

Wie schreibt man eine Hausarbeit?

**Leitfaden zur Erstellung einer Hausarbeit
im Fachbereich Computerlinguistik**

Prof. Manfred Pinkal Andrea Horbach

Universität des Saarlandes
Fachrichtung 4.7 Allgemeine Linguistik
Oktober 2012

Inhaltsverzeichnis

1. Über diesen Leitfaden	3
2. Vor dem Schreiben	3
2.1. Sinn und Zweck einer Hausarbeit	3
2.2. Adressat der Hausarbeit	3
2.3. Inhalt und Textart	3
2.4. Arbeitsschritte	4
2.5. Gliederung	4
2.6. Literaturrecherche	5
3. Während des Schreibens	5
3.1. Aufbau	5
3.2. Formatierung	6
3.3. Beispiele	6
3.4. LaTeX	7
3.5. Grafische Elemente: Tabellen, Formeln, Grafiken etc.	7
3.6. Schreibstil	8
3.6.1. In der ersten Person schreiben?	8
3.6.2. Unterschied zwischen der eigenen Perspektive und der eines anderen Autors	8
3.6.3. Kohärenz – Der rote Faden	9
3.6.4. Terminologie	9
3.6.5. Fußnoten	9
3.7. Wenn Sie nicht weiterkommen	10
4. Referenzen	10
4.1. Zitieren im Text	10
4.1.1. Direkte Zitate	11
4.1.2. Indirekte Zitate	11
4.1.3. Besondere Quellen	12
4.2. Das Literaturverzeichnis	13
5. Nach dem Schreiben	15
A. Nützliche Latex-Pakete	17
B. Weitere nützliche Tools	17

Dieser Leitfaden basiert auf einer Vorlage von Barbara Stiebels und Manfred Pinkal. Das Kapitel zum richtigen Zitieren basiert auf einer Vorlage von Caroline Sporleder. Außerdem verwendet der Leitfaden Ideen aus einem Vortrag von Ivana Kruijff-Korbayová.

1. Über diesen Leitfaden

Dieser Leitfaden soll Ihnen als Computerlinguistik-Studierenden dabei helfen, eine Hausarbeit zu einem (Pro-)Seminar zu schreiben. Wir möchten Ihnen Hilfestellung bei der formalen Gestaltung geben und Sie beim Schreibprozess unterstützen.

2. Vor dem Schreiben

2.1. Sinn und Zweck einer Hausarbeit

Eine Hausarbeit ist eine Arbeit, die Sie als Prüfungsleistung (meist auf Grundlage eines Referates) zu einem Seminar oder Proseminar schreiben. Die Hausarbeit soll einerseits zeigen, dass Sie Ihr Thema verstanden haben und seine Kern-Aussagen und Fragestellungen nachvollziehen und wiedergeben können, andererseits sollen Sie in der Hausarbeit das wissenschaftliche Schreiben in der Computerlinguistik üben.

2.2. Adressat der Hausarbeit

Ihr Dozent ist derjenige, der Ihre Hausarbeit lesen und bewerten wird. Bei der Frage, welche fachlichen Grundlagen Sie in Ihrer Hausarbeit voraussetzen können, welche Fachbegriffe Sie erläutern sollten usw. ist die Vorstellung, dass Sie an Ihren Dozenten gerichtet schreiben, aber wenig hilfreich. Er oder sie kennt das Thema natürlich, so dass Sie unter dem Gesichtspunkt der Wissensvermittlung die gesamte Hausarbeit nicht zu schreiben bräuchten. Stellen Sie sich besser vor, dass Sie Ihre Arbeit an Kommilitonen aus der Computerlinguistik richten, die das Seminar nicht besucht haben. Sie dürfen also beispielsweise grundlegende Fachtermini der Computerlinguistik als bekannt voraussetzen (z.B. „Was ist eine Nominalphrase?“), müssen aber die themenspezifische Terminologie erklären.

2.3. Inhalt und Textart

Machen Sie sich ganz am Anfang klar, was Sie schreiben. Ist Ihre Hausarbeit zum Beispiel eine zusammenfassende Darstellung eines bestimmten Themas oder Artikels, ein Vergleich zwischen verschiedenen Ansätzen, ein Bericht zu einer eigenen praktischen Arbeit etc.?

