

# Proseminar Maschinelles Lernen und Experimentelles Design, UdS, SS11 Kurzinfo zum richtigen Zitieren

Caroline Sporleder

7. Juli 2011

Dieses Dokument gibt nur einige kurze (und z.T. fachspezifische) Hinweise zum korrekten Zitieren. Es gibt zahlreiche gute Ratgeber zum wissenschaftlichen Arbeiten, die weiterführende Informationen enthalten. Zum Beispiel:

Susanne Kassel, Martina Thiele und Margit Böck (2006): Zitieren in wissenschaftlichen Arbeiten.<sup>1</sup>

Um mehr über die Gepflogenheiten unseres Fachs zu erfahren, schaut euch an, wie in guten computerlinguistischen Veröffentlichungen typischerweise zitiert wird.

## 1 Zitieren im Text

- in der CL wird die Quelle normalerweise **im Text** angegeben (nicht in einer Fußnote). Z.B. “As Smith (2000) showed, . . .” oder “Recent corpus studies have shown that the principles of Centering Theory are not always adhered (Smith 2008, Brown 2009)”.
- Quellenangaben sind notwendig für alle Sachverhalte, die nicht gedankliches Allgemeingut sind. Allgemeingut sind z.B. Aussagen wie “Human language is a complex system.” oder “Syntactic parsers analyse sentences and determine their grammatical structure.” Kein Allgemeingut sind u.a.:
  - direkte, wörtliche Zitate
  - indirekte Übernahme von fremdem Gedankengut
  - Forschungsergebnisse anderer Wissenschaftler

---

<sup>1</sup><http://www.daf.tu-berlin.de/fileadmin/fg75/PDF/Zitieren.pdf>

## 1.1 Direkte Zitate

In der CL **möglichst vermeiden!** Eigentlich nur sinnvoll bei:

- Definitionen
- Aussagen, die in irgendeiner Form denkwürdig, zweifelhaft etc. sind. Z.B.:

Smith (2000, p. 15) claims that his method “has, in practice, been found to be superior to all other methods that have been proposed so far”.

### Grundregeln:

- Das Zitat muss in doppelte Anführungszeichen gesetzt oder durch Einrückung etc. klar als wörtliches Zitat gekennzeichnet sein.
- Das Zitat muss von einer Referenz begleitet sein, die besteht aus:
  - dem Nachnamen des Autors (bzw. der Autoren oder des Erstautors plus “et al.”)
  - der Jahreszahl und
  - der Angabe der Seite, auf der das Zitat im Originaltext steht.  
**(Die Seitenangabe muss bei direkten Zitaten immer dabei stehen!)**
- Das Zitat muss mit dem Original übereinstimmen, sowohl bezüglich des Inhalts als auch bezüglich der Orthographie und des Satzbaus (Hervorhebungen, Tippfehler etc. müssen beibehalten werden). Wenn Abweichungen vom Original sinnvoll oder notwendig sind, müssen diese eindeutig gekennzeichnet sein.

Achtung: Um korrekt zitieren zu können, müß ihr **sicherstellen, dass euch tatsächlich die Originalquelle** in ihrer maßgeblichen Fassung **vorliegt**. Vorsicht ist geboten bei Publikation, die z.B. von dem Homepages ihrer Verfasser heruntergeladen werden. Besonders bei Zeitschriftenaufsätzen handelt es sich hier oft um Vorabdrucke (sogenannte “Preprints”), in denen der Text und die Seitenangaben mehr oder weniger leicht von der gedruckten Fassung abweichen können. Wenn ihr direkt zitieren wollt, besorgt euch die Originalquelle (aus der Bib oder wenn elektronisch, dann direkt von der Webseite der entsprechenden Zeitschrift bzw. aus der ACL Anthology etc.).

## 1.2 Indirekte Zitate

Indirekte Zitate kommen sehr häufig vor und geben den Inhalt nicht wörtlich, sondern nur **sinngemäß** (umschreibend) wieder. Indirekte Zitate müssen sich in der Formulierung **deutlich von der Originalquelle unterscheiden**, ein einfaches Ersetzen einzelner Wörter reicht nicht. (Der Begriff 'Zitat' ist daher etwas irreführend.) Im Gegensatz zu direkten Zitaten:

- werden indirekte Zitate nicht durch Anführungszeichen etc. hervorgehoben
- muss bei indirekten Zitaten keine Seitenzahl in der Referenz stehen (kann aber, wenn dies sinnvoll erscheint).

Es ist auch möglich mehrere Quellen in einem Satz zusammenzufassen. Z.B.: "Recent corpus studies have shown that the principles of Centering Theory are not always adhered to in coherent texts (Smith 2008, Brown 2009)".

