. Tübingen.
- BUBENHOFER, Noah (2009): *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin; New York.
- FILLMORE, Charles J. (1992): „Corpus linguistics“ or „Computer-aided armchair linguistics“. In: SVARTVIK, Jan (Hrsg.): *Directions in Corpus Linguistics. Proceedings of the Nobel Symposium 82, Stockholm, 4.–8. August 1991*. Berlin; New York, S. 35–60.
- GRIES, Stefan Th. (2008): *Statistik für Sprachwissenschaftler*. Göttingen.
- HAUSER, Bernhard/HUMPERT, Winfried (2009): *signifikant? Einführung in statistische Methoden für Lehrkräfte*. Zug.
- MEINDL, Claudia (2011): *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung*. Tübingen.
- SPITZMÜLLER, Jürgen/WARNKE, Ingo H. (2011): *Diskurslinguistik. Eine Einführung in Theorien und Methoden der transkulturellen Sprachanalyse*. Berlin; Boston.

- STAFFELDT, Sven (2011): Die phraseologische Konstruktionsfamilie [X Pröp *Hand* Verb]. In: *ZGL* 39, S. 188–216.
- STEYER, Kathrin (2002): Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotential der korpusbasierten Kookkurrenzanalyse. In: HASS-ZUMKEHR, Ulrike/KALLMEYER, Werner/ZIFONUN, Gisela (Hrsg.): *Ansichten der deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag*. Tübingen, S. 215–236.