Grenzen Sie Ihre Arbeit thematisch ein. Meistens werden Sie nicht alle Aspekte der von Ihnen bearbeiteten Papiere in gleicher Weise in Ihre Hausarbeit einfließen lassen. Machen Sie sich klar, worüber Sie schreiben möchten und worüber Sie explizit nicht schreiben

und welche Grundlagen Sie dazu voraussetzen. Sprechen Sie diese Überlegungen so früh wie möglich mit Ihrem Dozenten ab.

2.4. Arbeitsschritte

1. Thema eingrenzen, Fragestellung festlegen
2. zusätzliche Literatur finden und lesen
3. Gliederung, grober Draft, z.B. Inhaltsverzeichnis mit Unterüberschriften und jeweils Stichpunkten, Argumentation skizzieren
4. Hauptteil schreiben
5. Hauptteil überarbeiten, umstellen, zusammenstreichen und erweitern
6. Einleitung und Schluss erst nach dem Hauptteil schreiben
7. Korrekturlesen: Schreibfehler, Konsistenz etc. (→ nach dem Schreiben)

2.5. Gliederung

Typischerweise besteht Ihre Hausarbeit aus den folgenden inhaltlichen Teilen:

- In der Einleitung geben Sie das Thema Ihrer Arbeit an und erklären es kurz: Mit welcher Fragestellung haben Sie sich beschäftigt? Erläutern Sie Ihre Fragestellung kurz (hier kann gerne schon ein kurzes anschauliches Beispiel kommen). Auf welche Literatur beziehen Sie sich hauptsächlich? Was ist Ihre Motivation, d.h. warum ist Ihre Fragestellung interessant und/oder wichtig? Wie gehen Sie bei der Bearbeitung Ihrer Fragestellung vor? Geben Sie am Schluss Ihrer Einleitung eine kurze Zusammenfassung, was sie in jedem einzelnen Kapitel der Arbeit tun werden. Jemand, der nur die Einleitung Ihrer Hausarbeit liest, sollte danach verstanden haben, worum es in Ihrer Arbeit geht.
- Im Hauptteil Ihrer Arbeit stellen Sie die eigentlichen Inhalte dar. Unterteilen Sie Ihren Hauptteil in mehrere (vorläufige) Kapitel und machen Sie sich klar, wie Sie Ihre Argumentation aufbauen möchten, so dass ein „roter Faden“ in Ihrer Arbeit erkennbar ist. Wenn Sie während des Schreibens merken, dass Ihre ursprünglich geplante Gliederung nicht mehr mit Ihrer tatsächlichen Arbeit übereinstimmt, können Sie Ihre Kapiteleinteilung immer noch ändern.
- Im Schlussteil sollten Sie eine Zusammenfassung Ihrer Hausarbeit geben: Welches Problem/welche Fragestellung haben Sie besprochen. Mit welchen Methoden und Ansätzen wurde es gelöst oder bearbeitet? Wo sind Grenzen des diskutierten Ansatzes, Kritikpunkte, ungelöste Probleme? Geben Sie auch einen Ausblick, welche weiteren Fragestellungen sich aus dem Inhalt Ihrer Darstellung ergeben oder

welche neueren Forschungsansätze es zu diesem Thema gibt. Im Fall einer empirischen Arbeit gehört die Präsentation und Diskussion der Ergebnisse bereits in den Hauptteil und nicht in den Schluss.

2.6. Literaturrecherche

Ausgangspunkt für Ihre Hausarbeit sind der oder die Artikel, die Sie im Seminar für Ihr Referat gelesen haben. Damit sollten Sie sich nicht zufrieden geben. Um weitere relevante Literatur zu Ihrem Thema zu finden, haben Sie die folgenden Möglichkeiten:

- Die Artikel verweisen auf andere Literatur: Hier können Sie nachlesen, auf welchen vorherigen Ansätzen die Artikel, die Sie besprechen, aufbauen. Diese Artikel liefern Ihnen also Hintergrundinformationen zu ihrem Thema, beispielsweise zu anderen Ansätzen zur gleichen Fragestellung oder zu theoretischen Grundlagen.
- Finden Sie möglichst aktuelle Überblicksartikel zu Ihrem Thema, die auf andere spezifischere Artikel verweisen. Wenn Sie keinen Überblicksartikel finden, ist ein Lehrbuchkapitel auch ein möglicher erster Ausgangspunkt, z.B. ein entsprechendes Kapitel bei Jurafsky & Martin [JM08] oder Carstensen et al. [CEJ⁺09]. Aber denken Sie daran, dass Sie solche Lehrbuchkapitel nur im Ausnahmefall als Quelle zitieren dürfen (siehe Kapitel zu Referenzen) und im Normalfall nur als Grundlage für die weitere Literaturrecherche benutzen sollten.
- Suchen Sie über Schlagwörter nach verwandten Papieren. Dazu können Sie beispielsweise die Artikelsuche der Coli-Bibliothek, die ACL Anthology, DBLP, Cite-seer, Google Scholar oder Google benutzen.
- Schauen Sie sich an, in welchen aktuellen Veröffentlichungen auf die Artikel, die Sie bearbeiten, verwiesen wird (z.B. mit Hilfe von Google Scholar). Dieses Vorgehen hilft Ihnen dabei, neuere Ansätze zu diesem Thema zu finden, die auf Ihren Artikeln aufbauen.
- Sie können natürlich auch Ihren Dozenten fragen, aber bitte erst, wenn Sie die oben genannten Möglichkeiten genutzt haben.

Wenn Sie Artikel über das Internet suchen, kann es Ihnen passieren, dass Sie manche Artikel nur dann abrufen können, wenn Sie über die Uni eingeloggt sind. Die Saarländische Landesbibliothek hat Lizenzen mit manchen Verlagen, die es ermöglichen, auch nicht frei zugängliche Artikel abzurufen.

3. Während des Schreibens

3.1. Aufbau

Ihre Arbeit sollte die folgenden formalen Bestandteile haben (in der Reihenfolge wie gelistet): Deckblatt, Inhaltsverzeichnis, Abbildungsverzeichnis (wenn nötig), der Inhalt

(bestehend aus Einleitung, Hauptteil, Schluss), Literaturverzeichnis, evtl. Anhang. Jeder dieser Bestandteile sollte mit einer neuen Seite beginnen.

1. Deckblatt: Auf dem Deckblatt der Arbeit sollen die folgenden Informationen enthalten sein: Ihr Name, Ihre Matrikelnummer und coli-Emailadresse, der Titel Ihrer Hausarbeit, der Name der zugehörigen Veranstaltung und des Dozenten und eine Semesterangabe. Geben Sie Ihrer Hausarbeit einen Titel, der aussagt, worüber Sie in Ihrer Arbeit schreiben. Es reicht nicht, einfach den Titel des Artikels, über das Sie im Wesentlichen schreiben, als Titel zu übernehmen.
2. Inhaltsverzeichnis: Gliedern Sie Ihre Arbeit so, dass der Aufbau schon beim Lesen des Inhaltsverzeichnisses deutlich wird, aber gliedern Sie nicht zu feinkörnig. (2 bis 3 Gliederungsebenen sind ein guter Richtwert.) Geben Sie im Inhaltsverzeichnis die Überschriften genau so an, wie sie auch tatsächlich in Ihrer Arbeit stehen.
3. Abbildungsverzeichnis: Listen Sie alle Abbildungen, Diagramme usw. mit der jeweiligen Bildunterschrift auf.
4. Hauptteil: Der eigentliche inhaltliche Teil Ihrer Arbeit. Die einzelnen Bestandteile (Einleitung usw.) sollten nicht durch Seitenumbrüche getrennt sein.
5. Literaturverzeichnis: Dazu weiter unten mehr.
6. Anhang: Alles, was den Lesefluss in der Arbeit stören würde (größere Tabellen, lange Codebeispiele usw.), gehört in den Anhang.

3.2. Formatierung

Falls Ihr Dozent genaue Vorstellungen von Schriftgröße, Schriftart oder Zeilenabstand hat, wird er Ihnen dies mitteilen. Typischerweise gilt: Wählen Sie eine Schriftgröße zwischen 10 und 12 pt und einfachen Zeilenabstand. Bleiben Sie bei Ihrer einmal getroffenen Entscheidung und wechseln Sie die Formatierung nur, wenn es einen guten Grund dafür gibt. Hervorhebungen sollten Sie sparsam und konsistent einsetzen, also beispielsweise neu eingeführte Begriffe immer kursiv schreiben (aber nicht manchmal fettgedruckt). Unterstreichungen sollten Sie nur ausnahmsweise benutzen. Wenn es von Seiten Ihres Dozenten eine Vorgabe zur Anzahl der Seiten gibt, dann zählt dafür normalerweise die Seitenzahl im inhaltlichen Teil, also ohne Deckblatt, Inhaltsverzeichnis usw.