## 1.3 Besondere Quellen

### 1.3.1 Lehrbücher

Lehrbücher wie Jurafsky und Martin (2009)<sup>2</sup> sollten nur in Ausnahmefällen zitiert werden (bei Proseminararbeiten noch gerade ok, bei Masterarbeiten eher vermeiden). Wenn ihr auf spezielle CL-Methoden verweisen wollt, solltet ihr nach Möglichkeit die Original- oder Standardquelle ausfindig machen. Z.B. kann man für Centering Theory Grosz et al. (1995)<sup>3</sup> zitieren. Wenn ihr nicht sicher seid, welches die Originalquelle ist, schaut euch Publikationen an, die sich auf die Methode beziehen, und seht nach, welche Quelle(n) dort genannt wird. Viele Methoden, Techniken können zudem als Standard vorausgesetzt werden und benötigen daher nicht unbedingt eine gesonderte Referenz, z.B. TF.IDF. Als Faustregel: Wenn sich noch ein eindeutiger Urheber ermitteln lässt (wie bei Centering, TextTiling etc.), besser eine Quelle angeben, sonst nicht.

### 1.3.2 Sekundärquellen

Lehrbücher sind ein spezieller Fall von Sekundärquellen, d.h. Quellen, die andere Quellen lediglich zusammenfassen und aufbereiten. Sekundärquellen sind in der CL generell zu vermeiden. D.h. wenn Brown (2010) auf eine andere Quelle, Smith (2005), verweist, solltet ihr direkt auf Smith (2005) verweisen. Dies setzt voraus,

<sup>2</sup>Jurafsky, D. und J. Martin (2009): Speech and Language Processing.

<sup>3</sup>Grosz, B. J., A. K. Joshi und S. Weinstein (1995): Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21:2, S. 202-225.

dass ihr euch Smith (2005) im Original besorgt und lest! Ihr dürft euch nicht auf die Sekundärquelle, Brown (2010), verlassen. Eine Ausnahme von dieser Regel gibt es nur, wenn sich Smith (2005) partout nicht im Original auftreiben läßt (nicht in der Bib, nicht über Fernleihe, nicht im Internet, nicht durch direktes Kontaktieren des Autors etc.). In diesem Fall dürft ihr die Sekundärquelle verwenden, müßt jedoch im Text deutlich machen, dass euch die Primärquelle nicht vorlag. Z.B. “Smith (2005, S. 10) zitiert nach Brown (2010, S. 46)”.

### 1.3.3 Korpora, Software und Internetquellen

Als Quelle für Korpora und Software solltet ihr nach Möglichkeit eine Publikation angeben (nicht nur die URL der entsprechenden Webseite!). Um die korrekte Publikation ausfindig zu machen, geht ihr wie folgt vor:

- Schaut auf der Webseite für das Korpus/die Software nach. Dort wird oft angegeben, auf welche Publikation verwiesen werden sollte. Z.B. gibt es auf der Webseite des Stanford Parsers<sup>4</sup> einen Abschnitt “Citing the Stanford Parser”, in dem genau aufgelistet ist, welche Publikationen für welche Aspekte des Parsers zitiert werden sollten.
- Wenn die Webseite nicht weiterhilft, schaut euch die Dokumentation an, die mit der Software/den Daten ausgegeben wird. Oft finden sich z.B. Hinweise in der README Datei oder in der Dokumentation selber.
- Falls dies auch nicht hilft, findet heraus, wie andere Wissenschaftler auf diese Software / dieses Korpus verweisen.

Wenn ihr keine Publikation find könnt (manchmal gibt es keine) oder wenn ihr zusätzlich zur Publikation noch die Webseite angeben wollt, könnt ihr dies als Fußnote im Text machen (nicht die URL im Literaturverzeichnis auflisten).

In sehr, sehr seltenen Fällen kann es sein, dass ihr auf eine Internetquelle im Literaturverzeichnis verweisen müsst, z.B. weil ein richtungsweisender Aufsatz nie veröffentlicht wurden ist oder weil ihr auf unveröffentlichte Annotationsrichtlinien etc. verweisen wollt. In diesem Fall solltet ihr soviel Information wie möglich ins Literaturverzeichnis übernehmen, insbesondere Autor, Titel und Entstehungsdatum. Falls sich diese Informationen ganz oder teilweise nicht finden lassen, ist die Quelle ohnehin zu unzuverlässig, um zitiert zu werden. Zusätzlich solltet ihr dann noch die URL angeben, mit einem Verweis darauf, wann ihr sie zum letzten Mal besucht habt.

---

<sup>4</sup><http://nlp.stanford.edu/software/lex-parser.shtml>

## 2 Das Literaturverzeichnis

Die Quellen im Literaturverzeichnis sollten geordnet sein, entweder alphabetisch nach dem Nachnamen des Erstautors oder numerisch danach, wenn die Quelle zuerst im Text erwähnt wurde. Es gibt verschiedene Möglichkeiten, die Quellenangaben zu formatieren und die Informationen anzuordnen. Wichtig ist:

- Die Angaben im Literaturverzeichnis müssen konsistent sein.
- Alle notwendigen Angaben müssen vorhanden sein.