3.3. Beispiele

Ein Beispiel sagt oft mehr als tausend Worte. Geben Sie deshalb, wann immer es sich anbietet, passende Beispiele an, um Ideen, Begriffe oder Algorithmen zu veranschaulichen. Sie können die Beispiele, die in einem Papier genannt werden, übernehmen (und müssen Sie dabei natürlich zitieren). Versuchen Sie, wenn möglich, aber immer auch eigene Beispiele zu finden, die in Ihrem Kontext vielleicht sogar besser passen als die des Autors. Wenn Sie auf Deutsch über einen Artikel in englischer Sprache schreiben (und es

nicht um ein typisches Phänomen des Englischen geht) versuchen Sie deutsche Beispiele zu finden. Wenn Sie das englische Beispiel des Originalpapiers ins Deutsche übertragen, muss dies als indirektes Zitat kenntlich gemacht werden.

Beispielsätze sollten Sie im Text absetzen und fortlaufend durchnummerieren. Bei fremdsprachlichen Beispielen hilft eine Interlinearübersetzung. Ungrammatische Sätze werden mit einem Stern kenntlich gemacht. Kürzere Sprachbeispiele im laufenden Text (z.B. einzelne Wörter) setzen Sie entweder in doppelten Anführungszeichen oder kursiv. Passen Sie im letzteren Fall auf, dass Sie dann Kursivschrift nicht gleichzeitig zum Hervorheben neuer Begriffe verwenden.

Auch andere längere Beispiele (z.B. Beispielrechnungen) sollten Sie im Text absetzen und nummerieren.

3.4. LaTeX

Latex ist ein Textsatzprogramm, das eine Alternative zu beispielsweise Word darstellt und dessen Benutzung wir Ihnen empfehlen.

Latex bietet Ihnen insbesondere dadurch Vorteile, dass es vieles automatisch einheitlich erledigt, was Sie in Word manuell tun müssten und was oft eine Quelle für Inkonsistenzen darstellt. Dazu zählen insbesondere die Erstellung des Inhaltverzeichnis und der Literaturliste (mit Bibtex), die einheitliche Formatierung von Überschriften und Aufzählungen und die konsistente Durchnummerierung von Formeln, Tabellen usw.

Latex bietet Ihnen außerdem die Möglichkeit, grafische und formale Elemente wie Baumstrukturen, Formeln oder AVMs zu erstellen. Eine Liste der dazu geeigneten Latex-Pakete finden Sie im Anhang A. Ein gutes Einführungsbuch ist außerdem Das LaTeX-Handbuch von Leslie Lamport [Lam95]

Wenn Sie Latex benutzen, werden auch viele Formatierungsaspekte automatisch in einer meist brauchbaren Weise geregelt. Es ist also oft nicht nötig (und führt dann auch zu deutlich weniger schönen Ergebnissen), dass Sie z.B. Ränder manuell ändern. Allgemein lassen sich in LaTeX unkompliziert Texte mit einem professionellen Layout erstellen.

In der Computerlinguistik ist LaTeX so weit verbreitet, dass viele Konferenzen Beiträge nur in LaTeX-Format akzeptieren. Es hilft Ihnen also später, wenn Sie sich frühzeitig damit vertraut machen.

3.5. Grafische Elemente: Tabellen, Formeln, Grafiken etc.

Tabellen, Grafiken, Formeln usw. helfen dabei, Sachverhalte anschaulich zu präsentieren. Setzen Sie diese Gestaltungsmöglichkeiten gezielt und zur Verdeutlichung, aber nicht als Ersatz für Ihren Text ein. Sie sollten also immer auch im Text erklären, was z.B. die Daten in einer Tabelle repräsentieren. Fassen Sie dabei die wesentlichen Erkenntnisse oder Ergebnisse, die in den Daten deutlich werden, zusammen. Sie müssen und sollen nicht die Werte jeder einzelnen Zeilen einer Tabelle paraphrasieren.

- Tabellen: durchnummerieren (Tabelle 1...), (Unter-)Titel und kurze Erklärung, sehr große Tabellen in den Anhang
- Grafiken: durchnummerieren, Titel geben, in Abbildungsverzeichnis listen.
- Charts: zur Veranschaulichung statistischer Daten, z.B. Evaluationsergebnisse
- Listen: Listen sind nur dann sinnvoll, wenn Sie mehrere (> 2) gleichartige Objekte anordnen wollen. Wenn eine Rangordnung oder Reihenfolge besteht, benutzen Sie eine nummerierte Liste, sonst eine nichtnummerierte (wie diese hier). Versuchen Sie sehr lange Listen (ungefähr > 10 Elemente) aufzuspalten oder weiter zu untergliedern. Sätze beginnen auch in Listen mit einem Großbuchstaben und enden mit einem Punkt am Ende.
- Auch Formeln sollten Sie im Text absetzen und durchnummerieren, so dass Sie sich im Text darauf beziehen können. Haben Sie mehrere Gleichungen, dann sollten Sie am Gleichheitszeichen ausgerichtet werden.
- Code: Algorithmen werden oft in Pseudocode angegeben. Überlegen Sie bei „echtem“ Programmcode immer, ob er tatsächlich angegeben werden sollte. Meist bietet es sich an, längere Code-Abschnitte, wenn überhaupt, gesondert in einem Anhang anzugeben.

3.6. Schreibstil

3.6.1. In der ersten Person schreiben?

Wenn Sie Ihre eigene Arbeit, Ideen oder Bewertungen darlegen, sollte dies in der Arbeit immer deutlich zu erkennen sein. Es ist dabei Ihnen überlassen, ob Sie in der Ich-Form oder in der Wir-Form schreiben. Die Wir-Form soll meist sowohl nicht explizit genannte Vordenker einschließen, als auch den Leser in die Perspektive des Autors miteinbeziehen und ist vor allem im Englischen sehr verbreitet. Eine dritte Möglichkeit ist die unpersönliche Schreibweise, bei der durch Passiv und andere Konstruktionen die Verwendung von „ich“ oder „wir“ komplett vermieden wird. Wenn Sie ausschließlich diese Ausdruckweise verwenden, klingt das meistens künstlich, daher empfehlen wir sie nicht zum alleinigen Gebrauch. Es spricht nichts dagegen, trotzdem gelegentlich unpersönliche Ausdrücke zu verwenden („Im folgenden Beispiel sieht man, das...“), solange Sie es damit nicht übertreiben. Ansonsten bleiben Sie bei der einmal gewählten Variante, also entweder nur „ich“ oder nur „wir“ aber nicht gemischt.

3.6.2. Unterschied zwischen der eigenen Perspektive und der eines anderen Autors

Unterscheiden Sie immer deutlich zwischen Ihrer eigenen Perspektive und der des Autors, über den Sie schreiben, vor allem in zeitlicher Hinsicht. Deiktische Ausdrücke wie „in letzter Zeit“ oder „der aktuelle Forschungsstand“ sind immer im Entstehungszeitraum

des Papiers zu sehen. Wenn ein Autor beispielsweise in den 90ern geschrieben hat, dass sein Ansatz besser als alle bisherigen Ansätzen ist, dürfen Sie dies in Ihrer Arbeit nicht einfach auf die Gegenwart übertragen, sondern könnten etwa schreiben, dass der Ansatz XY im Jahr 1997 allen anderen Ansätzen überlegen war.

Auch bei der Übernahme wertender Aussagen von anderen Autoren müssen sie aufpassen. Wenn Sie die Bewertung unkommentiert übernehmen, bedeutet das, dass Sie mit ihr einverstanden sind. Wenn Ihre eigene Meinung von der des Autors abweicht, machen Sie dies kenntlich und kommentieren Sie diese Bewertung kritisch.