Hier ist ein kurzer Überblick, wie die einzelnen Publikationsformen behandelt werden sollten:

- **Monographien:** Nachname, Vorname (Jahr): Titel (plus Untertitel, wenn vorhanden). Erscheinungsort: Verlag.

Kamp, H. und U. Reyle (1993): *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht, The Netherlands: Kluwer.

Die Angabe von Erscheinungsort und Verlag ist optional (wenn, dann aber beide). In manchen Geisteswissenschaften ist diese Angabe verpönt, in der CL aber durchaus Standard.

- **Sammelbände:** Nachname, Vorname des Herausgebers (Hg.|Ed.) (Jahr): Titel (plus Untertitel, wenn vorhanden). Erscheinungsort: Verlag.

Fellbaum, C. (Hg.) (1998): *WordNet: An Electronic Database*. Cambridge, MA: MIT Press.

Durch “Hg.” (dt.: Herausgeber) bzw. “Ed.” (engl.: Editor) wird angezeigt, dass es sich um einen Sammelband handelt. Dies ist wichtig.

- **Beiträge in Sammelbänden:** Name, Vorname des Autors/der Autoren (Jahr): Titel. Untertitel des Beitrags. In: Name, Vorname der HerausgeberInnen des Sammelwerks (Hg.—Ed.): Titel. Untertitel des Sammelwerks. Erscheinungsort: Verlag, Seiten (die erste und letzte Seite des Beitrags).

Fellbaum, C. (1998): A Semantic Network of English Verbs. In: Fellbaum, C. (Hg.): *WordNet: An Electronic Database*. Cambridge, MA: MIT Press, S. 69-104.

Wichtig: Es wird zuerst der Beitrag selber genannt, dann das Sammelwerk.  
Die Seitenangabe ist unbedingt notwendig.

- **Beiträge in Fachzeitschriften:** Nachname, Vorname des/der Verfasser(s) (Jahr): Titel (plus Untertitel, wenn vorhanden). In: Zeitschrift Jahrgang:Heftnummer, Seitenangabe.

Grosz, B. J., A. K. Joshi und S. Weinstein (1995): Centering: A framework for modeling the local coherence of discourse. In: *Computational Linguistics*, 21:2, S. 202-225.

Die Seitenangabe ist auch hier obligatorisch. (Auch Aufsätze in rein elektronischen Zeitschriften, haben i.d.R. Seitenangaben.)

- **Konferenz- oder Workshopbeiträge:** Nachname, Vorname des/der Verfasser(s) (Jahr): Titel (plus Untertitel, wenn vorhanden). In: Proc. of Name des Workshops/der Konferenz, Seitenangabe.

Klein, D. und C. D. Manning (2003): Accurate Unlexicalized Parsing. In: *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, S. 423-430.

- Optional kann zwischen Konferenz-/Workshopnamen und Seitenangabe noch der Ort und das Datum der Veranstaltung angegeben werden.
- Die Seitenangabe ist nicht immer zu ermitteln und kann daher weggelassen werden.
- Der Name der Veranstaltung kann ausgeschrieben oder abgekürzt werden (besonders bei bekannten Konferenzen). Also: *Proc. of ACL 2010* oder *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Achtet aber auf eine einheitliche Handhabung.

- **Unveröffentlichte Doktor-, Masterarbeiten etc.:** Nachname, Vorname des/der Verfasser(s) (Jahr): Titel (plus Untertitel, wenn vorhanden). Art der Abschlussarbeit. Universität.

Collins, M. (1999): *Head-Driven Statistical Models for Natural Language Parsing*. Doktorarbeit, University of Pennsylvania.

Optional kann vor der Universität noch die Fachrichtung mit angegeben werden.

- **Technical Reports:** Nachname, Vorname des/der Verfasser(s) (Jahr): Titel (plus Untertitel, wenn vorhanden). Technical Report ID-Nummer des Reports. Institut, Ort.

Mann, W. C. und S. A. Thompson (1987): *Rhetorical structure theory: A theory of text organization*. Technical Report ISI/RS-87-190, ISI, Los Angeles, CA.

Die ID-Nummer ist optional, sollte aber nach Möglichkeit angegeben werden.

**Ermittlung von Seitenangaben:** Die Ermittlung von Seitenangaben ist bei elektronischen Workshop- und Konferenzbeiträgen nicht immer ganz einfach. Hier einige Tipps:

- Schaut direkt im elektronischen Originaldokument nach (d.h. dem Dokument aus der ACL Anthology — die Dateien auf den Webseiten der Verfasser enthalten oft keine Seitenzahlen).
- Schaut im elektronisch verfügbaren Inhaltsverzeichnis der Veranstaltung nach (unter der sog. “Front Matter” z.B. in der Anthology verfügbar).
- Falls beides nicht hilft: Schaut auf der Webseite der Autoren nach (Publikationsliste), ob dort die Seitenangaben vorhanden sind.
- Falls das auch nicht hilft oder falls ihr widersprüchliche Angaben bekommt, lässt man die Seitenangaben besser weg. (Gilt aber nicht für Zeitschriftenaufsätze!)