3.6.3. Kohärenz – Der rote Faden

In Ihrer Arbeit sollte durchgängig ein „roter Faden“ zu erkennen sein. Das erreichen Sie, indem Sie sich zunächst selbst bewusst machen, welche Rolle die einzelnen Abschnitte innerhalb Ihrer Arbeit spielen und wie Sie miteinander zusammenhängen. Bemühen Sie sich dann, diese Zusammenhänge auch dem Leser deutlich zu machen.

Zu Beginn eines neuen Abschnitts können Sie diesen in einem Satz einleiten (z.B. „Im folgenden Abschnitt machen wir uns mit den mathematischen Grundlagen des Algorithmus aus Kapitel 3.5 vertraut.“). Innerhalb eines Abschnitts können Sie einzelne Sätze durch geeignete Diskurskonnectoren verbinden, die die Relation zwischen den Sätzen deutlich machen. Wenn Sie also zwei konträre Aussagen von zwei Autoren einander gegenüberstellen, schreiben Sie nicht einfach „Müller sagt... Schmidt sagt...“, sondern schließen Sie den zweiten Satz z.B. durch „Im Gegensatz dazu behauptet Schmidt...“ an.

Als Faustregel können Sie sich überlegen, ob ein Leser, der nur die Einleitungssätze jedes Kapitels liest, danach grob verstanden hat, worum es in Ihrer Hausarbeit geht.

3.6.4. Terminologie

Übernehmen Sie nicht einfach Fachbegriffe aus den Artikeln, über die Sie schreiben. Bei allen Fachtermini in Ihrer Arbeit sollten Sie sich fragen, ob sie (im Bereich der Computerlinguistik) Allgemeingut sind, oder durch den zitierten Autor eingeführt wurden. Falls die Begriffe kein Allgemeingut sind, müssen Sie sie in Ihrer Arbeit einführen und angeben, wer sie in welcher Bedeutung benutzt oder geprägt hat. Achten Sie insbesondere darauf, ob verschiedene Autoren denselben Begriff mit unterschiedlichen Bedeutungen verwenden. In diesem Fall müssen Sie klarstellen, nach wessen Terminologie Sie sich richten.

3.6.5. Fußnoten

Versuchen Sie die Verwendung von Fußnoten nach Möglichkeit zu vermeiden. Geben Sie insbesondere keine ausführlichen Anmerkungen etc. in Fußnoten an. Sie können Fußnoten beispielsweise verwenden, um auf Webseiten zu verweisen. In der Computerlinguistik ist die Benutzung von Fußnoten für Literaturverweise nicht üblich (vgl. Kapitel 4).

3.7. Wenn Sie nicht weiterkommen

Wenn Sie mit Ihrer Arbeit nicht weiterkommen:

- Wenn Sie mit einem Ihrer Papiere Probleme haben, weil beispielsweise Grundlagen nicht klar sind, kann es helfen, in einem Standard-Lehrbuch nach dem entsprechenden Thema zu suchen oder auf der Literaturliste der Veranstaltung nach Grundlagenliteratur zu suchen.
- Sprechen Sie Ihren Dozenten an. Er kann Ihnen sowohl fachlich, als auch formal helfen. Er weiß z.B. am besten, welchen Grad der Ausführlichkeit er bei einem Thema erwartet und kann Ihnen Tipps geben, wenn Sie bei der Literaturrecherche steckenbleiben.
- Vernetzen Sie sich mit Kommilitonen aus dem gleichen Seminar, die an einem ähnlichen Thema arbeiten.

4. Referenzen

Dieses Kapitel gibt nur einige kurze Hinweise zum korrekten Zitieren in der Computerlinguistik. Es gibt zahlreiche gute Ratgeber zum wissenschaftlichen Arbeiten im Allgemeinen, die weiterführende Informationen enthalten. Zum Beispiel:

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

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

4.1. Zitieren im Text

- In der Computerlinguistik 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

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

- Das Zitat muss von einer Referenz begleitet sein, die besteht aus
 - dem Nachnamen des Autors (bzw. der Autoren oder des Erstautors plus “et al.”) und
 - der Jahreszahl.

4.1.1. Direkte Zitate

In der Computerlinguistik möglichst vermeiden! Eigentlich nur sinnvoll bei:

- Definitionen oder anderen Ideen, bei denen es auf den exakten Wortlaut der Formulierung ankommt.
- 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 mit dem Original übereinstimmen, sowohl bezüglich des Inhalts als auch bezüglich der Orthografie und des Satzbaus (Hervorhebungen, Tippfehler etc. müssen beibehalten werden). Wenn Abweichungen vom Original sinnvoll oder notwendig sind, müssen diese eindeutig gekennzeichnet sein.
- Bei einem direkten Zitat müssen Sie zusätzlich zu Autoren und Jahreszahl die Seite angeben, auf der das Zitat im Originaltext steht.

Achtung: Um korrekt zitieren zu können, müssen Sie **sicherstellen, dass Ihnen tatsächlich die Originalquelle** in ihrer maßgeblichen Fassung **vorliegt**. Vorsicht ist geboten bei Publikationen, die z.B. von den 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 von der gedruckten Fassung abweichen können. Wenn Sie direkt zitieren wollen, besorgen Sie sich die Originalquelle (aus der Bib oder wenn elektronisch, dann direkt von der Webseite der entsprechenden Zeitschrift bzw. aus der ACL Anthology etc.).

4.1.2. Indirekte Zitate

Indirekte Zitate kommen sehr häufig vor und geben fremdes Gedankengut oder Forschungsergebnisse anderer Wissenschaftler 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 und
- 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)”.

4.1.3. Besondere Quellen

Lehrbücher Lehrbücher wie Jurafsky und Martin [JM08] sollten nur in Ausnahmefällen zitiert werden (bei Proseminararbeiten noch gerade ok, bei Masterarbeiten eher vermeiden). Wenn Sie auf spezielle computerlinguistische Methoden verweisen wollen, sollten Sie nach Möglichkeit die Original- oder Standardquelle ausfindig machen. Z.B. kann man für Centering Theory Grosz et al. (1995)² zitieren, und nicht das entsprechende Kapitel aus Jurafsky und Martin. Wenn Sie nicht sicher sind, welches die Originalquelle ist, schauen Sie sich Publikationen an, die sich auf die Methode beziehen, und sehen Sie nach, welche Quelle(n) dort genannt wird. Einige Methoden und Techniken können zudem als Standard vorausgesetzt werden und benötigen daher nicht unbedingt eine gesonderte Referenz, z.B. TF.IDF.

Sekundärquellen Lehrbücher sind ein spezieller Fall von Sekundärquellen, d.h. Quellen, die andere Quellen lediglich zusammenfassen und aufbereiten. Sekundärquellen sind in der Computerlinguistik generell zu vermeiden. D.h. wenn Brown (2010) auf eine andere Quelle, Smith (2005), verweist, sollten Sie direkt auf Smith (2005) verweisen. Dies setzt voraus, dass Sie sich Smith (2005) im Original besorgen und die relevanten Passagen lesen! Sie dürfen sich 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 dürfen Sie die Sekundärquelle verwenden, müssen jedoch im Text deutlich machen, dass Ihnen die Primärquelle nicht vorlag. Z.B. “Smith (2005, S. 10) zitiert nach Brown (2010, S. 46)”.

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

- Schauen Sie 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 Parsers³ einen Abschnitt “Citing the Stanford Parser”, in

²Grosz, 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.

³<http://nlp.stanford.edu/software/lex-parser.shtml>

dem genau aufgelistet ist, welche Publikationen für welche Aspekte des Parsers zitiert werden sollten.

- Wenn die Webseite nicht weiterhilft, schauen Sie sich 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, finden Sie heraus, wie andere Wissenschaftler auf diese Software/dieses Korpus verweisen.

Wenn Sie keine Publikation finden können (manchmal gibt es keine) oder wenn Sie zusätzlich zur Publikation noch die Webseite angeben möchten, können Sie dies als Fußnote im Text machen (nicht die URL im Literaturverzeichnis auflisten). In sehr, sehr seltenen Fällen kann es sein, dass Sie auf eine Internetquelle im Literaturverzeichnis verweisen müssen, z.B. weil ein richtungsweisender Aufsatz nie veröffentlicht worden ist oder weil Sie auf unveröffentlichte Annotationsrichtlinien etc. verweisen wollen. In diesem Fall sollten Sie 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 sollten Sie dann noch die URL angeben, mit einem Verweis darauf, wann Sie sie zum letzten Mal besucht haben.

4.2. Das Literaturverzeichnis

Die Quellen im Literaturverzeichnis sollten nach dem Nachnamen des Erstautors geordnet sein. 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:

- **Monografien:** 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 Computerlinguistik 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 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.
- 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. Achten Sie 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
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:

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

5. Nach dem Schreiben

Lassen Sie idealerweise (wenn der Abgabetermin es zulässt) Ihre Arbeit ein paar Tage liegen und lesen Sie sie mit etwas Abstand noch einmal. Wenn Sie etwas dann nicht auf Anhieb verstehen, versteht es jemand anderes mit großer Wahrscheinlichkeit auch nicht. (Und denken Sie nicht, Ihr Dozent wird schon wissen, was Sie gemeint haben. Das ist an dieser Stelle irrelevant, Ihr Dozent kennt das Thema sowieso.)

Oft findet man seine eigenen Fehler nicht selbst; es ist sinnvoll, seine Hausarbeit von Kommilitonen Korrektur lesen zu lassen. Die folgenden Punkte können Sie selbst durchgehen:

- Sind alle wichtigen formalen Teile (siehe oben) vorhanden?
- Lassen Sie noch einmal eine Rechtschreibkorrektur über Ihr Programm laufen.
- Vereinfachen Sie gegebenenfalls „Bandwurmsätze“.
- Kontrollieren Sie nochmal, ob Sie alle Quellen angegeben haben und ob das Literaturverzeichnis vollständig ist.

- Sind Grafiken, Anhänge etc. richtig formatiert und eingebunden?
- Schriftbild: Vermeiden Sie einzelne Zeilen eines Absatzes, die durch einen Seitenbruch vom Rest des Absatzes getrennt sind. (Auch das macht LaTeX automatisch.)
- Einheitlichkeit überprüfen: Wie oben bereits geschrieben, gibt es sehr oft verschiedene Möglichkeiten etwas zu tun (Formatierung, Literaturverzeichnis, Schreibstil...), ohne dass eine davon „richtiger“ wäre als die andere. Wichtig ist dann vor allem, einheitlich zu bleiben und die einmal getroffene Entscheidung konsequent für die ganze Hausarbeit umsetzen. (Auch hier hilft LaTeX)

A. Nützliche Latex-Pakete

Hier finden Sie eine Liste von nützlichen Latex-Paketen. Die genauen Befehle entnehmen Sie bitte den Dokumentationen der Pakete:

- avm für Attribut-Wert-Matrizen:
<http://nlp.stanford.edu/~manning/tex/avm-doc.pdf>
- qtree oder pstricks/pst-tree für Bäume:
<http://www.ling.upenn.edu/advice/latex/qtree/qtreenotes.pdf>
<http://www.essex.ac.uk/linguistics/external/clmt/latex4ling/trees/pstrees/index.pdf>
- gb4e für linguistische Beispiele:
<ftp://ftp.dante.de/tex-archive/macros/latex/contrib/gb4e/gb4e-doc.pdf>
- tipa für phonetische Symbole (IPA):
<http://www.l.u-tokyo.ac.jp/~fkr/tipa/tipaman.pdf>
- finomaton für Automaten: <http://www.ctan.org/pkg/finomaton>
- koma-script für verschieden Layoutformate:
<http://tug.ctan.org/tex-archive/macros/latex/contrib/koma-script/>

B. Weitere nützliche Tools

- Ein Applet, das Bäume aus Klammernotationen erstellt. Die Ergebnisse kann man als Grafik abspeichern und dann einbinden:
<http://ironcreek.net/phpsyntaxtree/>
- Graphen aller Art (Bäume, Automaten, ...) lassen sich mit dot erstellen. (sehr umfangreich, erstellt Vektorgrafiken):
<http://www.graphviz.org/doc/info/lang.html>

Literatur

- [CEJ⁺09] Kai-Uwe Carstensen, Christian Ebert, Susanne Jekat, Cornelia Ebert, Hagen Langer, and Ralf Klabunde. *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Springer, 2009. ISBN 3827420237.
- [JM08] Daniel Jurafsky and James H. Martin. *Speech and Language Processing (2nd Edition) (Prentice Hall Series in Artificial Intelligence)*. Prentice Hall, 2 edition, 2008.
- [Lam95] Leslie Lamport. *Das LaTeX-Handbuch*. Addison-Wesley, 1995.